



Journal of Philosophical Investigations



University of Tabriz

Special Issue:

# **Kant's Philosophy in 21<sup>st</sup> century**

Guest Editor of This Issue:

**Prof. Robert Hanna**

**Journal of Philosophical Investigations**

**Volume 18, Issue 47, Summer 2024**

# *Journal of Philosophical Investigations*

ISSN (print): **2251-7960**

ISSN (online): **2423-4419**

**Department of Philosophy- University of Tabriz- Iran**

Founder: **University of Tabriz**

General Manager: **Dr. Hassan Fathi**, *Associate Professor of western philosophy, University of Tabriz*

Editor in chief: **Dr. Muhammad Asghari**, *Professor of philosophy, University of Tabriz*

Editorial board:

|                               |                                                                                          |
|-------------------------------|------------------------------------------------------------------------------------------|
| <b>Susan Haack</b>            | <i>Professor of law &amp; philosophy, Miami University- USA</i>                          |
| <b>Galen Strawson</b>         | <i>Professor of philosophy, Texas University at Austin- USA</i>                          |
| <b>Mark Van Dan Bossche</b>   | <i>Professor of western philosophy, Vrije University- Brussel</i>                        |
| <b>Slavoj Žižek</b>           | <i>Professor of philosophy at The European Graduate School/ EGS - Switzerland</i>        |
| <b>Rabert J. Dostal</b>       | <i>Professor of western philosophy of Bryn Mawr College - USA</i>                        |
| <b>Simon Critchley</b>        | <i>Professor of philosophy at The European Graduate School/ EGS - Switzerland</i>        |
| <b>Lawrence Cahoon</b>        | <i>Professor of philosophy at College of The Holy Cross- USA</i>                         |
| <b>Thomas Sheehan</b>         | <i>Professor of religious studies and philosophy, University of Stanford - USA</i>       |
| <b>James M Magrini</b>        | <i>Professor of philosophy and religious studies at College of DuPage - USA</i>          |
| <b>Mahmoud Navali</b>         | <i>Professor of western philosophy, University of Tabriz</i>                             |
| <b>Morteza Shajari</b>        | <i>Professor of Islamic philosophy, University of Tabriz</i>                             |
| <b>Mansour Imanpour</b>       | <i>Associate Professor of Islamic philosophy, Shahid Madani University of Azarbaijan</i> |
| <b>Abdolrazzaq Hessamifar</b> | <i>Professor of western philosophy, Imam Khomeini University of Qazvin</i>               |
| <b>Abdolkarim Rashidian</b>   | <i>Associate Professor of western philosophy, Shahid Beheshti's University</i>           |
| <b>Mostafa Shahrayini</b>     | <i>Associate Professor of western philosophy, University of Tabriz</i>                   |
| <b>Majid Sadremajles</b>      | <i>Associate Professor of western philosophy, University of Tabriz</i>                   |
| <b>Hassan Fathzade</b>        | <i>Associate Professor of western philosophy, University of Zanjan</i>                   |
| <b>Yousef Nouzohoor</b>       | <i>Associate Professor of western philosophy, Allameh Tabataba'i University</i>          |
| <b>Alireza Farnam</b>         | <i>Professor of psychiatry, Tabriz University of Medical Sciences</i>                    |
| <b>Mahmoud Soufiani</b>       | <i>Assistant Professor of western philosophy, University of Tabriz</i>                   |

Internal Manager: **Dr. Mahmoud Soufiani**, *Assistant Professor of western philosophy, University of Tabriz*

Publisher: **University of Tabriz**

Address: **Faculty of Persian Literature & Foreign Languages, University of Tabriz, Tabriz, Iran.**

Tel: **00984133392131**

Fax: **00984133342490**

Website: **<https://philosophy.tabrizu.ac.ir>**

E-mail: **[philosophy@tabrizu.ac.ir](mailto:philosophy@tabrizu.ac.ir)**

Indexing Databases: **Scopus, ISC, Doaj, Philpapers, Oaji, Accademia, Sindexs, Magiran, Noormags, Erihplus, Cosmos, Scirev, Portal.issn, Civilica, Europub, Researchbib, Ensani.ir, Road, Indexcopernicus**

## Table of Contents

|                                                                                                     |     |
|-----------------------------------------------------------------------------------------------------|-----|
| <b>Editorial Note</b>                                                                               |     |
| Kant's Philosophy in the 21 <sup>st</sup> Century .....                                             | I   |
| <b>Robert Hanna</b>                                                                                 |     |
| Kantian Futurism .....                                                                              | 1   |
| <b>Stephen R. Palmquist</b>                                                                         |     |
| Interpreting Kantian Religious Judgment for the Twenty-First Century .....                          | 9   |
| <b>Daniel Dal Monte</b>                                                                             |     |
| Kant's Humanism: A Loophole in the Principle of Sufficient Reason .....                             | 29  |
| <b>Hemmo Laiho</b>                                                                                  |     |
| Does Kantianism Imply Some Sort of Conceptual Creationism? .....                                    | 49  |
| <b>Willem McCloud</b>                                                                               |     |
| Kant, Niels Bohr and Quantum Spontaneity .....                                                      | 63  |
| <b>Mohamad Mehdi Moghadas &amp; Ali Fath Taheri</b>                                                 |     |
| The Kantian Self versus Pattern Theory of the Self .....                                            | 89  |
| <b>Otto Paans</b>                                                                                   |     |
| Grasping the Grounds of Thought: The Thing-in-Itself, Actancy and Ecology .....                     | 111 |
| <b>Pereira Roberto</b>                                                                              |     |
| Consciousness and Cognition in Kant' First Critique .....                                           | 139 |
| <b>Bernard Riley</b>                                                                                |     |
| The Mathematical Basis of the Phenomenal World .....                                                | 161 |
| <b>Romeu Rossi Junior &amp; Patricia Kauark-Leite</b>                                               |     |
| A Kantian Solution for the Freedom of Choice Loophole in Bell Experiments .....                     | 189 |
| <b>Edgar Valdez</b>                                                                                 |     |
| On the Architectonic Idea of Mathematics .....                                                      | 203 |
| <b>Maria Borges</b>                                                                                 |     |
| Many Shades of Love in Kant .....                                                                   | 219 |
| <b>Reena Kumari; Madhu Mangal Chaturvedi &amp; Ravi Kumar</b>                                       |     |
| Reason(s) have Weight with the Evidence of Practical Reason .....                                   | 233 |
| <b>Alireza Mansouri</b>                                                                             |     |
| Kant's Moral Theory Meets Evolutionary Theory .....                                                 | 251 |
| <b>Ana Cristina Falcato</b>                                                                         |     |
| Shame and 'Shame Instinct' in Kant's Pre-Critical Texts; RH: Shame and shame instinct in Kant ..... | 265 |
| <b>Dennis Schulting</b>                                                                             |     |
| Maturity, Freedom of Thought and Emancipation on Kant's What Is Enlightenment? .....                | 281 |
| <b>Vadim Chaly</b>                                                                                  |     |
| Kantian Fallibilist Ethics for AI Alignment .....                                                   | 303 |



Journal of Philosophical Investigations

## Journal of Philosophical Investigations

Print ISSN: 2251-7960    Online ISSN: 2423-4419

Homepage: <https://philosophy.tabrizu.ac.ir>



University of Tabriz

### Editorial Note

#### Kant's Philosophy in the 21<sup>st</sup> Century

Robert Hanna 

**Guest Editor,** *Journal of Pazhoeshha-ye Falsafi* (The Journal of Philosophical Investigations)

Independent Philosopher, Canada, Email: [bobhannahbob1@gmail.com](mailto:bobhannahbob1@gmail.com)

The 21<sup>st</sup> century is a turning point for humankind. Its principal existential challenges or threats include disastrous climate change; addicted and uncontrolled use of digital technology, and the hegemony of what I have called *the myth of artificial intelligence*; the return of fascist politics and war mongering; and the ongoing threat of nuclear annihilation. If humankind is to survive and flourish, then it must have an overarching plan that is guided by an adequate, fundamental, and synoptic philosophy. What sort of philosophy? It is arguable that only an updated version or versions of the 18<sup>th</sup> century Critical philosophy of Immanuel Kant can do the job. Correspondingly and to that end, this issue of the *Journal of Philosophical Investigations* is devoted to the special topic "Kant's Philosophy in the 21<sup>st</sup> Century."

The core of Kant's Critical philosophy is *transcendental idealism*: the thesis that necessarily, the manifestly real world structurally conforms to our innately specified cognitive (intellectual and sensible), affective, and practical powers, and not the converse. If transcendental idealism is true, then the the natural universe is our metaphysical, epistemic, and practical home, and it cannot be adequately explained without reference to the real possibility of rational human animals.

Against that backdrop, the articles in this collection amply testify to the intrinsic significance of Kant's Critical philosophy, and above all to its contemporary and future relevance. There are articles on Kant's metaphysics, epistemology, and philosophy of science (articles 2-9); on his ethics and practical philosophy (articles 10-12); and on his social and political theory (article 13-14). Taken collectively, they demonstrate that Kant's Critical philosophy still has defensible and original things to say about the principle of sufficient reason and the thing-in-itself, cognitive



semantics and consciousness, free will and quantum mechanics, the nature of the self, the nature of human love, practical reasons, meta-ethics, the nature of shame, and enlightenment. Indeed, perhaps surprisingly, Kant's 18<sup>th</sup> century Critical philosophy is uniquely suited to be *the philosophy of the future* (article 1).

This is because *all* Anglo-American-&-European philosophy since Kant—i.e., since the end of the 18<sup>th</sup> century—is *post-Kantian*, in the substantive sense that all Anglo-American-&-European philosophy since the end of the 18<sup>th</sup> century falls within a single comprehensive *Ur*-framework, according to which Kant's philosophy is either (1) wholly accepted without revision-or-updating, (2) at least partially accepted but also significantly revised-&-updated, or (3) outright rejected. However, whether Kant's philosophy is wholly accepted, partially accepted, or outright rejected, *it is inescapable*. This is simply because Kant's philosophy determines the total logical space of relevant philosophical options for all post-Kantian Anglo-American-&-European philosophy. In this sense, all post-Kantian Anglo-American-&-European philosophy, including of course all contemporary philosophy up to 6am this morning, *has come out from under Kant's wig*, whether positively (pro-) or negatively (anti-). Therefore, all foreseeably future philosophy Anglo-American-&-European philosophy will be a series of positive or negative footnotes to Kant, and the articles in this special topics collection provide fourteen different previews of that philosophy.



Journal of Philosophical Investigations

## Journal of Philosophical Investigations

Print ISSN: 2251-7960 Online ISSN: 2423-4419

Homepage: <https://philosophy.tabrizu.ac.ir>



University of Tabriz

### Kantian Futurism

Robert Hanna 

Independent philosopher, Canada, Email: [bobhannahbob1@gmail.com](mailto:bobhannahbob1@gmail.com)

#### Article Info

##### Article type:

Research Article

##### Article history:

Received 12 July 2024

Received in revised form  
14 July 2024

Accepted 14 July 2024

Published online 07  
August 2024

##### Keywords:

Kant, meta-philosophy,  
humankind, futurism

#### ABSTRACT

The future of philosophy and the future of humankind-in-the-world are intimately related, not only (i) in the obvious sense that all philosophers are “human, all-too-human” animals—i.e., members of the biological species *Homo sapiens*, and also finite, fallible, and thoroughly normative imperfect in every other way too—hence the natural fate of all human animals is also the natural fate of all philosophers, but also (ii) in the more profound and subtle sense of what I’ll call *philosophical futurism*. Philosophical futurism is a critical, synoptic, and speculative reflection on the fate of humankind-in-the-world, with special attention paid not only to what humankind-in-the-world (including philosophy itself) will most likely be, if things continue to go along in more or less the same way as they have been and are now going, or could conceivably be, as in science fiction or other forms of imaginative projection, but also to what what humankind-in-the-world (including philosophy itself) *ought* to be, and therefore (assuming that “ought” entails “can”) *can* be, as the direct result of our individual and collective free agency, for the purpose of rationally guiding humankind in the near future. In this essay, I very briefly present, defend, and strongly recommend a version of philosophical futurism that I call *Kantian futurism*.

**Cite this article:** Hanna, R. (2024). Kantian Futurism. *Journal of Philosophical Investigations*, 18(47), 1-8.  
<https://doi.org/10.22034/jpiut.2024.18254>



© The Author(s).

<https://doi.org/10.22034/jpiut.2024.18254>

Publisher: University of Tabriz.

The future of philosophy and the future of humankind-in-the-world are intimately related, not only (i) in the obvious sense that all philosophers are “human, all-too-human” animals—i.e., members of the biological species *Homo sapiens*, and also finite, fallible, and thoroughly normative imperfect in every other way too—hence the natural fate of all human animals is also the natural fate of all philosophers, but also (ii) in the more profound and subtle sense of what I’ll call *philosophical futurism*. Philosophical futurism is a critical, synoptic, and speculative reflection on the fate of humankind-in-the-world, with special attention paid not only to what humankind-in-the-world (including philosophy itself) *will most likely be*, if things continue to go along in more or less the same way as they have been and are now going, or *could conceivably be*, as in science fiction or other forms of imaginative projection, but also to what what humankind-in-the-world (including philosophy itself) *ought to be, and therefore* (assuming that “ought” entails “can”) *can be*, as the direct result of our individual and collective free agency, for the purpose of rationally guiding humankind in the near future. In what follows, I will very briefly present, defend, and strongly recommend a version of philosophical futurism that I call *Kantian futurism*.<sup>1</sup>

Immanuel Kant’s *Critique of Pure Reason*, aka *Kritik der reinen Vernunft*, aka “the first *Critique*,” and also “CPR” for short, is the single most brilliant, difficult, and important book in modern philosophy; although to be sure, the CPR has always provoked as much controversy, criticism, and disagreement as it has inspired agreement, application, and creative revision-&-updating. The CPR was first published in 1781 (the edition), followed six years later by a revised second edition (the B edition) in 1787. Whole libraries could be filled with the secondary literature—books and essays—published on the first *Critique*, since its explosive appearance on the philosophical scene in the twilight decades of the 18<sup>th</sup> century. Correspondingly, during those 240+ years, the impact of the CPR, via Kant’s comprehensive *Critical Philosophy*, has steadily deepened, increased, ramified, and widened, in three directions: **first**, as a positive source for foundational work in metaphysics, epistemology, philosophy of mind, ethics, aesthetics and the philosophy of art, philosophical anthropology, the philosophy of religion and philosophical theology, political philosophy, the history and philosophy of the formal and natural sciences, and the philosophy of nature, **second**, as a negative foil for *all* alternative developments in those basic philosophical subjects, since *all* philosophers in the post-Kantian era, whether pro-Kant, anti-Kant, or officially Kant-neutral, must define themselves and their work in relation to Kant and the Critical Philosophy (Hanna, 2020), and **third**, as a broad, deep, and powerful sociopolitical force, under the much-controverted rubric of *enlightenment* (Hanna, 2016, 2017a, 2021a).

---

<sup>1</sup> This short essay has two different and longer versions: **first**, a same-named but much longer essay, “Kantian Futurism” (Hanna, 2023), in which I present, defend, and strongly recommend three futuristic Kantian ideas: (i) *moderately anthropic cosmology* aka *weak transcendental idealism* aka *top-down cosmology*, (ii) *cosmological and cognitive organicism*, and (iii) *the moral and sociopolitical primacy of human dignity*; and **second**, “Kant, Williamson, and The Future of Analytic Philosophy” (Hanna, 2024a), in which I explicitly relate Kantian futurism to classical and post-classical Analytic philosophy.

Indeed, in this very year, 2024, the 300<sup>th</sup> anniversary of Kant's birth, it's self-evident that the impact of the CPR, via the larger Critical Philosophy and post-Kantian enlightenment thinking, is now truly *cosmopolitan*, hence worldwide, and also by no means restricted to what Arthur Schopenhauer in the mid-19<sup>th</sup> century so sarcastically and presciently called "university philosophy" (Schopenhauer, 2014), i.e., *professional academic philosophy*, nowadays known simply as "professional philosophy." This is an oxymoron if ever there was one, since that is precisely how Plato defined *sophistry*, as philosophy for pay and inherently controlled by conventional social and political norms. By 2024, Kant's philosophical thinking and writing are fully and globally *in the real world*, and thereby woven as efficaciously and tightly into the warp and woof of modern life as *real philosophy* can ever be. By "real philosophy", I mean *authentic, serious* philosophy, namely, philosophy done wholeheartedly for its own sake and focused on fundamental issues and problems flowing from the rational human condition in a nonideal natural and social world. That's as opposed to *inauthentic, superficial* philosophy, namely, philosophy done halfheartedly and essentially for the sake of one's career, and basically focused on scholastic, trivial issues—for example, contemporary professional academic philosophy (Hanna, 2022; see also Haack, 2021).

In that connection, it is ironic that Kant was the first truly great real philosopher who was *also* a professional academic philosopher, an irony presciently noted by Schopenhauer too. In his 1798 essay, "The Conflict of the Faculties," Kant rightly asserted the rational autonomy of the faculty of philosophy, not only from all the other faculties of the university, especially including the faculty of theology, but also from the government (CF 7: 27). But at the same time, Kant failed to anticipate the future heteronomy and subordination of mainstream 20<sup>th</sup> and 21<sup>st</sup> century philosophy *to the formal and natural sciences* and above all *to its own academicization and professionalization* (Hanna, 2022).

In any case, given the conception of real philosophy, I spelled out two paragraphs above, my metaphilosophical view about the CPR in particular and the Critical Philosophy in general is this:

Like all philosophers Kant sometimes errs, or anyhow nods. However, we respect him most by critically noting and then setting aside his slips, and by promoting his deepest and most powerful doctrines.... Kant's Critical Philosophy is fully worth studying, critically analyzing, charitably explicating, defending, and then independently developing in a contemporary context. This is because, in my opinion, more than any other single-authored body of work in modern philosophy the Critical Philosophy most doggedly pursues and most profoundly captures some non-trivial fragment of the honest-to-goodness truth about about rational human animals and the larger natural world that surrounds them (Hanna, 2006a, 7).



Lying behind and scaffolding those claims, my take on the relationship between contemporary philosophy and the history of philosophy is based on what I call *The No-Deep-Difference Thesis*:

There is no fundamental difference in philosophical content between the history of philosophy and contemporary philosophy (Hanna, 2017b: section 3).

In other words, in doing contemporary philosophy one is thereby directly engaging with the history of philosophy, and in doing the history of philosophy one is thereby directly engaging with contemporary philosophy. There is no fundamental difference in philosophical content between philosophy that was thought, spoken, written, or published 2000 or 200 or 20 years ago, or 20 months ago, and philosophy that was thought, spoken, written, or published 20 weeks or 20 days or 20 minutes or 20 seconds ago.

Relatedly, it's a truth *not* generally acknowledged, that *all* Anglo-American-&-European philosophy since Kant—i.e., since the end of the 18<sup>th</sup> century—is *post-Kantian*. This is of course trivially true, in that all Anglo-American-&-European philosophy since the end of the 18<sup>th</sup> century literally temporally succeeds the publication and dissemination of Kant's philosophical writings. But it's also profoundly true, in that all Anglo-American-&-European philosophy since the end of the 18<sup>th</sup> century falls within a single comprehensive *Ur*-framework, according to which Kant's philosophy is either (1) wholly accepted without revision-or-updating (*ortho-Kantianism*), (2) at least partially accepted but also significantly revised-&-updated (*quasi-Kantianism*, *crypto-Kantianism*, and classical 19<sup>th</sup> and early 20<sup>th</sup> century *neo-Kantianism*, whose original rallying cry was: *back to Kant!*), or (3) outright rejected (*anti-Kantianism*) (Hanna, 2008, 2020).

The paradigmatic example of ortho-Kantianism is mainstream late 18<sup>th</sup>, 19<sup>th</sup>, 20<sup>th</sup>, and 21<sup>st</sup> century Kant-scholarship, allowing of course for many and various domestic or in-house scholarly disagreements about how best or correctly to interpret Kant's writings. Paradigmatic examples of quasi-Kantian philosophy include: classical German idealism (Fichte, Schelling, Hegel, etc.); British neo-Hegelianism (Bradley, McTaggart, etc.); realistic phenomenology, transcendental phenomenology, and existential phenomenology (Brentano, Husserl, Heidegger, Sartre, Merleau-Ponty, etc.); other varieties of post-phenomenological "Continental" philosophy (existentialism, hermeneutics, post-structuralism, deconstructionism, postmodernism, etc.); New England transcendentalism (especially Emerson); classical American pragmatism (especially Peirce); process philosophy (especially Bergson and Whitehead); and Pittsburgh neo-Hegelianism (especially Sellars, McDowell, and Brandom). The paradigmatic example of crypto-Kantianism is Wittgenstein's philosophy, both early and late (Hanna, 2017c). In addition, obviously, classical 19<sup>th</sup> and early 20<sup>th</sup> century German and French neo-Kantianism are paradigmatic examples of neo-Kantianism. As to anti-Kantian philosophy, paradigmatic examples are classical Analytic philosophy and post-classical Analytic philosophy (Hanna, 2001, 2006a, 2021b).

Nevertheless, whether Kant's philosophy is wholly accepted, partially accepted, or outright rejected, *it is inescapable*. This is simply because Kant's philosophy determines the total logical

space of relevant philosophical options for all post-Kantian Anglo-American-&-European philosophy. In this sense, all post-Kantian Anglo-American-&-European philosophy, including of course all contemporary philosophy up to 6am this morning, *has come out from under Kant's wig*, whether positively (pro-) or negatively (anti-).

Now, it must be admitted that it is at least possible that some old or new *non*-Anglo-American-&-European philosophical framework will unexpectedly stride into the center of the global intellectual and sociocultural scene like an all-conquering Colossus, and henceforth dominate philosophy worldwide. Let us call this *the extra-Kantian-philosophy possibility*. But the extra-Kantian-philosophy possibility seems to me *extremely unlikely*, in view of the bumpy (to put it very mildly) yet relentless *Americanization* of world culture, driven by the USA's militaristic adventures and misadventures, music, movies, television, and digital technology, and the correspondingly equally bumpy yet equally relentless *neoliberalization* of world politics, driven by technocratic capitalism, whether corporate capitalism or State-capitalism, and whether democratic or not-so-democratic, since the end of World War II.

Therefore, leaving aside the extra-Kantian-philosophy possibility, then all foreseeably future philosophy worldwide *will be a series of positive or negative footnotes to Kant*. Moreover, as regards negative footnotes, the 140-year-long anti-Kantian tradition of Analytic philosophy is in fact now coming to an end, as post-classical Analytic philosophy crashes, burns, and goes down forever into the ash-heap of history. In addition, as regards positive footnotes, obviously ortho-Kantianism is historically and philosophically backward looking, not forward-looking. Therefore, (i) the times they are a-changing, and (ii) the near-future emergence of some or another creatively revised-&-updated version of Kant's philosophy, as the central and dominant world philosophy, is historically inevitable. For all these reasons, *forward to Kant!* Must be humankind's philosophical futurist rallying cry.

Now, briefly, what does Kant's Critical Philosophy say? Focusing for the purposes of this short essay on Kant's *theoretical* philosophy, especially including logic, metaphysics, the theory of cognition, and epistemology (see, e.g., Hanna, 2021c, 2021d, 2021e, 2021f, 2021g, 2021h)—as opposed to his practical philosophy, including his moral theory, his political theory, and his theory of enlightenment—what I call Kant's *real metaphysics* (Hanna, 2017d), aka his “transcendental idealism,” is an a priori conceptual science describing the essential nature of the manifestly real world, insofar as that world necessarily conforms to the a priori formal constraints and normative structures of our two basic innately-specified cognitive faculties or powers: (1) “sensibility” (*Sinnlichkeit*), the capacity for pure or empirical sensory intuition (*Anschauung*), sense-perception, and imagination, and (2) “understanding” (*Verstand*), the capacity for conceptualizing and logical thinking (CPR A713/B741, A723/751, A841-842/B869-870).

Kant's real metaphysics of transcendental idealism is also a *cognitive semantics*, that is, a general theory of rational human representational content (C 10: 129-130; see also Hanna, 2001,

2006a). According to Kant, a representational content is *richly* meaningful, *empirically* meaningful, or “objectively valid,” if and only if it is either directly grounded on human sensible intuition or is an a priori necessary condition of a representational content that is directly grounded on human sensible intuition, hence *indirectly* grounded on human sensible intuition (CPR A38-40/B55-57, A239-240/B298-299). A judgment or proposition can have a truth-value, and be true or false, according to Kant, if and only if all of its representational components are objectively valid (CPR A58/B83). If any representational component of a judgment or proposition is *not* objectively valid, *not* empirically meaningful, or “empty” (*leer*), then the whole judgment or proposition is *not* objectively valid, not richly or empirically meaningful, and empty—“thoughts without content are empty” (CPR A 51/B 75)—and it has no truth-value, hence it is neither true nor false. By sharp contrast, if a judgment or proposition is objectively valid, hence richly or empirically meaningful, and thereby has a truth-value, then (i) it is synthetic if and only if (ia) its denial is logically and conceptually consistent and (ib) its meaning and truth are grounded on either pure or empirical human sensible intuition (CPR B73, A721/B749, C 11: 38, OD 8, 245), and (ii) it is analytic if and only if (iia) its denial entails either a formal logical contradiction, a formal contradiction in “intension” or *Inhalt*, or a mutual exclusion of “comprehension” or *Umfang* (CPR A150-153/B189-19, *Prol* 4, 267), (iib) it is necessarily true and a priori (CPR B4, A6/B10), and (iic) its meaning and necessary truth are grounded on essential connections between objectively valid, richly or empirically meaningful concepts. In other words, by means of his theoretical philosophy, Kant worked out a richly detailed, robust, and (I strongly believe) defensible theory of the analytic-synthetic distinction and philosophical analysis (Hanna, 2001, 2006a: chs. 5-7, 2015: chs. 4-8; see also Lewin, 2023a, 2023b).

That all being so, then self-evidently, this question must also be asked:

Precisely *how* should we creatively revise-&-update Kant’s 18<sup>th</sup> century philosophy for the purposes of bringing about the Kantian philosophy of the near future?

Humankind’s existential predicament—including its epistemic, metaphysical, logical, mathematical, natural-scientific, moral, sociopolitical, and religious or spiritual modes—in this nonideal natural and social world is what I call *the rational human condition*. Correspondingly, I call the general philosophical theory of the rational human condition *rational anthropology*. Since 2001, for better or worse, I have been developing, presenting, and defending a version of rational anthropology as the comprehensive Kantian philosophy of the near future (Hanna, 2015, 2018a, 2018b, 2018c, 2018d, and 2024b). Therefore, if I am right, then humankind’s philosophical futurist rallying cry—the rallying cry of Kantian futurism—must also be *forward to rational anthropology!*

## References

For convenience, I cite Kant's works infratextually in parentheses. The citations include both an abbreviation of the English title and the corresponding volume and page numbers in the standard "Akademie" edition of Kant's works: *Kants gesammelte Schriften*, edited by the Königlich Preussischen (now Deutschen) Akademie der Wissenschaften (Berlin: G. Reimer [now Walter De Gruyter], 1902). For references to the first Critique, I follow the common practice of giving page numbers from the A (1781) and B (1787) German editions only. Because the Akademie (AK) edition contains only the B edition of the first Critique, I have also consulted the following German composite edition: *Kritik der reinen Vernunft*, ed. W. Weischedel, Immanuel Kant Werkausgabe III (Frankfurt: Suhrkamp, 1968). Here's a list of the relevant abbreviations and English translations of the works I've cited in this essay, together with the year of each work's first publication in German, its Ak volume number, and (if relevant) its Ak page-range:

- Kant, I. (1973). *OD* "On a Discovery Whereby Any New Critique of Pure Reason Has Been Made Superfluous by an Older One. Translated by H. Allison. In H. Allison, *the Kant-Eberhard Controversy*. Pp. 107-160. Johns Hopkins University Press.
- Kant, I. (1979). *CF The Conflict of the Faculties*. Translated by M. Gregor. Lincoln, University of Nebraska Press.
- Kant, I. (1997). *CPR Critique of Pure Reason*. Translated by P. Guyer & A. Wood. Cambridge University Press. (1781 or an edition: AK 4: 1-251; 1787 or B edition: Ak 3)
- Kant, I. (1999). *Immanuel Kant: Correspondence, 1759-99*. Translated by A. Zweig, Cambridge University Press.
- Kant, I. (2004). *Prol Prolegomena to Any Future Metaphysics*. Translated by G. Hatfield. Cambridge University Press.
- Haack, S. (2021). Philosophy as a Profession, and as a Calling. *Syzetesis*, 8, 33-51.
- Hanna, R (2008). Kant in the Twentieth Century. In *Routledge Companion to Twentieth-Century Philosophy*. pp. 149-203, Routledge.
- Hanna, R. (2023). Kantian Futurism. Unpublished MS.
- Hanna, R. (2001). *Kant and the Foundations of Analytic Philosophy*. Oxford University Press.
- Hanna, R. (2006a). *Kant, Science, and Human Nature*. Oxford University Press.
- Hanna, R. (2006b). Rationality and the Ethics of Logic. *Journal of Philosophy*, 103, 67-100.
- Hanna, R. (2006c). *Rationality and Logic*. MIT Press.
- Hanna, R. (20115). *Cognition, Content, and the A Priori: A Study in the Philosophy of Mind and Knowledge*. Vol. 5, *The Rational Human Condition*, Oxford University Press.
- Hanna, R. (2016). Radical Enlightenment: Existential Kantian Cosmopolitan Anarchism, With a Concluding Quasi-Federalist Postscript. In *Join, Or Die: Philosophical Foundations of Federalism*. pp. 63-90, Edited by Heidemann & K. Stoppenbrink.
- Hanna, R. (2017a). Exiting the State and Debunking the State of Nature, *Con-Textos Kantianos*, 5, 167-189.
- Hanna, R. (2017b). Life-Changing Metaphysics: Rational Anthropology and its Kantian Methodology." In *The Cambridge Companion to Philosophical Methodology*. Edited by G. D'Oro & S. Overgaard, pp. 187-210, Cambridge University Press.

- Hanna, R. (2017c). Wittgenstein and Kantianism. In *Blackwell Companion to Wittgenstein*. Edited by H.-J. Glock, pp. 682-698, Blackwell.
- Hanna, R. (2017d). Kant, the Copernican Devolution, and Real Metaphysics. In *Kant Handbook*. pp. 761-789, Edited by M. Altman, Palgrave Macmillan.
- Hanna, R. (2018a). Preface and General Introduction, Supplementary Essays, and General Bibliography. *The Rational Human Condition*, Vol. 1. Nova Science.
- Hanna, R. (2018b). Deep Freedom and Real Persons: A Study in Metaphysics. *The Rational Human Condition*, Vol. 2. Nova Science.
- Hanna, R. (2018c). Kantian Ethics and Human Existence: A Study in Moral Philosophy. *The Rational Human Condition*, Vol. 3. Nova Science.
- Hanna, R. (2018d). Kant, Agnosticism, and Anarchism: A Theological-Political Treatise. *The Rational Human Condition*, Vol. 4. Nova Science.
- Hanna, R. (2020). The Kant Wars and the Three Faces of Kant. *Contemporary Studies in Kantian Philosophy*, 5, 73-94.
- Hanna, R. (2021a). The New Conflict of the Faculties: Kant, Radical Enlightenment, the Hyper-State, and How to Philosophize During a Pandemic, *Con-Textos Kantianos*, 13, 209-233.
- Hanna, R. (2021b). *The Fate of Analysis: Analytic Philosophy from Frege to the Ash-Heap of History*. Mad Duck Coalition.
- Hanna, R. (2021c). Axioms. In *Wuerth*, 58.
- Hanna, R. (2021d). Experience. In *Wuerth*, pp. 183-186.
- Hanna, R. (2021e). Representation. In *Wuerth*, pp. 388-390.
- Hanna, R. (2021f). Synthesis. In *Wuerth*, pp. 429-432.
- Hanna, R. (2021g). Synthetic a Priori. In *Wuerth*, pp. 432-434.
- Hanna, R. (2021h). Jäsche Logic. In *Wuerth*, pp. 707-711.
- Hanna, R. (2022). Six Studies in the Decline and fall of Professional Academic Philosophy, and a Real and Relevant Alternative. *Borderless Philosophy*, 5, 48-130.
- Hanna, R. (2023). *Kantian Futurism*. Unpublished MS.
- Hanna, R. (2024a). *Kant, Williamson, and the Future of Analytic Philosophy*. Unpublished MS.
- Hanna, R. (2024b). *Science for Humans: Mind, Life, the Formal-&-Natural Sciences, and a New Concept of Nature*. Springer Nature.
- Lewin, M & Williamson, T. (2023). Kant and Analysis. *Kantian Journal*, 42 (3), 49-73.
- Lewin, M. (2023a). Kant and Analysis. In *Lewin and Williamson*.
- Lewin, M. (2023b). Kant on Philosophy as Conceptual Analysis. *Con-Textos Kantianos*, 18, 11-20.
- Schopenhauer, A. (2014). On University Philosophy. In *Parerga and Paralipomena: Short Philosophical Essays*. Translated by S Roehr & C. Janaway. Vol. 1. pp. 125-176. Cambridge University Press.
- Wuerth, J. (Ed.). (2021). *Cambridge Kant Lexicon*. Cambridge University Press.



Journal of Philosophical Investigations

## Journal of Philosophical Investigations

Print ISSN: 2251-7960 Online ISSN: 2423-4419

Homepage: <https://philosophy.tabrizu.ac.ir>



University of Tabriz

# Interpreting Kantian Religious Judgment for the Twenty-First Century

Stephen R. Palmquist 

Independent Scholar, Los Angeles, California, USA. Email: [stevepq@associate.hkbu.edu.hk](mailto:stevepq@associate.hkbu.edu.hk)

### Article Info

#### Article type:

Research Article

#### Article history:

Received 23 July 2024

Received in revised form  
28 July 2024

Accepted 28 July 2024

Published online 07  
August 2024

#### Keywords:

Immanuel Kant, moral  
religion, affirmative  
interpretations, religious  
judgment, Douglas  
McGaughey and  
Lawrence Pasternack,  
religious practices,  
hermeneutics

### ABSTRACT

This article clears a path for employing the affirmative way of interpreting Kant's general theory of religion as a guidepost for twenty-first century religious practitioners. Before attempting such an employment, I correct several misconceptions regarding the affirmative way of interpreting Kant's theory that, if adopted as mainstream, would risk weakening its universal applicability. With this purpose in mind, I consider and respond to a misplaced criticism of my interpretation of Kant's theory of grace, advanced by Douglas McGaughey in 2013. I then assess the current state of a debate that has arisen between several key affirmative interpreters, concerning Kant's reference to the two "experiments" that guide the argument in his *Religion*. I give special attention to Lawrence Pasternack's 2017 attempt to "dismantle" this whole line of interpretation as one that "may very well not be worth our time." After identifying several serious flaws in Pasternack's argument, I conclude by backing a better alternative developed by Brandon Love. By clearing away two very different sorts of potential roadblocks that could stand in the way of a reader of Kant gaining a fair understanding of the affirmative approach to interpreting his theory of religion, these arguments prepare the way for a concluding attempt to catch a glimpse of how Kant's theory of religious judgment (especially his four guidelines for designing the constitution of a church), if adopted widely by practitioners of various historical religions, could benefit the fractured world we find ourselves in, a quarter of the way through the twenty-first century.

**Cite this article:** Palmquist, S. R. (2024). Interpreting Kantian Religious Judgment for the Twenty-First Century. *Journal of Philosophical Investigations*, 18(47), 9-28. <https://doi.org/10.22034/jpiut.2024.18292>



© The Author(s).

<https://doi.org/10.22034/jpiut.2024.18292>

Publisher: University of Tabriz.



## 1. The affirmative revolution in interpreting Kant's theory of religion

Of all the many and varied aspects of Immanuel Kant's theory of judgment, his theory of *religious* judgment has been among the least discussed by his interpreters. Those who mention his theory of religion at all often portray it in a thoroughly reductionistic way, as merely an application of his theory of how to make correct *ethical* judgments. In various publications over the past 40 years, I have made a concerted effort to offer an alternative to this reductionistic approach to interpreting Kant's theory of religion--an alternative I sometimes call an "affirmative" approach. Since the mid-1990s, the literature on Kant's theory of religion has mushroomed, to the point where we now witness a previously unprecedented number of internal debates between those interpreters who agree with the broad claim that Kant sets out to identify and affirm a specifically religious type of judgment and that the resulting theory of religion is worthy of our affirmation.

Before we proceed, a word of explanation is in order regarding my use of "religious judgment" in this article's title. As far as I am aware, Kant himself never used this expression; it certainly never appears in his influential 1793/1794 book, *Religion within the Bounds of Bare Reason*, the main work that sets out the details of his theory of religion. However, words with the root "judg" do appear over 100 times in the text of *Religion*.<sup>1</sup> I'll say more about some of these uses in §4, below. For now, note that these references to judgment in the context of Kant's theory of religion support a claim that many affirmative interpreters have backed: that the role filled by *Religion* in the overall architectonic structure of Kant's system is properly regarded as being a supplement or complement to the *third Critique* rather than to the second. That is, if we were to give *Religion* a title that more explicitly highlights its role in the overall structure of Kant's philosophical system, the title would not be something like *The Ethical Nature of All True Religion*, as traditional reductionist interpreters would have it, but instead something like *The Role of Religious Judgment in Empowering Moral Improvement*. For at no point in *Religion* is Kant's focus on expanding or even refining the moral philosophy put forward in his explicitly ethical writings (especially *Groundwork* and the second *Critique*); rather, his focus is on expanding and refining the reach of *judgment* (beyond the sort detailed in the third *Critique*), so that the role of certain types of judgment in assisting us to fulfill our moral mandate can be seen to include not only aesthetics and purposiveness in nature, but also religious beliefs, symbols, and rituals.

---

<sup>1</sup> I base this word count on the English translation of Kant's text that I cite throughout this article, which appears interspersed throughout the main text of my *Comprehensive Commentary on Kant's Religion within the Bounds of Bare Reason* (Chichester: Wiley, 2016). My translation revises Pluhar's 2008 translation. I count 111 words containing the root "judg" in the text of *Religion*: 33 refer not to the action of judging but to a (human or divine) person who serves as judge. That leaves 78 references to the action of judging in *Religion*. Most of these refer to specifically religious forms of judgment, such as the hypothetical situation whereby a divine being is believed to pass (or to be able to pass) judgment on humans, or where human beings judge a certain act or belief to be either in conformity with or in contradiction to the teachings of their historical faith.

In the prefaces to both editions of *Religion* Kant refers to an "experiment" he intends to conduct in the book's main text. In the second Preface he also introduces a metaphor of concentric circles to clarify the difference between the main experiment (which attempts to distinguish between historical/biblical religion and a set of pure conditions for the *possibility* of any true religion) and a "second experiment" (6:12), which takes fragments from one particular historical religion and holds them up to the system of pure religious reason to see if they overlap. When Kant refers to "judging" or "judgment" (78 times) in the text of *Religion*, he usually has this second experiment at least at the back of his mind, if not at the forefront. That is, religious judgment, in its philosophical (Kantian) sense, attempts to assess whether or not a particular historical religious belief, symbol, or ritual, aligns with the underlying principles of pure religion that will qualify that belief, symbol, or ritual as genuinely religious, if it passes the test. More of this later.

This article's main purpose is to clear a path for employing the affirmative way of interpreting Kant's general theory of religion, especially his account of religious judgment, as a guidepost for twenty-first century religious practitioners, regardless of the particular religious tradition they may belong to. Before attempting such an employment, we must correct several misconceptions regarding the affirmative way of interpreting Kant's theory that, if adopted as mainstream, would risk weakening its universal applicability. With this purpose in mind, the following three sections will proceed as follows.

In §2 I shall consider and respond to a criticism of my interpretation of Kant's theory of grace, as put forward in several publications but primarily in a 2010 paper entitled "Kant's Ethics of Grace" (hereafter "KEG") (Palmquist, 2010).<sup>1</sup> It would not be an understatement to say that on my interpretation, Kant's theory of divine grace serves as the cornerstone of his whole approach to religious judgment. As such, it is crucial to respond to a misplaced criticism of my interpretation advanced by Douglas McGaughey (2013).<sup>2</sup> In §3 I shall then assess the current state of a debate that has arisen between several key affirmative interpreters, concerning Kant's reference to the two "experiments" that guide the argument in his *Religion*. Special attention will be given to Lawrence Pasternack's influential 2017 attempt to "dismantle" this whole line of interpretation as one that "may very well not be worth our time." (Pasternack 2017).<sup>3</sup> After identifying several serious flaws in Pasternack's argument, I conclude by backing a better alternative developed by Brandon Love. By clearing away two very different sorts of potential roadblocks that could stand in the way of a reader of Kant gaining a fair understanding of the affirmative approach to interpreting his theory of religion, my arguments in §§2-3 will prepare the way for a concluding attempt to catch a glimpse

---

<sup>1</sup> See my article "Kant's Ethics of Grace: Perspectival Solutions to the Moral Difficulties with Divine Grace", *The Journal of Religion* 90 (2010), 530-53.

<sup>2</sup> Douglas McGaughey, "Historical and Pure Religion: A Response to Stephen Palmquist", *The Journal of Religion* 93 (2013), 151-176.

<sup>3</sup> Lawrence Pasternack, "The 'Two Experiments' of Kant's *Religion*: Dismantling the Conundrum," *Kantian Review* 22 (2017), 107,108.

of how Kant's theory of religious judgment, if adopted widely by practitioners of various historical religions, could benefit the fractured world we find ourselves in, a quarter of the way through the twenty-first century.

## 2. Kant's religious hermeneutics as the antidote to McGaughey's misplaced criticism

A key feature of Kant's theory of religious judgment, often neglected even by the growing number of affirmative interpreters, is his claim that in religious matters, everything depends on how believers/practitioners *interpret* the beliefs and rituals of their preferred historical religion. This emphasis is evident in a variety of ways throughout the text of *Religion* but reaches a crescendo in §VI of Division One of the Third Piece (*Stück*), entitled "Church Faith Has Pure Religious Faith as Its Highest Interpreter" (6:109). Kant there distinguishes between three ways of interpreting religious texts in a church (e.g., in sermons or childhood education): the theoretical/historical, the practical/ethical, and the judicial/experiential interpret according to empirical facts, moral lessons, or personal *feelings*, respectively. In a single concluding paragraph, he dismisses the third approach (later revived by Schleiermacher), leaving the first two as the only serious contenders to serve as the highest hermeneutic principle to guide the religious judgment of churchgoers. Kant concedes, in a nutshell, that matters of historical fact can serve as important contributors to religious education in a church--indeed, what church leaders teach must not *contradict* known scientific or historical facts--but the main thrust of one's teaching should be moral. When facing scriptural texts that pose potential ethical difficulties, such as when the Bible portrays God as commanding Abraham to kill his child, Kant says (6:110) that imposing a moral interpretation onto the text may often seem forced, [and] may often really be forced: and yet this interpretation must, if only it is possible that the text accepts it, be preferred to a literal interpretation that either contains within itself absolutely nothing for morality, or perhaps even acts counter to morality's incentives.

Nothing in my past publications, provided they are read in line with my own stated biases and intentions, challenges or counteracts the basic principle of Kantian religious hermeneutics that can be gleaned from the previous paragraph: for Kant "pure religion," or "rational religion" as he sometimes calls it, is necessarily grounded in humanity's *moral* nature. Nevertheless, in 2013 Douglas McGaughey's article in *The Journal of Religion* accused me of imputing to Kant 11 claims, all of which are--I actually agree!--contrary to Kant's actual intentions. In a few publications since that time, I have responded briefly to McGaughey's astonishing criticism of my position; (Palmquist, 2015: xin).<sup>1</sup> these brief previous responses could give the impression that I regard a simple "No!" as sufficient to refute McGaughey's claims. The rest of this section fills that lacuna by attempting to do what all good Kantian interpreters should do, whenever such a seemingly ungrounded dispute arises: I will highlight what is correct about McGaughey's interpretation, then identify why he needs to work harder to take on board some key features of the affirmative

---

<sup>1</sup> Most notably, see my *Comprehensive Commentary*, p.xin.

approach to interpreting Kant's theory of religion, if he is to avoid the dangers of the alternative, reductionist interpretation, which always stands in danger of throwing out the proverbial "baby" of moral improvement with the "bathwater" of historical religion.

Without providing any preliminary description of my approach to interpreting Kant or any textual evidence from my publications (except repeated general references to "KEG" and one short paragraph quoting a few passages out of context), McGaughey accuses me of defending 11 false claims about Kant's theory of religion. Without explanation, he simply lists these claims in the lefthand column of a table under the heading, "Palmquist". In the righthand column, under the heading "Another Kant", McGaughey (2013, 155-6) lists 11 corresponding claims that he presents as the correct versions of what he believes I have defended. Here are his 11 sets of claims, with numbers and letters ("A" for McGaughey's left column and "B" for the right) added by me:

1. A. "Religion is *historical* religion based on particular revelation" / B. "Religion is *pure* religion based on the universal gift of moral capacity that can neither be derived from nor reduced to the senses, which can be the core of any and all historical religion"
2. A. "Grace is a divine gift without an account of the 'ground' of humanity's moral capacity to which grace would add" / B. "Grace is speculative opinion, neither subjectively nor objectively sufficient, that undermines morality, which in turn is the internal, self-legislation of a noncodified moral order."
3. A. "Religion reduced by Kant to 'mere' morality if no Noumenal (divine) perspective" / B. "Religion is necessary for any and all human experience and anchored in the Noumenal gift of the capacity of creative freedom"
4. A. "'Exclusive Christology' of morally perfect individual as the historical archetype of the Logos external to the individual" / B. "'Inclusive Christology' of possible, universal, internal, moral improvement by which it is impossible to determine the moral status of an other"
5. A. "Christ is to be imitated" / B. "Morality, when at all, is at best served by examples, never by imitation, which only fosters discouragement"
6. A. "Religious hope: concerned with the speculative opinion (neither subjectively nor objectively sufficient) of 'parerga' of religion (i.e., based on ignorance as in Pascal's Wager; leads to personal torment)" / B. "Religious hope: Concerned with the mystery (subjectively sufficient) of *pure* religion (i.e., the capacity of goodness that is creative freedom can never be lost and remains the condition for any and all possible moral transformation)" [*sic*]
7. A. "Substitutes God's hypothetical decisions (as divine 'timeless choices') for humanity's categorical decisions" / B. "Embraces humanity's categorical capacity to create and add things (as 'timeless choices') that are not otherwise found in phenomenal reality"
8. A. "Moral improvement is concerned merely with the individual" / B. "Moral improvement is communal (involves Kingdom of Ends and culture that promotes the goodness of humanity) and is concerned with the improvement of the species as an open-ended process"

- 
9. A. "Morality measured by external consequences (consequentialism)" / B. "Morality measured by internal moral disposition known only to the individual (deontology)"
10. A. "Alternative: philosophy or religion (literal anthropomorphism, which we must embrace speculatively)" / B. "Alternative: none between pure philosophy and pure religion--regulative ideas of God, freedom, and the soul with respect to moral improvement in this life (symbolic anthropomorphism at most, which we embrace as a necessary assumption)"
11. A. "Moral perfectionism made possible by God" / B. "Moral improvement made possible by extraordinary capacity given by God to humanity as the 'final end' of creation/nature"

Anyone familiar with my writings on Kant's theory of religion<sup>1</sup> will share my perplexity in trying to understand how McGaughey could have possibly gleaned the 11 foregoing "A" claims from my work. A careful look at the first few pages of his article, however, provides some helpful insight here. McGaughey's main distinction is between historical religion and pure religion, and in the first few pages he defines this distinction in some detail, appealing to Kant's aforementioned concentric circles metaphor as an absolute, either/or choice. McGaughey apparently thinks Kant tells us that a religious person must choose *either* to accept the inner core of the concentric circles (i.e., pure rational/moral religion, the religion of reason) *or* to accept some historical religion (i.e., the outer circle, the part *outside* of the inner circle): McGaughey consistently writes as if one *cannot* accept both. He quotes Kant's metaphor of the hopeless attempt to mix oil and water (6:13) as evidence that the two circles can never have anything in common. If one imagines Kant's concentric circles imagery to be a large wheel, then McGaughey's foundational assumption is that the wheel's *hub* (the oil of pure religion) can have nothing at all in common with the wheel's outer *tire* (the water of historical religion). If this were indeed Kant's meaning, much of the rest of McGaughey's argument would follow, and much of my approach to interpreting Kant would indeed be deeply flawed.

But McGaughey's key assumption cannot be what Kant intended, for he misquotes the oil and water metaphor. Kant introduces this metaphor (6:13) as part of his discussion of how "between reason and Scripture" (i.e., between rational religion and revealed religion) we can find "not merely compatibility but also unity": Kant states that if one follows revealed religion "under the guidance of moral concepts", then one *will also be following* rational (moral) religion. He then appeals to the oil and water metaphor to explain what happens if one follows revealed religion *without* grounding it in such moral guidance. Revealed religion is like water only when it is released from the requirement of being moral: Kant's point is that if a religious person thinks revealed truths are *distinct from* and (especially) *superior to* moral truths, then pure (moral) religion and revealed religion cannot enjoy the aforementioned "unity"; instead, they amount to "one *religion* and one

---

<sup>1</sup> See especially my *Kant's Critical Religion* (Ashgate, 2000; Routledge, 2019) and *Comprehensive Commentary*. The latter, admittedly, had not yet been published when McGaughey wrote his 2013 article.

*ritual worship*", such that "the two would often have to be shaken together, in order to associate for a short time" yet will inevitably separate once again. McGaughey, by contrast, reads Kant's metaphor as meaning that any and all expressions of historical religion are necessarily "water" and thus fundamentally incompatible with the "oil" of pure religion--a claim Kant never makes, though reductionist interpreters routinely read it into the text at every turn.

McGaughey's misreading of Kant's oil-and-water metaphor is conveniently consistent with his aforementioned, wheel-like reading of Kant's concentric circles metaphor. Once we recognize that Kant's warning about oil and water refers only to historical religions that do *not* hold moral concepts at their core, we can see that Kant's reference to concentric circles is also an attempt to show how historical religion and revealed religion can be *unified*, not (as McGaughey reads it) an attempt to argue that moral religion can *never* be manifested in historical religion. McGaughey's position is extremely odd, for if he seriously maintains that historical religion and pure religion can never unite in one and the same person, then he is essentially condemning Kant's theory to be just that--an idealistic theory that, by definition, can never be put into practice--whether in the 21st century or in any other age. The textual evidence against such a claim (popular though it was when McGaughey first cut his philosophical teeth, prior to the appearance of affirmative interpretations) is so overwhelming that I cannot begin to present it all in a single article. Anyone who doubt this can find all the details spelled out in my *Comprehensive Commentary*, thus demonstrating that Kant's intended his theory of pure religion to make a real impact on the actual practice of historical religions of all types, around the world.

What exactly was it about my affirmative way of interpreting Kant's theory of religion that prompted McGaughey to respond to it in such an overwhelmingly negative, one-sided way? I may never know the answer to this question, though I have attempted to solve the mystery in private correspondence with him. Perhaps he read the 11 claims into my writings because the title of my first publication on this topic (Palmquist 1989)<sup>1</sup> misled him to think that my lifelong focus on the question of whether or not Kant's theory of religion is compatible with Christian belief and practice amounts to the claim that Kant *himself* thought of his theory as a defense of historical Christianity. (In that early article, I concluded that Kant's theory is consistent with Christianity, *if* properly

---

<sup>1</sup> See Stephen Palmquist, "Immanuel Kant: A Christian Philosopher?", *Faith and Philosophy* 6 (1989), 65-75. The origin of this title is noteworthy: based on a reviewer's recommendation, the editor required me to change my original title, which was "Can A Christian Be A Kantian?" As a young philosopher who felt the pressure of "publish or perish", I accepted the editor's requirement, especially because he did not require me to change the paper's content to fit the new title. The article never argues that Kant himself was a Christian; it argues only that Kant's theory of religion is not *incompatible* with Christian belief, such that Christians can also be Kantians without being forced to give up their Christian faith (or vice versa). Had I been able to foresee how often readers of my subsequent work have incorrectly assumed that I was making a claim about Kant's personal religious faith, I would have more courageously defended my original title as preferable.

interpreted; however, despite the misleading title that the journal's editor imposed on the article [see note 1 on p.15, above], I have never argued that Kant himself was, or thought of himself as, a Christian.) Or, perhaps McGaughey's motivation for writing such a vitriolic denunciation of my work, without even attempting to engage with its actual arguments, was because he knew I was then teaching at a university with "Baptist" in its name. If this was his reason, he was badly misinformed, since the only thing genuinely Baptist about the university in Hong Kong where I taught for 34 years is its heritage: prior to taking government funding in 1983, the institution was indeed funded and largely run by a group of American Baptists, working with the Baptist Convention of Hong Kong. But by the time I joined in 1987, it had already cut these religious ties and was on the path to being the thoroughly secular university it is today. Whatever his reason, nothing in "KEG" ever makes any of the 11 patently false claims that McGaughey attributes to me. While some interpreters (such as Firestone and Jacobs) *have* argued that Kant's primary goal in *Religion* was to promote historical Christianity, I have never defended such a claim--though I have claimed that interpreters *could* employ Kant's theory in this way.

In light of this detailed response to McGaughey's poorly argued article, and in a spirit of Kantian unity, I shall conclude §2 by responding to each of the 11 claims that McGaughey imputed to me (i.e., the "A" claims in the foregoing list) and to his preferred alternatives (i.e., the previously listed "B" claims), regarding the issue of how Kant intended his theory of pure (rational) religion to relate to historical religions. The following 11 numbered paragraphs correspond to the above list, with each item being labelled as claim "C" to indicate that each constitutes my attempt at a synthesis between McGaughey's "A" (his straw-man Palmquist) and "B" (his view of Kant's actual position).

1C. I fully agree with McGaughey's 1B. Perhaps he thought I would uphold 1A because 1A seems quite close to what Kant calls his "second experiment" (6:12), which I take very seriously. Kant says this second aspect of the book's aim will start with the fragments of a particular historical faith (i.e., Christianity) and examine the compatibility of each fragment with the corresponding element of pure religion. The fact that I have written a number of articles, including "KEG", that explore how this "second experiment" works, does not mean I ever claim that Kantian religion is *based on* any given historical religion.

2C. My *actual* argument in "KEG" is that Kant intends his discussion of divine grace, in subsection C of Section One of the Second Piece, to explain how a Christian *could* believe in the doctrine of prevenient grace *without* denying its rational grounding in "humanity's moral capacity". Claim 2B, that "Grace is speculative opinion," is not something Kant ever states; as such, I struggle to understand what McGaughey might mean here. In any case, Kant certainly never claims that belief in divine grace *necessarily* "undermines morality"; it only has this result if one believes in an tenet of historical faith in such a way that the belief supersedes or even negates all moral concerns. By no means do Christians necessarily interpret the doctrine in this



way. A main goal of Kant's theory of religious judgment was to urge Christians to avoid committing such a laziness-inducing error.

- 3C. Claims 3A and 3B are not contradictory but complement each other in Kant's theory of religion. As I will clarify in §4, Kantian religious judgment is (at least in part) anchored in freedom, as B requires, because genuine religion is anchored in morality and morality presupposes free choice. However, this does not change the fact that, when Kant discusses divine grace in *Religion*, he grounds it in the same noumenal gift (namely, the archetype of perfect humanity) that is the source of human freedom. "KEG" interprets Kant as explaining how Christians can therefore believe in their tradition's account of divine grace without abandoning pure rational religion, provided they retain this grounding in morality. McGaughey's misunderstanding of the concentric circles metaphor as requiring an either/or *choice* between historical and pure religion, has blinded him to this compatibility.
- 4C. "KEG" never states what McGaughey reports in claim 4A, that Christology is "exclusive", nor that it is solely grounded in a particular historical individual who is "external" to religious believers. Rather, I argue that Kant's account of the archetype (which uses explicitly Christian language and therefore surely counts as one of the "fragments" that he referred to as his "second experiment") seeks to explain how a person who *does* believe that Jesus Christ was a unique historical "prototype" of this archetype can do so without becoming morally lazy. McGaughey is therefore correct to insist in claim 4B that for Kant Christology is "inclusive," "internal," etc.
- 5C. "KEG" never makes claim 5A; it mentions imitation only once in passing, in reference to Kant's view that the person who experiences a change of heart will interpret suffering in a new way, a way that is similar to (and only in this *symbolic* sense "imitates") Jesus' attitude toward suffering. I fully agree with McGaughey's claim (5B) that Kant would regard rote imitation as fostering discouragement.
- 6C. "KEG" nowhere refers to Pascal's Wager, nor does it ever mention Kant's term "parerga". I do refer often to the role hope plays in Kant's theory, and I generally agree with the claim that increasing our awareness of our moral capacity is the rational ground for hope. But Kant also argues that we human beings, who do not have knowledge of God's noumenal perspective, must therefore *also* ground our hope in some empirical evidence that our moral improvement has progressed over time.
- 7C. McGaughey's claim 7A completely ignores the main argument in "KEG", that Kant's theory of grace has *both* a noumenal and a phenomenal side, both being crucial to his argument. In no way do I interpret Kant as claiming that the noumenal must be *substituted* for the phenomenal. Rather, I argue that, according to Kant, Christians who believe in divine grace (the noumenal) must base their personal assessment of their own condition on phenomenal evidence of their

moral improvement. McGaughey's preferred claim (7B) seems relevant to Kant's moral philosophy, but it is not something Kant explicitly argues in *Religion*.

8C. Neither in "KEG" nor in any of my other publications have I ever defended claim 8A. Rather, I interpret Kant as arguing that moral improvement is a matter of concern *both* to individuals and to the whole human community. I thus agree that claim 8B accurately describes (part of) Kant's position.

9C. A reader of "KEG" might impute claim 9A to me, because of the way I interpret the phenomenal/historical side of belief in divine grace. But again, my interpretation is two-eyed (perspectival), not one-sided in the way McGaughey's wording of claim 9A implies. I agree that claim 9B describes how Kant thinks we should measure morality; the phenomenal side of my argument in "KEG" does not claim that we are to measure *morality* by external consequences; rather, I read Kant as claiming that those who believe in divine grace (something human beings can never know for certain) should assess whether God's hypothetical action applies in their individual case through a process of internal self-examination, attempting to discern a tendency to improve (morally) over time. Kant's pragmatic claim cannot mean he is defending "consequentialism", because the latter is a theory of *moral* judgment, not of religious judgment.

10C. McGaughey's two claims here are so sketchy that I am unsure of what he intends by the word "Alternative". In any case, what is clear from Kant's text (and from a fair reading of "KEG") is that Kant opposes literal anthropomorphism but approves of symbolic anthropomorphism. In the prefaces to *Religion*, he also explicitly states that religion and philosophy should work together, in "unity". According to my best guess of McGaughey's intended meaning, Kant's (and my) reading could therefore be expressed accurately by 10B.

11C. My position in "KEG" and elsewhere is that God and human beings must work together as partners toward the cause of moral improvement. God founds the invisible church (i.e., pure religion), whereas human beings are charged with the task of building visible churches (i.e., historical religions) (see 6:100-102). As such, McGaughey's 11B is certainly correct, while his imputation of 11A to me expresses only one side of my two-sided interpretation.

Once again, as far as I can tell, all of McGaughey's erroneous assumptions about my interpretation of Kant's theory of religion stem from his radical interpretation of Kant's concentric circles metaphor, whereby historical religion and pure religion can, by definition, never be expressions of one and the same truth. If McGaughey were correct on this point, the whole notion of an "experiment," whose purpose is to explore the possibility of finding *unity* between the historical and the pure, would be a failure before it starts. However, now that I have clarified in some detail what I actually argue, in response to each of McGaughey's imputed claims, we can turn to a more explicit way of challenging my focus on "two experiments" as the hermeneutic key to interpreting Kant's *Religion*.

### 3. Pasternack's attempt to dismantle the "two experiments" hermeneutic

When I was writing my doctoral thesis in Oxford during the mid-1980s, I included several chapters setting out a new way of interpreting the theory of religion Kant defends in *Religion*. One of the several key hermeneutic principles I proposed to guide one's reading of Kant's (then) much-neglected book was to focus attention on defining and locating the "experiment(s)" (*Versuch[en]*) that, as Kant appears to say in the Preface to the second edition of *Religion* (6:12), he will be conducting throughout the book. A 1992 *Kant-Studien* article, which I later revised as Chapter VI of *Kant's Critical Religion* (2000/2019), is where I first defined Kant's two experiments as a key hermeneutic principle and proposed tracing Kant's alternating attention to each experiment throughout *Religion*. While a few earlier interpreters had mentioned Kant's stated intention to conduct his "experiment(s)" in the book, I know of no previous scholar who made Kant's statements in the prefaces, interpreted in terms of his appeal to the metaphor of concentric circles, as a "map" to guide one's reading of the entire text.

A few subsequent scholars, most notably John Hare in his 1996 *The Moral Gap*, also highlighted the importance of Kant's reference to the two experiments/circles, yet none focused on this theme as resolutely as I had done, until Chris Firestone and Nathan Jacobs published their *In Defense of Kant's Religion*, in 2008. Contrasting several possible "rubrics" that scholars had previously used to interpret Kant's text, they proposed an altogether novel way of defining and locating the two experiments--but without ever mentioning my thoroughgoing treatment of the same guiding metaphor in *Kant's Critical Religion*. Six years later, Lawrence Pasternack followed with a commentary that also made numerous references to Kant's metaphor, but tended to write as if Kant was really employing only one experiment throughout *Religion*. I have previously engaged with the Firestone/Jacobs interpretation at some length, (Palmquist 2012)<sup>1</sup> so here I will respond for the first time to Pasternack's new approach.

My focus in this section will be on Pasternack's most mature work on the two-experiments hermeneutic: his 2017 *Kantian Review* article offers an overview of the various options that had been proposed up to that point (including the approach in his own 2014 commentary on *Religion*<sup>9</sup>), (Pasternack 2014)<sup>2</sup> then proposes a new argument that attempts to dismiss the whole exercise as "an utterly banal misunderstanding" (122), little more than a waste of everyone's time (108). Before I summarize and respond to his new argument, let me share some details regarding the historical context in which Pasternack was prompted to write the 2017 article. Before doing so, let me add that from 2012 until his untimely death in 2024 Pasternack sent me early drafts of most of his

<sup>1</sup> See "Cross-Examination of In Defense of Kant's Religion", *Faith and Philosophy* 29 (2012), 170-180, and "To Tell the Truth on Kant and Christianity: Will the Real Affirmative Interpreter Please Stand Up?", *Faith and Philosophy* 29 (2012), 340-346.

<sup>2</sup> Lawrence Pasternack, *Routledge Philosophy Guidebook to Immanuel Kant's Religion within the Boundaries of Mere Reason: An Interpretation and Defense* (London: Routledge, 2014).

publications, and I gave detailed feedback on many of them, including his 2014 book and the 2017 article that is the main topic of this section. As such, I am painfully aware of the sad fact that Pasternack will be unable to publish a rebuttal or other clarification in response to my comments here. I shall therefore aim to be as fair-minded and accurate as possible, sticking only to known facts and obvious implications.

Not long after he started work on his doctorate in Hong Kong, my former student Brandon Love wrote what was undoubtedly the best paper a graduate student has ever submitted to me: in it, he proposed an ingenious way of resolving the debate that had at that point been raging for only a few years, regarding the nature and location of Kant's two experiments in *Religion*. (Firestone supervised Love's graduate work for his master's degree, so Love's interest in this debate was personal as well as professional: honorably, he hoped to make peace between his two beloved supervisors.) At my encouragement, Love submitted his paper to a prominent Kant journal in 2014. In it, he proposed an ingenious resolution of the dispute, showing how all past interpretations are correct, each in its own way, once we notice that Kant uses the term "natural religion" in two distinct ways (6:154-6): as referring *either* to the way practical reason *justifies* religion in general (or any specific religious belief or practice) by grounding religion in morality, *or* to the way religious beliefs and practices (whether pure or historical) are *spread* in history. (Note that this is parallel but not identical to Kant's distinction between pure and historical religion, which McGaughey 2013 interpreted in such a misleading way.) In the first case, Kant contrasts natural religion with *revealed* religion, as two ways of understanding the *basic nature* of religion, while in the second, he contrasts it with *scholarly* religion, as two ways of conceiving how religion is best *propagated*. In a nutshell, Love argues that my way of distinguishing between the two experiments focuses on the first sense of "natural religion", while the Firestone-Jacobs reading focuses on the second: Kant really *does* move from natural religion (in the first sense) to *revealed* religion in each of *Religion's* four pieces, as I have demonstrated elsewhere; yet he also really *does* focus only on natural religion (in the *second* sense) throughout the first three pieces, while turning to *scholarly* religion only in the Fourth Piece. Love's paper adds that Hare's translation thesis (as interpreted in Firestone-Jacobs 2008) also correctly conveys certain aspects of the relation between the second *Critique* and *Religion*. Of course, much more could be said about Love's claims; but let it suffice to say that, as a stellar example of good philosophy, his insight has the potential to resolve the debate once and for all.

Unfortunately, the journal did not accept Love's paper for publication: their reviewer composed an extraordinarily long response, arguing that there was no need to resolve this dispute because Love's whole valiant attempt to make peace between the interlocutors was simply a waste of time. The reviewer's stated reasons for this reaching this conclusion closely matched the arguments later presented in the last two sections of Pasternack 2017, which ended up being published in the same journal. In 2023, Love presented a much-revised version of his 2014 paper, one that incorporates

an equally empathetic response to Pasternack 2017, at a Notre Dame colloquium on Kant's affirmative approach to religion. As all the iterations of Love's paper remain unpublished today, I will not risk rendering its publication irrelevant by saying anything further about the details of his brilliant argument. Instead, the remainder of §3 will be a response to Pasternack 2017.

The first 70% of Pasternack's article is devoted to a generally helpful overview of the various ways affirmative interpreters have explained Kant's references to the two "experiments" that he intends to conduct in the main text of *Religion*. (Love's unpublished 2014 paper also included such an overview, though of course his conclusions were very different from Pasternack's. Pasternack's summaries say nothing about the new solution suggested in Love's 2014 paper, because that position had of course not been published. Given that detailed summaries of the various positions are available elsewhere, I see no need to summarize the options here. Any reader who is unfamiliar with this debate can consult §§1-2 of Pasternack 2017.) Perhaps the main weakness of his summary of the debate is that he borrows the potentially misleading "Religion as Translation" versus "Religion as Symbol" distinction that Firestone-Jacobs 2008 had introduced. Pasternack also employs two additional (but equally misleading) rubrics: "Religion as Apologetics" (referring to the new approach defended by Firestone-Jacobs 2008) and "Religion as Vehicle" (referring to the approach in Pasternack 2014).

These rubrics are all misleading because they use the word "religion" in a thoroughly unKantian way. Kant himself is careful to reserve "religion" for the forms of religious judgment that *properly* prioritize the relationship between moral precepts and the beliefs, symbols, and rituals of a particular historical tradition. The more neutral word Kant uses for historical religious traditions, quite apart from any consideration of their moral content, is "faith". Thus, "rational faith" refers to the beliefs, symbols, and rituals that are (as it were) "commanded" by reason itself, whereas "historical faiths" are the systems of beliefs, symbols, and rituals to which people adhere at various times and places down through human history, many (if not most) of which are not explicitly moral. By contrast, Kant modifies his technical use of "religion" with words like "pure", "moral", or "rational" when it refers to his philosophical theory of the elements that must be exhibited by any *genuine* (i.e., rationally grounded) faith; likewise, the term "historical religion" refers to any historical faith that has properly preserved the elements of rational/pure religion at its core. Given Kant's official definitions of these technical terms, Firestone-Jacobs 2008 and Pasternack 2017 should have used the rubric "*Faith* as x" in order to distinguish between the various ways of treating Kant's two-experiments language as the primary hermeneutic tool for interpreting the text of *Religion*. This may seem like nit-picking, but it has important implications for how one assesses the relevance of Kant's theory to the way religion is actually practiced during any given historical period.

When this correction is made, the fancy labels for each hermeneutic approach align more closely with Kant's own usage. In *Kant's Critical Religion* and other early publications, for example, I did

indeed argue that the elements of any given *historical faith* (not just Christianity) can and should serve as symbols of pure moral religion. Likewise, Pasternack 2014 argues that Kant portrays *historical faiths* (not pure religion!) as "vehicles" for moral religion. (The label "*Religion* as Vehicle" thus substitutes the proverbial philosophical "horse" for the historical "cart"!) Perhaps most importantly, Pasternack's extreme impatience and premature dismissal of the arguments advanced by Firestone-Jacobs 2008 could have been significantly tempered, had he recognized that they are not arguing that *rational religion* is "apologetic", but rather that the Christian *faith* stands (or, they argue, should stand) in an apologetic relationship with pure moral religion. The exception to my suggested terminological correction of these labels is the label they give Hare's hermeneutic assumption: if we choose to read Hare 1996 as Firestone/Jacobs 2008 and Pasternack 2017 do, as adopting a "Religion as translation" hermeneutic, then in this case alone, the label is best read as arguing that Kant's *Religion book* is a project of translating the tenets of a particular *faith* (namely Christianity) into the terms of rational religion.

The last main section of Pasternack 2017 offers a careful exegesis of the sentences in the two prefaces of *Religion* where Kant refers to the "experiment" (6:10) and/or "second experiment" that guide(s) Kant's writing of the book. In a nutshell, Pasternack reads Kant as using "*Versuch*" in a non-technical, "banal" sense, whereby it refers simply to the "essay" (one of the possible meanings of Kant's German word) that Kant presents to the public, along with the Preface. Pasternack claims there is no textual evidence to support the supposition that Kant ever had any "first experiment" in mind. Instead, he hypothesizes, the "experiment [*Versuch*]" Kant mentions in the first Preface (6:10) is *identical* to the "second experiment [*Versuch*]" he mentions in the second Preface (6:12);<sup>1</sup> the only difference is that in the 1794 edition Kant is telling the reader that he is now making a *second attempt* to put forward his argument regarding the nature of religion. In other words, the whole debate over the nature and location of Kant's alleged "two experiments" is all based on a sad translation mistake: Kant's use of *Versuch* at 6:12 refers simply to the "second iteration" of the book's singular project (127). Although Pasternack amasses an impressive array of evidence to support his new reading--most notably, he is undoubtedly correct that the word *Versuch* can and did often refer to an "essay" in Kant's day, so it is not impossible that Kant might have intended to use it in this way--his argument remains far from conclusive, for the following reasons.

First, although Pasternack acknowledges that Kant's reference to a "second experiment" occurs in the same paragraph as his concentric circles metaphor, he does not seem to recognize that Kant employs both metaphors for the same expressed purpose: explaining the nature of the book's key task(s). What Pasternack 2017 never considers is that those who use the "two experiments"

---

<sup>1</sup> As such, the conclusion of Pasternack 2017 is consistent with the claim that Kant intended to propose a single task (or experiment) for *Religion*, but one that takes place in two distinct steps. Pasternack's own tendency in his 2014 commentary was to use the singular term "experiment", in much the same way.

language are referring to something that is far more deeply embedded in the text of *Religion* than merely what we find in Kant's use of the term "Versuch". Aside from one brief reference in an early quote from Hare, Pasternack mentions Kant's use of the concentric circles metaphor in only one passage (p.122). He is correct that Kant employs this metaphor (and his infamous reference to the "second experiment") in the context of a clarification of the meaning of his book's title. What Pasternack leaves unstated is that the simple measure of changing "experiment" to "attempt" in our English translation of Kant's words, or even the more radical measure of regarding the use of "*Versuch*" in both prefaces to be references to one and the same thing, does nothing to change the fact that Kant's concentric circles metaphor itself beckons the reader to perform a *two-step* task: first, distinguish between historical and pure religion, then assess the extent to which fragments of some specific historical faith contain within them the kernel of moral truth that reason teaches. In other words, Pasternack's attempt to silence all future discussion of Kant's two experiments is not only futile, but potentially detrimental to the goal that Pasternack himself sets forth: a fuller understanding of what Kant intended to accomplish in the main text of *Religion*.

Even if Pasternack's argument were correct and Kant does not intentionally distinguish between two distinct "experiments" that he plans to conduct in the text, this fact would not free interpreters from the need to understand what the concentric circles are all about. From the very beginning of my work on interpreting Kant's *Religion* I have regarded any talk of the two experiments to be integrally bound up in an understanding of the concentric circles metaphor. So, in fact, nothing crucial would change in my argument, if I were to grant Pasternack's new argument. All that would change is that interpreters would no longer be able to talk in terms of experiments. We would still be duty-bound to answer two questions: (1) What is the difference between the nature of the inner and outer circles? and (2) Precisely how and to what extent does the outer circle relate to the inner circle? These just *are* the two experiments, as I understand them. Interpreters frequently introduce new terminology in order to clarify the arguments and theories they believe a great thinker has put forward. I would not protest if, in the wake of Pasternack 2017, future interpreters begin to describe the continued use of a two-experiments hermeneutic as "Palmquist's" model for interpreting *Religion*. However, I arrived at the name for this model because the translators used the word "experiment" to translate Kant's *Versuch*, so I still continue to maintain that the term I use to describe the tasks defined by the two concentric circles are Kant's, not mine. In other words, the first problem with Pasternack's new approach is that it doesn't really solve any problem; as he himself ironically admits at the end of the article, everything remains as it is, and interpreters are still left with the task of figuring out *what* tasks Kant intends to carry out and *where* in his text he accomplishes whatever the task(s) may be.

The second drawback of Pasternack 2017 is that it remains completely silent about the very significant arguments put forward in the aforementioned paper by Brandon Love, even though that paper predated (and, according to my observation as a personal friend of both scholars, was *the*

motivating factor prompting) Pasternack's decision to write an article on this topic. (Again, the anonymous review of Love's paper expresses many of the same points defended in Pasternack 2017, and in the several discussions of Love's paper that I had with Pasternack, he seemed to have a prior familiarity with its claims.) Of course, his silence is in a sense justified, because Love to this day has not published any of the several versions of his paper, so Pasternack was under no scholarly obligation to cite it. However, I believe Love's paper (and later, his doctoral dissertation<sup>1</sup>) demonstrates beyond reasonable doubt, through an extensive analysis of a tradition in early chemistry that had a deep influence on Kant's thinking, that in the earlier decades of the eighteenth century chemists used just such "experiment" terminology to describe the two-step process they employed to conduct experiments on chemical compounds: in short, chemists first *separate* a compound into its constituent *elements* (a process sometimes called "analysis"), then attempt to reunite them (a process sometimes called "synthesis"). Kant uses exactly the same terminology elsewhere in his writing, most notably in the prefaces to the first *Critique*, and he knew about the chemists who made such distinctions, because he referred to them from time to time. This makes it far more likely than Pasternack admits for Kant to think in terms of a pair of experiments when he appeals in the first Preface to the need to *distinguish* biblical theology from philosophical theology and when he appeals in the second Preface to a second step of *uniting* these. As a close friend of both Love and Pasternack during the years when they were writing their respective essays, it saddens me that up to now, the less profound work of the senior scholar has eclipsed the more profound work of the junior scholar--a fate Pasternack himself had to endure at certain stages of his career. Once Love's paper is eventually published, all Kant scholars will be able to see for themselves that there is much more to Kant's use of *Versuch* than Pasternack was willing to admit.

#### 4. Conclusion: What can Kant do for religion in the 21<sup>st</sup> century?

In a lengthy footnote, after summarizing some of the more influential approaches to interpreting Kant's theory of religion by several prominent German interpreters, McGaughey 2013 (156-7n) quotes approvingly from a 2007 article by Andreas Urs that claims, in commenting on *Kant and the New Philosophy of Religion*<sup>2</sup> (40), that proponents of "the current 'affirmative' interpretation of Kant's philosophy of religion in the English speaking world...are more ready to fill in the ditch between Christian theology and Kant's philosophy of religion in contrast to what is done in similar German publications. As a consequence, Kant's critical approach loses its bite." If McGaughey's

---

<sup>1</sup>Brandon Love, *Kant's Baconian Method as a Transformation of Aristotelian Transcendental Philosophy—A Propaedeutic* (Hong Kong Baptist University doctoral dissertation, 2018). <https://scholars.hkbu.edu.hk/en/studentTheses/kants-baconian-method-as-a-transformation-of-aristotelian-transce>

<sup>2</sup>Chris L Firestone and Stephen Palmquist, *Kant and the New Philosophy of Religion* (Bloomington: Indiana University Press, 2006). McGaughey 2013 (157n) cites this review article by Andreas Urs as "*Neuerscheinungen zu Kants Religionsphilosophie*," *Philosophische Rundschau* 54 (2007), 31-53



rough-and-ready parody of my position were even close to being accurate, then Urs' claim might have some merit. However, Urs' assessment proves to be off target for any affirmative interpreter who portrays Kant as attempting to revolutionize *historical religion* not by blurring the difference between the pure and the historical (as McGaughey misconstrues my interpretation as doing) but by requiring the historical to *conform itself* to the pure. As such, which type of interpretation actually causes Kant's theory of religion to "lose its bite"? Is it not those such as McGaughey's, which end up converting Lessing's "ugly ditch" into a chasm that is forever and always unbridgeable, such that Kant's pure religion can *never* become relevant to *anything* historical? As anyone who has ever actually been bitten is sure to testify, the *pain* of being bitten is *always* an empirically real *experience*, not merely an abstract transcendental condition! How could *any* philosophical theory have "bite" in the empirical world if it defines itself into a corner by claiming that it can *never* have any resonance with anything historical?

In §IV of Division One of the Third Piece, Kant sets out four guidelines for the constitution of a "true church" (6:101-2): universality, integrity (*Lauterkeit*), freedom, and unchangeableness (of these four guidelines). Just before introducing these guidelines, he distinguishes between the "invisible church", of which only God could serve as the founder, and the "visible church", whose founding can only be the task of us human beings (6:100-101). I read these four guidelines as Kant's explanation of how his theory of religious judgment will (or at least should) have "bite" in relation to any and all historical religious faiths. For, as I demonstrate in Part Three of my *Comprehensive Commentary*, immediately after this discussion of the four categorial guidelines for designing a constitution for a visible church, each of the next four main sections of *Religion's* Third Piece focuses on one of the four guidelines. This is no accident. Kant typically presents his theory of judgment in the third main part of the relevant work, so this is precisely where we should expect to find Kant introducing a set of "categories" to guide our judgment in matters of religious faith.

With this in mind, let us look briefly at the way Kant uses words with the "judg" root in the Third Piece (6:93-147). Division One starts out with a Hobbesian distinction between the "state of nature", wherein "each person is his own judge [*Richter*]" (6:92) and "each human being wants to be himself the judge [*Richter*] concerning what is to be a right for him" (97). The problem is that "this state is a continual wounding of the rights of all others through the presumption of one's being the judge [*Richter*] in one's own cause" (97) --a problem whose solution, Kant suggests, is to set up an "ethical state". But this solution immediately runs into a problem, because "a human judge [*Richter*] cannot fathom the inwardness of other human beings" (95) and this means that our ethical judgments about each other are bound to be inaccurate. This is why human beings inevitably end up appealing to "a divine judge [*Richter*]" (116), a rational postulation that allows us to cope with our inability to understand various perplexities of our ethical situation, such as why some human beings end up being good while others end up being evil: on such matters, "we must leave the judgment [*Urtheil*] to the All-seeing" (121). In other words, the twofold key to Kantian religious

judgment is, on the one hand, to interpret our relations with other people ethically, while on the other hand *deferring* judgment of others in humble recognition of our own ignorance, leaving the ultimate judgment to God, who alone knows the heart of each person.

A parenthetical word is in order here about the importance of hermeneutics to the scholarly task of interpreting the writings of great philosophers. Both McGaughey and Pasternack have accused me of imposing an "agenda" onto Kant's text that is not merited by the text itself. As we have seen, they level this accusation in different ways and with different aims in view, but they both agree that a good interpreter will resist the temptation to impose any such agenda onto a text and should instead just read the text for what it is. The problem, of course, as Gadamer made so clear in his monumental writings, is that it is virtually impossible to interpret *without* reading one's own assumptions into the text. The art and science of interpretation does not (or should not) succeed if we think of ourselves as doing pure, objective exegesis, because *everyone* who writes *anything* about another text must (if he or she has anything remotely interesting to say) *have an agenda that is to some extent foreign to the text*. Post-Gadamer, it should no longer be controversial to affirm that eisegesis is an essential part of good interpretation; what matters is not *whether or not one has an agenda*, but rather whether one's agenda (regarding what the text doesn't say) is *justified* by what the text *does say* and therefore provides *insight* into what the original author wrote and might have meant by what he or she wrote. McGaughey's and Pasternack's articles are themselves excellent illustrations of two serious errors that philosophers typically make when they critique other people's interpretations: Pasternack's error is to assume that an interpretive agenda cannot bring out *genuine* insights if it is not literally and explicitly based solely on the original text; McGaughey's error is to assume that an interpreter who *seems* to be disagreeing with you thereby has nothing valuable to teach you, but should instead just be scoffed at and dismissed as a ridiculous buffoon.

I hope the foregoing article practices what it preaches by responding to McGaughey 2013 and Pasternack 2017 without committing these twin hermeneutic errors that both result in prematurely dismissing one's interlocutors' insights. That is, even though McGaughey's list of 11 errors initially struck me as total fabrications that do not reflect anything I actually wrote, I should read them as evidence that their author was bringing an (unspoken) agenda to the text of my article, and that if I can identify and empathize with that agenda, then I can *tweak* each claim in such a way that the revised list becomes instructive. By reading our opponent's work with humility, rather than with the knee-jerk assumption that "you disagree, so you're just wrong!", we can actually learn something about how our own work was not as clear as we thought it was. By appropriately tweaking such seemingly erroneous claims, *new* insights can be gleaned, and as I illustrated in §2, I was able to reassess and, in some cases, tweak the items in McGaughey's *right-hand column*, which (when I first read his article) seemed mostly correct. In other words, one of my goals in this

article has been to illustrate how a good interpretation draws insight from one's interlocutor and uses the disagreement to move to a higher level of understanding.

Where the rubber hits the road, when it comes to the influence Kant's theory of religious judgment could have on religious belief and practice in the last three-quarters of the twenty-first century, is in his analysis of the proper relationship between what he calls (as we have already seen) pure/moral/rational religion and specific historical faiths. This is where his aforementioned four guidelines for religious judgment can have significant "bite": whereas pure religion must be universal, historical faiths are inevitably tied to specific traditions; whereas pure religion focuses on a person's moral integrity, historical faiths typically focus on currying favor through divine worship; whereas pure religion is grounded in freedom (both of individuals, between each other, and of the religious community in its relation to the state), historical faiths are usually based on a hierarchical relationship between the clergy and the laity and are often unduly concerned with gaining and maintaining political power in the state; and whereas pure religion holds these basic forms of judgment to be categorically necessary (unchangeable), historical faiths properly consist of various empirical forms that are best seen as evolving with the changing times yet which can *seem* immutable by their adherents.

Unlike McGaughey, who writes as if each of these "whereas" comparisons entails that pure religion must keep itself distanced from any and all historical faiths, a close look at Kant's references to judgment in the Third Piece (and throughout *Religion*) reveal quite a different story. While I do not have space here for a detailed analysis, suffice it to say that many of Kant's references to judges, judging, and judgment in the Third Piece are concerned with identifying a way of understanding the Christian doctrine of the Trinity in such a way that it can be *consistent with* the principles of pure religion. Suffice it to say that Kant interprets the different persons of the Trinity as different aspects of *moral judging*. Empowered by belief in such a moral Trinity, a genuinely religious person (6:145) "must constantly test himself as one summoned to give account before a judge [*Richter*]." This is "bite" if ever there was any.

Why should believers in any historical faith care about conforming their beliefs and practices to Kantian pure religion? One good reason, Kant says, is that this enables the historical faith to be spread more widely to a broader range of people. As he puts it in 6:103: "a historical faith based merely on facts can spread its influence no farther than however far the reports relating to the capacity of judging [*zu beurtheilen*] their credibility can reach out to, according to circumstances of time and place." By contrast, a historical *religion* (i.e., a faith that embraces the principles of pure religion at its core) will have *universal* appeal.

A more detailed analysis of Kant's theory of religious judgment would reveal that it, too, entails a two-step process that parallels the way early chemistry experiments were conducted, and this same two-step process lies at the basis of Kant's (or if we follow Pasternack, some interpreters') suggestion that two experiments (or perhaps a single, twofold experiment) operate(s) throughout

*Religion*. As such, I fully agree with the final conclusion of Pasternack 2017, that what really matters is what the main text says about the nature and function of religion. What it says is that religious people *must*, if they are to avoid falling into a false form of religious belief and practice, attempt to accomplish two rational aims: first, distinguish the elements of your religious tradition that are grounded in moral reason from those that are not so grounded; and second, assess how best to interpret the latter elements in such a way that they can stand in *unity* with the former. This is precisely what I had in mind when, over 35 years ago, I first interpreted the second Preface to *Religion* as urging readers to be on the lookout for the two experiments that appear throughout the main text. To that claim I would now add a prediction: a thorough study of Kant's use of terms relating to judgment in *Religion* would confirm that religious judgment for Kant must follow a similar two-step (or two-sided) process. The fate of religion in the remainder of the 21st century will depend, to a large extent, on the extent to which religious believers and practitioners are willing to follow Kant's lead as they consider how to make real, historical religious judgments

## References

- Brandon, L. (2018). Kant's Baconian Method as a Transformation of Aristotelian Transcendental Philosophy—A Propaedeutic, *Hong Kong Baptist University doctoral dissertation*. <https://scholars.hkbu.edu.hk/en/studentTheses/kants-baconian-method-as-a-transformation-of-aristotelian-transce>
- Firestone, C. L. & Palmquist, S. (2006). *Kant and the New Philosophy of Religion*, Indiana University Press.
- McGaughey, D. (2013). Historical and Pure Religion: A Response to Stephen Palmquist, *The Journal of Religion*, 93 (2), 151-176. <http://dx.doi.org/10.1086/669206>
- Palmquist, S. R. (1989). Immanuel Kant: A Christian Philosopher? *Faith and Philosophy*, 6 (1), 65-75, <http://dx.doi.org/10.5840/faithphil1989619>
- Palmquist, S. R. (2012). Cross-Examination of In Defense of Kant's Religion', *Faith and Philosophy*, 29 (2), 170-180, <http://dx.doi.org/10.5840/faithphil201229217>
- Palmquist, S. R. (2012). To Tell the Truth on Kant and Christianity: Will the Real Affirmative Interpreter Please Stand Up?", *Faith and Philosophy*, 29 (3), 340-346. <http://dx.doi.org/10.5840/faithphil201229334>
- Palmquist, S. R. (2016). *Comprehensive Commentary on Kant's Religion Within the Bounds of Bare Reason*, Wiley-Blackwell
- Palmquist, S. R. (2018). *Kant's Critical Religion: Volume Two of Kant's System of Perspectives*, Routledge,
- Palmquist, S. (2010). Kant's Ethics of Grace: Perspectival Solutions to the Moral Difficulties with Divine Grace, *The Journal of Religion*, 90 (4), 530-53. <http://dx.doi.org/10.1086/654821>
- Pasternack, L. (2014). *Routledge Philosophy Guidebook to Immanuel Kant's Religion within the Boundaries of Mere Reason: An Interpretation and Defense*, Routledge.
- Pasternack, L. (2017). The 'Two Experiments' of Kant's *Religion*: Dismantling the Conundrum, *Kantian Review*, 22 (1), 107,131. <http://dx.doi.org/10.1017/S136941541600039X>



Journal of Philosophical Investigations

## Journal of Philosophical Investigations

Print ISSN: 2251-7960 Online ISSN: 2423-4419

Homepage: <https://philosophy.tabrizu.ac.ir>



University of Tabriz

# Kant's Humanism: A Loophole in the Principle of Sufficient Reason

Daniel Dal Monte 

Professor of Philosophy, Rowan University, United States. Email: [ddalmonete@lincoln.edu](mailto:ddalmonete@lincoln.edu)

### Article Info

#### Article type:

Research Article

#### Article history:

Received 01 June 2024

Received in revised form  
01 July 2024

Accepted 08 July 2024

Published online 07  
August 2024

#### Keywords:

Kant, free will, humanism,  
principle of sufficient  
reason

### ABSTRACT

I consider the principle of sufficient reason (henceforth, PSR) as it functions in both Leibniz and Kant. The issue separating these thinkers is a modal status of absolute contingency, which is exempt from PSR insofar as it is neither logically necessary, nor does it necessarily follow from the given causal series. Leibniz's ambitious metaphysics applies PSR even to God's choices, which, since they must rest on a reason that makes sense of them, necessarily tend to the creation of the best of all possible worlds. Through PSR, the exercise of human freedom represents the unfolding of a concept God already has chosen, with an eye to the best possible world aligned with the universal intelligibility enjoined by PSR. PSR, in Kant's critical period, is not a principle of being, but one of mere experience, since any extension of thought beyond possible experience can yield no knowledge. Human agency, for Kant, has an intelligible aspect that is beyond possible experience. Since PSR is only a principle of experience for Kant, the agent in its intelligible aspect is not subject to it. Human free will introduces a special modal category of absolute contingency. Kant provides impetus for a humanism that makes the absolute freedom of the human will a competitor with the sovereignty of God, and also liberates the human will from contemporary ideologies that would subordinate it to natural determinism or group dynamics.

**Cite this article:** Dal Monte, D. (2024). Kant's Humanism: A Loophole in the Principle of Sufficient Reason. *Journal of Philosophical Investigations*, 18(47), 29-48. <https://doi.org/10.22034/jpiut.2024.62265.3800>



© The Author(s).

<https://doi.org/10.22034/jpiut.2024.62265.3800>

Publisher: University of Tabriz.

## Introduction

I establish how Kant carves out a special modal category of absolute contingency, that is free from determination by both natural causality as well as God's creative activity. This space of absolute contingency is intolerable within the metaphysics of Leibniz, in which the principle of sufficient reason (PSR)<sup>1</sup> orients all of reality towards the best of all possible worlds (Leibniz 1965, 53). My stipulative definition for *absolute* contingency is a modal status that might be false in another possible world, so that what is absolutely contingent is not logically necessary. Moreover, what is absolutely contingent is underdetermined by the prior causal sequence in the actual world. We can only account for what is absolutely contingent, in virtue of its actuality, not its logical necessity that makes its contrary a contradiction, nor in terms of a deterministic sequence leading up to it. We find the modal category of absolute contingency in Kant's contemporary, Crusius, who maintained that we can account for certain events in terms of their actuality, without reference to any antecedent ground (Kant, 1996, 1:397). Kant rejects absolute contingency in the pre-critical period, maintaining instead that "nothing which exists contingently can be without a ground which determines its existence antecedently" (Kant, 1996, 1:396). Yet, in the critical period, Kant endorses the notion of Crusius, carving out a space for free will that is independent of determinism of natural causes, and of the sovereignty of God (Kant, 1998, A540/B568).

Kant's liberation of free will from antecedent causes is significant in the 21<sup>st</sup> century, as the impetus for a new humanism. Theological determinism, which subordinates human free will to the pre-determination of God, is still a viable ideology, but even more so is a denial of human agency emerging in the natural sciences. Natural sciences seek to account for human choices in terms of antecedent natural processes, and so deny the absolute contingency of the human will (Wegner, 2002). Contemporary political ideologies view the individual as a product of group identity and economic and political structures, and so presuppose an understanding of society that is irreducible to individuals capable of independent action (Durkheim, 1982).

I contrast absolute contingency with hypothetical contingency, in a way that parallels Leibniz's distinction between absolute and hypothetical necessity (Leibniz, 1988). Absolute necessity entails existence in any possible world, since its non-existence is contradictory. Entities that are hypothetically necessary might not exist in another possible world, because their contraries do not imply a contradiction. Hypothetically necessary entities must exist, given the causal sequence in the actual world. I stipulate, as a parallel to hypothetical necessity, a hypothetical contingency that may not exist in other possible worlds, but must exist given the causal structure of the actual world. Hypothetical contingency is only contingent *on the condition that* we are not in the actual world with its causal sequence. Absolute contingency is contingent in *all* worlds, actual and possible, and

---

<sup>1</sup> *Discourse on Metaphysics*, Section 30. Leibniz, G.W. (1988) *Discourse on Metaphysics and Related Writings*. Edited and translated by R. Niall D. Martin and Stuart Brown. Manchester: Manchester University Press.

so even the actual sequence of causes in the world underdetermines that which is absolutely contingent.

The thesis of this paper is that Kant, in his critical system, carves out a space of absolute contingency that confers on the human will an absolute sovereignty over its actions that is competitive both with nature and with God. This space of absolute contingency would be anathema to Leibniz. “For it is necessary to analyze everything into some reason, and not to stop until we arrive at a first reason—or else it must be admitted that something can exist without a sufficient reason for its existence, and this admission destroys the demonstration of the existence of God and of many philosophical theorems.”<sup>1</sup> Leibniz is a precursor to contemporary thinkers, in both the natural and human sciences, that seek to explain human agency, not in terms of free choices, but in terms of external forces. In conferring on the human will an absolute sovereignty over its actions, independent of both natural and theological determinism, Kant introduces a new humanism that makes the human agent a locus of independent activity. For Leibniz, the human agent simply follows its individual concept pre-installed by God (Leibniz, 1988), participating in a universal harmony that God pre-established. “...Every single substance is a perpetual living mirror of the universe.” (Leibniz, 1965, 56)

In the first section, I will show how PSR in Leibniz serves to rebut both Spinozism and Cartesianism, insofar as both of the latter involve a contingency that threatens universal intelligibility according to God’s choice of the best of all possible worlds. I conduct this inquiry to show how Leibniz uses PSR to close any loopholes of incomprehensibility in his system. This aspiration to universal intelligibility is a precursor to contemporary philosophies that prefer a holistic order that strips human agency of any ability to stray from it. Human agency must not compromise the explicability and predictability of reality in natural or social terms.

In section two, I show how Leibniz subjects the human will to the order of things expressing PSR. Each choice follows from the complete concept of the individual already contained in the mind of God (Adams, 1994,12).

In the third section, I shift to Kant’s system, with the ultimate aim of showing how he establishes a modal space of absolute contingency. I begin by focusing on Kant’s modesty in metaphysics (de Jung 2013, 553), in contrast to the ambitious metaphysics of Leibniz in which there is a universal ontological order according to PSR. The analytic of the understanding, i.e. the articulation of the formal structure of possible experience, replaces the once proud discipline of ontology, the science of being in itself (Kant 1998, A247/B304). PSR becomes a principle establishing a causal law in experience rather than absolute being that, moreover, is not a teleological orientation towards the best of all possible worlds. Teleology shifts, in Kant, from an ontological principle organizing all things according to what is best, to a mere principle of the faculty of judgment that organizes cognition, not nature itself (Kant 2000, 20:236). In Kant’s critical philosophy, the unconditioned,

---

<sup>1</sup> From a letter to legal scholar Magnus Wedderkopf, cited in Adams (1994): 10.

which is supposed to provide ultimate satisfaction for PSR, permanently escapes our capacity for knowledge (Kant, 1998, A481/B509). Kant's epistemological modesty breaks down the architectonic of Leibniz in which all of reality finds its root in the will of God, and, by implication, challenges the contemporary ideologies that would explain human society in terms of natural causes or sociological forces.

In the fourth section, I will focus on how the modest metaphysics of Kant's critical period carves out a space for absolute contingency with respect to human free will. Kant achieves this special modal space through the distinction between the empirical and the intelligible characters. The intelligible character is not subject to the causal law characterizing experience (Kant, 1998, A540/B568). Since the intelligible character could choose differently than it does in other possible worlds, and since it is not subject to determination by antecedent causes, the intelligible character attains the special modal status of absolute contingency. The free will of the intelligible character in Kant represents a loophole to PSR.

At stake in this discussion of absolute contingency is the intelligibility and order of the world. Does our free will undermine, not only the natural order subject to law and so intelligible to science, but the sovereignty of God in virtue of which he has control over all events? Carving out a space for absolute contingency is a coup for humanism, since it imparts to human free will an ability to rise above the natural forces converging upon it, as well as any extension of the divine sovereignty over human actions. Kant, in *KU*, identifies the human being as the *final end* of nature, which requires no further condition to justify it (Kant, 2000, 20:432). Humanity is the final end in nature insofar as it is able, through its intelligible character, to exercise absolute freedom above any natural causes, in a supersensible way, according to an unconditional moral law that binds us regardless of our natural circumstances. "His existence contains the highest end in itself, to which...he can subject the whole of nature, or against which at least he need not hold himself to be subjected by any influence from nature..." (Kant 2000, 20:436). This humanism exalts the human individual above contemporary views that reduce human agency to expressions of natural causes, or to group dynamics that take on their own reality irreducible to individual decisions.

### **1. PSR in Leibniz**

PSR's centrality to the metaphysical order in Leibniz prevents the realization of a world that "flouts and baffles the reason that is in man" (Lovejoy, 1936, 168). Without the universal application of PSR, there would be events in the world that lack a full explanation. Any expectations we might form, based on observed regularities in the world, would be subject to doubt, because anything can happen at any time, without reason.

Leibniz uses PSR to reject two philosophical systems, Spinozism and Cartesianism. I show Leibniz's rejection of these systems in order to show the predominance of PSR in his system, which accounts for everything in terms of God's choice of the best of all possible worlds (Leibniz 1965,55). First, Spinozism adopts a full necessitarianism that invokes PSR, but in a reductive way.



Spinoza maintains that “for each thing there must be a cause, or reason, both for why it exists and for why it does not exist” (Spinoza, 2018,10). God exists, according to Spinoza, not because of any external cause, but because of his nature as the supremely perfect and absolutely infinite being, and so God exists of himself (Spinoza, 2018,12). Spinoza goes on to argue that God is the only substance that exists, capable of existing on its own and not as an attribute or mode dependent on an underlying substance. “By substance I mean that which is in itself and is conceived through itself...” (Spinoza, 2018, 3). An attribute, on the contrary, is the essence of a substance, and so is inconceivable without the substance underlying it, while a mode is an affection of a substance (Spinoza, 2018, 3). Since Spinoza maintains that God is the only substance, everything else must be either an attribute or a mode of God: “No substance can be or be conceived besides God” (Spinoza, 2018,14).

Spinoza approximates the universal intelligibility characteristic of PSR, insofar as everything is rooted in the single substance of God. “Whatever is, is in God, and nothing can be, or be conceived, without God.” (Spinoza, 2018, 15). Since everything exists as a feature of God’s unique substantiality, everything happens “solely by the laws of God’s infinite nature and follows from the necessity of his essence.” (Spinoza, 2018, 18). The idea that God’s activity depends on the laws having to do with a necessity of his essence, though, and all that happens follows from these laws, is the crucial point at which the version of PSR we see in Spinoza diverges from what Leibniz maintains as the correct version. The divine nature in Spinoza does not unfold its infinite attributes according to purpose and choice, but according to a necessity based on its essence. “There is no cause besides the perfection of his own nature which prompts God...to act.” (Spinoza, 2018, 19).

PSR in Spinoza, therefore, does not account for events in the world in terms of what is best, because events flow out of God’s nature, not from a choice of the best possible world, but from the necessity flowing from God’s perfection. We can use Kant’s distinction between mechanical and teleological causation, in the third *Critique*, to classify the version of PSR we find in Spinoza. In mechanical causation, “the idea of the effect must not be taken as the ground of the possibility of their cause, but *vice versa*...” An aspiration towards the idea of the ultimate effect of a mechanical process is not the sufficient reason that explains why the parts exist and act the way they do. In a mechanical process the parts account for the effect (Kant 2000, 20:236). In a teleological process, the idea of the whole, the final outcome of the process, constitutes an end that accounts for the existence and activity of the parts.

When God chooses the best of all possible worlds, there is teleological causation, because the idea of the best possible world constitutes an end that drives the existence and evolution of the parts of the forming world. When God exercises mechanical causation, not from purposes but from the necessity of his nature, the idea of the best possible world is not driving the process. The final outcome of a mechanical process does not account for the process, but the parts do in piecemeal fashion.

Since “God acts by the laws of his own nature,” (Spinoza 2018,19) in Spinoza, and not from an idea of what is best, God in Spinoza exercises mechanical causation. God is not a free cause, in Spinoza, in virtue of an ability to select a particular possible world that serves as an aim driving processes forward. Spinoza denies to God both intellect and will, by which God might choose his idea of what is best (Spinoza 2018, 19). Instead of portraying God as sovereign over an array of possible realizations, choosing one with deliberate purpose, God acts with the same non-purposive necessity as we find in geometrical deductions. “...Just as from eternity to eternity it follows from the nature of a triangle its three angles are equal to two rights angles,” (Spinoza 2018, 20). the necessity of God’s nature dictates the processes that flow from him.

In Spinoza, we have a kind of satisfaction of PSR, insofar as we can account for everything in terms of the necessity of God’s nature, and so there are no gaps in nature involving unaccountable swerves. But universal causation does not mean that things have a *reason*. A series of events linked causally, driven by the laws of the nature of a necessary being, still has an element of arbitrariness because we do not know *why* this series of events is proceeding as it does. Leibniz maintains that universal causation, without a reason that assigns an end to it indicating an ultimate purpose, is still arbitrary.

To truly satisfy PSR, we must complement the series of causes with an overarching reason identifying a purpose driving the series of causes. Leibniz seeks an understanding of the grounds of the laws of mechanics in terms of the wisdom of the Author of nature (Leibniz, 1969, 478). Leibniz unites both final and efficient causation in his understanding of nature. We can understand the “skill of the workman,” i.e. God, in terms both of his purposive designs (final causation), but also in terms of the tools he uses to achieve his designs (efficient causation) (Leibniz, 1988, 22). Spinoza severs efficient causation from final causation, since he denies that effects flow from God’s nature according to a purpose. Instead, they flow from God’s nature with the same blind necessity that dictates the essential properties of triangles, which do not purposively elect to have their properties.

Explanations that appeal only to efficient causes, without an appeal to an end, operate in the “realm of power,” which appeals only to blind forces to account for things, without any appeal to purposes (Leibniz 1969, 479). In the realm of wisdom, we integrate things into an architectonic that explains them in terms of a purpose-oriented drive towards the realization of a whole. Instead of operating in the realm of power, which can give us only antecedent causes that have no ultimate purpose, Leibniz provides us with an ambitious metaphysics that orients everything according to a purposive wisdom. The architecture of nature perfectly expresses God’s just judgments, not mere blind necessity. “God as the architect satisfies in all respects God as the legislator.”( Leibniz 1965, 89). Instead of the *only* possible world, necessitated by the laws of God’s nature, we find in Leibniz the *best* of all possible worlds, chosen freely in the wise design of God. The laws of nature “originate in the wisdom of their Author...” (Leibniz 1969, 478).

We can sum up Leibniz's position, *vis-à-vis* Spinozism, in the following quote: "Spinoza...appears to have explicitly taught a blind necessity, having denied to the author of things understanding and will, and imagining that good and perfection relate only to us and not to him..." (Leibniz 1952, 173). Spinoza's God only mechanically follows the principles of his own nature, in complete indifference to the categories of good and bad. This indifference to questions of value cannot satisfy PSR.

*Critique of Cartesian Voluntarism.* In the Cartesian philosophy, we find the notion that God is sovereign creator of necessary truths regarding essential natures, like the properties of triangles. The truth value of propositions depends on the will of God, since God is omnipotent. Truth is not an order of being that God must respect, like we must, since truth is something we discover, and do not create. Our ideas are not true when we invent them, but in virtue of a correspondence with a reality independent of us. So, Leibniz insisted that no act of will, human or divine, can play a part in determining truth (Leibniz, 1981, 66). *Theological voluntarism* is the view that aspects of reality, which we might consider to be independent of the will of God, such as truth and goodness, are actually products of God's will. God, in the view of theological voluntarism, does not act in light of knowledge of what is good and what is true, receiving knowledge of an objective order independent of him. Instead, God creates the orders of goodness and truth *ex nihilo*, beginning from a position of complete indifference (Descartes, 1964,151-2). God chooses the norms of perfection, instead of accepting them as the external constraints of a moral order that pre-exists God's choice. God also chooses the norms of logic, even retaining sovereignty over the principle of non-contradiction (Descartes, 1964,151-2).

If God's will is sovereign even over the principle of non-contradiction, then the will of God is incomprehensible. Theological voluntarism means that God's will could completely invert the moral and intellectual order, making contradictories co-exist and turning what is good into what is evil. We cannot establish why God made certain things good and true, because there is no order of goodness and truth prior to God's creation of the moral and intellectual order by which we might account for it. Universal intelligibility in Leibniz means that PSR constrains God's most fundamental choices about the creation of the order of things. But, if God's will generates even PSR, there is no way to place limits on God's sovereignty according to PSR. God's sovereign power of choice can withdraw endorsement from PSR, and then the entire intelligible order would collapse, leaving nothing but arbitrary occurrences that happen for no reason. If God creates the principle requiring him to choose what is best, he can redefine what is best, or generate a new fundamental principle permitting arbitrary choices.

Leibniz rejects voluntarism, claiming that "the choice made by God must have a sufficient reason which determines him to the one rather than to another [possible universe]." (Leibniz, 1965, 53). It is the order established by PSR that determines God, not God who determines the order of PSR. God's wisdom discloses the rational order of PSR, and his goodness determines his choice

of the best possible world (Leibniz, 1965, 53). This is a very different direction of causality than what we find in voluntarism, which maintains that God determines what is good and creates, rather than receives a disclosure of, what is true.

If we allow the order of what is true to be subject to the will of God, so that it is manipulable, then PSR collapses. We will not be able to establish a sufficient reason for the existence of things, because reality will be subject to constant incursions by arbitrary choices of the divine will. Theological voluntarism is unacceptable, in Leibniz's view, because it makes what is true a malleable order, and so it collapses the stable comprehensibility that is the hallmark of Leibniz's metaphysics.

In this section, I have shown how PSR constitutes grounds for rejection in Leibniz, of both the mechanistic necessitarianism without final causes, as well as the *ex-nihilo* generation of goodness and truth in theological voluntarism. The employment of PSR in Leibniz amounts to a war on the absolutely contingent. Events in the Spinozistic universe have absolute contingency in the sense that there is no fully sufficient reason, in the order of value at least, for them. We find absolute contingency in Descartes, in the initial choice of God to found the moral and intellectual orders in *ex nihilo* fashion, without need to pay respects to a pre-existing moral and intellectual order independent from God's will. Leibniz demands that God's choices have a fully sufficient reason, and so are purposive instead of necessitarian, and also be answerable to an objective order it does not arbitrarily create.

## 2. Human Volition in Leibniz

Leibniz's ambition for universal intelligibility encompasses human volition. Each human action follows from the individual notion that God has of the individual. God is able to see, in *a priori* fashion through this individual notion, all the actions a person will ever perform (Leibniz 1988, Section 13). God does not have to experience, in *a posteriori* fashion, the actions a person takes. The individual notion of a person contains all the actions of the person, with the same *a priori* rational necessity with which we can derive properties from the definition of a circle (Leibniz, 1988, Section 13). The scripted individual concept of a person follows from God's plan of the best of all possible worlds (Leibniz, 1988, Section 13).

Leibniz, while conceding that all the actions of a human are implicit in his or her *a priori* concept, still seeks to carve out a space for freedom. Leibniz seeks to allay the fear that, by making all human actions follow with *a priori* necessity from the individual notion of the human agent, "the difference between necessary and contingent truths will be destroyed [and] the fate of the Stoics will take the place of liberty..." (Leibniz, 1988, Section 13).

Leibniz's rebuttal to the fear of the fatalism of the Stoics, smothering any options for the human will in light of infallible divine foreknowledge, is to differentiate between "what is certain and what is necessary." (Leibniz, 1988, Section 13)

Future actions, implicit in the individual concept, are *certain* to happen, insofar as God can deduce them in a *a priori* fashion with certainty. Yet, future actions are contingent, not necessary, because God's choice of the best of all possible worlds, in which the future actions figure, is itself contingent. God has the prerogative of "free decrees," i.e. he could have chosen otherwise than to actualize the best of all possible worlds. The certainty of the future actions of a human being are based on their participation in the causal sequence characteristic of the best of all possible worlds. But, the certainty of these future actions is not necessity, because the initial choice of the best of all possible worlds is contingent.

Another way of separating the certainty of human action, from a necessity characteristic of Stoic fatalism, is the distinction between absolute necessity, and hypothetical necessity (Leibniz, 1988, section 13). Absolute necessity is a logical necessity that exists because its contrary is a contradiction. Hypothetical necessity is not necessary by logic, but only on the condition of existing within a certain causal sequence. The contrary of what is hypothetically necessary is possible in itself, apart from a given causal series.

Human actions are only hypothetically necessary because, though God's choice of the best of all possible worlds necessitates them, their contrary is not a contradiction. It is possible for human actions to have been different, though not so long as we remain within the causal sequence established by God's choice. The idea of an individual concept containing *a priori* all the future actions of an agent does not usher in full Stoic fatalism, because future actions are still contingent, insofar as their contraries are not a contradiction, and they depend on the *free* decree of God.

But, establishing human actions as hypothetically necessary is a narrow space for human freedom. A logical conceivability of a different action that is not a contradiction does not change the fact of God's pre-determination of each individual concept according to his choice of the best of all possible worlds. The notion of Julius Caesar contains his action of becoming a perpetual dictator, so that he cannot escape this fate. The fact that another action is logically conceivable, apart from the precondition of God's decree, does not alleviate the fact that it is necessary for Caesar to conform to God's imposition (Leibniz, 1988, Section, 13). A thinker like Crusius maintains that even if we can establish the contrary of our individual concept as possible, the individual concept is still necessary in the actual causal structure of the world. "For of what avail is it if the opposite of an event, which is precisely determined by antecedent grounds, can be conceived when it is regarded in itself, since the opposite still cannot occur in reality..." (Kant, 1996, 1:339).

This unsatisfactory state of affairs for human freedom motivates the project of carving out a space for *absolute* contingency. The category of the absolutely contingent is able to escape both logical necessitations, since its contrary is conceivable, and determination by antecedent grounds, so that these grounds provide space for agents like Julius Caesar to choose otherwise than some pre-determined fact imposed by God. In the following section, I will show how Kant moves away

from the ambitious architectonic of Leibniz, in which all of reality has intelligibility in terms of the choice of the best of all possible worlds. Kant's philosophical system permits unintelligibility at its limits, and this facilitates the creation of a space of absolute contingency that eludes PSR. By carving out a space of absolute contingency, Kant avoids an anthropology in which human beings are nothing more than marionettes acting out a pre-determined plan of God (Kant, 1996, 5:101).

### 3. Kant's Modesty

Kant's system curtails the universal intelligibility of Leibniz's system. Kant imports PSR into his critical system, but not as an ontological principle, and not as a teleological version of PSR in which purposive final causes provide adequate explanations. Instead of an ontological principle, Kant limits PSR to a principle of possible experience. The discipline of ontology, which provides an account of being in itself, is no longer possible under the epistemological restrictions in Kant's system. To do ontology requires using metaphysical concepts, like substance and causation, to describe things *in general*, "without taking regard of the way in which we might intuit them." (Kant, 1998, A247/B303). But, in the critical philosophy, it is illegitimate to apply metaphysical concepts to being in itself, beyond possible experience<sup>1</sup>. Experience provides the only manner in which the understanding can make contact with objects. Any pure use of the understanding, extended beyond possible experience to being in itself, lacks epistemic grounding. "...The pure concepts of the understanding can never be of transcendental, but always only of empirical use..." (Kant, 1998, A247/B303).

Kant's epistemological restrictions transform the role of metaphysical categories. The metaphysical categories, in Aristotle, constitute the highest kinds into which we can divide being (Aristotle 1984, 25). Instead of articulating the structure of being in itself, as metaphysical categories, the categories in Kant become the *a priori* form of possible experience. The pure understanding cannot articulate the basic structure of being, but can only "anticipate the form of a possible experience in general..." (Kant, 1998, A247/B303).

Since the *a priori* use of the categories can only give us the form of possible experience, and not the structure of being, ontology as a discipline dies in the critical philosophy. The modest discipline of the "analytic of the understanding," i.e. the articulation of the basic structures of experience, replaces ontology (Kant, 1998, A247/B304). Ontology is supposed to provide "*a priori* cognitions of things in general in a systematic doctrine (e.g., the principle of causality) ...," (Kant, 1998, A247/B304) not things as they appear to us. This systematic doctrine of being in general includes a role for the principle of causality, i.e. the PSR by which nothing happens without a sufficient reason. Leibniz uses the category of causation as a metaphysical category to articulate the structure of being in general, not just in terms of how we experience it.

---

<sup>1</sup> KrV A539/B567. Kant, Immanuel (1998) *Critique of Pure Reason*. Edited and translated by Paul Guyer and Allen Wood. Cambridge: Cambridge University Press.

This sort of ontological ambition is not possible in the critical philosophy. What reality is like in general, apart from any experiential context associated with our sensibility, is unknown. “What may be the case with objects in themselves...remains entirely unknown to us.” (Kant, 1998, A42/B59). We have to remain agnostic about the ultimate metaphysical structure of things. The structure of causality in experience does not necessarily transfer over to the structure of being, and so Kant permits some gaps in the universally intelligible architectonic of Leibniz.

The categories in Kant constitute a *transcendental logic*, which has to do with the most general structures in virtue of which can provide categorical classification of the manifold given by sensibility, i.e. the diverse packet of intuitions prior to mental synthesis (Kant, 1998, A77/B102). Transcendental logic underlying the content sensibility provides is distinct from a transcendental ontology that classifies structures of being. Kant also inverts the transcendentals of scholastics (Kant, 1998, B113). The transcendentals in scholasticism were supposed to be features that pertained to every category of being, bridging even the distinction between finite and infinite being (de Boer 2020, 82-3). The transcendentals of medieval scholasticism, like the metaphysical categories of ancient philosophy, become, in Kant, structures of the cognition of things, rather than structures of things in general, apart from cognition. “These supposedly transcendental predicates of things are nothing other than logical requisites and criteria of all cognition of things in general...” (Kant, 1998, B114).

PSR, rather than a metaphysical principle bridging the gap between finite and infinite being, constraining even God himself, as PSR does in Leibniz, becomes in Kant an analogy of experience. An analogy of experience provides a necessary relational structure in experience, but does not provide specific determination, in terms of the content and quantity of phenomena in experience (Kant, 1998, B223). The second analogy of experience maintains that “all alterations occur in accordance with the law of the connection of cause and effect.” (Kant, 1998, A188/B233). Nothing can arise, apart from a law connecting it to an antecedent ground, according to the second analogy of experience, and this principle at least approximates PSR. But, an analogy of experience pertains to “empirical consciousness (of perception),” (Kant, 1998, A177/B220) not to being in itself.

*Second Analogy Prevents Full Realization of PSR.* The second analogy of experience, though it is a form of PSR, actually prevents the full realization of PSR that we see in Leibniz, in the form of an ontological architectonic of the best of all possible worlds. The second analogy (hereafter, 2A) forbids the occurrence of an event in experience that lacks a cause. An unconditioned event, that lacks an antecedent ground, “conflicts with the dynamic law of the determination of all appearances in time...” (Kant, 1998, A453/B481).

2A precludes what is unconditioned in experience, because what is unconditioned does not follow from an antecedent event according to a causal law (Kant 1998, A533/B559, fn). Yet, Kant is aware that PSR cannot continue infinitely with a series of entities, arising from a causal law from an antecedent ground, forever. An infinite series cannot constitute a total explanation, because an

infinite series is never completed. PSR, in Kant, requires, for every conditioned thing, an ultimate grounding in what is *unconditioned*, so as to prevent an infinite regress. “If the conditioned is given, then the whole series of all conditions for it is also given...” (Kant, 1998, A497/B525).

The demand for the unconditioned, to avoid the infinite regress, exists also in Leibniz’s metaphysics. “The ultimate reason of all things must subsist in a necessary substance...” (Leibniz, 1965, 38). The necessary substance, containing the reason for its own existence within itself, is an unconditioned entity that does not rest on anything external to itself.

Yet, PSR in Kant, in its demand both for continuity within experience, which precludes anything unconditioned in experience, and the demand for an unconditioned that avoids an infinite regress, creates a tension. The demand for continuity in experience, enjoined by 2A, precludes the demand for completeness in PSR, since experience cannot include what is unconditioned, only what follows from a prior ground. “For the understanding does not permit among appearances any condition that is itself empirically unconditioned” (Kant, 1998, A531/B559). 2A, in its application of PSR, precludes the very unconditioned ground necessary to fully satisfy PSR by avoiding an infinite regress. PSR’s demand for completeness, then, necessarily exceeds experience, and so, in the critical philosophy, PSR’s demand for completeness can never be an object of knowledge. The necessary being cannot figure in experience, because the necessary being is an unconditioned reality that clashes with the structure of experience based on 2A. By extending the necessary being beyond possible experience, Kant removes it from any experiential validation. Experiential validation is the only way in the critical philosophy to establish existence of objects. “If I take all thinking...away from an empirical cognition, then no cognition of any object at all remains...” (Kant, 1998, A253/B309). The necessary being, since it cannot figure in any experience, is an empty idea, not associated with any object. “...If one searches for the unconditioned among conditioned things, then one will seek forever and always in vain, since no law of any empirical synthesis will ever give an example of such a thing...” (Kant, 1998, A621/B649).

We cannot infer from conditioned realities in experience, to an unconditioned reality, because “all laws of transition from effects to causes...are directed to nothing other than possible experience...” (Kant, 1998, A622/B650). The only significant application of the transition from effects to causes is within experience, and so the transition to an unconditioned being that surpasses experience is without significance.

PSR, in Kant, contains a *universality* requirement, expressed in the second analogy of experience that all events must follow an antecedent ground according to a causal law (Kant, 1998, A188/B233). At the same time, PSR expresses a *completeness* requirement, requiring a total explanation terminating in what is unconditioned that avoids an infinite regress that never establishes a sufficient reason. The completeness requirement, in Kant, clashes with the universality requirement, because the unconditioned reality demanded by the completeness requirement exceeds the universality requirement that is the formal structure of experience in Kant



(Allison, 1990, 22). The only way the unconditioned reality can exist is as an intelligible object, beyond possible experience, but there can be no knowledge of an intelligible object<sup>1</sup>. “If...one would separate it [the highest being] from this chain, and, as a merely intelligible being, not include it within the series of natural causes, then what bridge can reason build so as to reach it?” (Kant, 1998, A621/B649).

Kant breaks down the architectonic that Leibniz constructed using PSR. Within experience, we only seek the unconditioned as a problem, and not as something given (Kant, 1998, A499/B527). As for the existence of an unconditioned reality, we can only remain agnostic, not knowing for sure of its existence, because we cannot have knowledge apart from experience.

*Non-Teleological PSR in Kant.* There is a diminishment of PSR in Kant, in the sense that it is merely mechanical, not teleological. The universality requirement for possible experience contained in 2A, that “all alterations occur in accordance with the law of the connection of cause and effect,” (Kant, 1998, A188/B233) does not include any teleology. The fact that an event follows an antecedent cause does not include the idea of the effect as the ground of the cause, animating the cause an end (Kant, 2000, 20:236). A rule of causal relationships, established in 2A, is separable from a value-driven orientation to what is best.

Teleology, in the critical philosophy, is the prerogative of reflecting, not determining judgment. We posit the idea of an end driving causation as a principle guiding investigation, not as a property of an object. Reflecting judgment posits teleology merely for its own behalf, as a convenient tool for classification, and not as some objective structure of being guiding all things towards what is best (Kant, 2000, 20:236). Reflecting judgment presupposes an order in nature such that it lends itself to an orderly classification under species and genera. Without this presupposition of order, we would face the possibility of a nature of such diversity that we could find no continuity of classes. But, this presupposition of order does not arise from a cognition of nature in itself. Instead, the presupposition of order that guides reflecting judgment is “its own subjective law, in accordance with its need...” (Kant, 2000, 20:214). As Neiman points out, “The point of Kant’s appeal to teleology can only be understood if his insistence that it does not give us knowledge of the world is fully appreciated” (Neiman, 1994, 82). Teleology in Kant is not constitutive, i.e. an objective feature of the world, but a subjective principle of our cognitive faculties.

Whereas Leibniz’s metaphysics gives us a full architectonic in which everything has its place in the best of all possible worlds, Kant’s system is incomplete, forcing us to search for the unconditioned without ever being able to find it, and only posits teleology as a subjective guide. I explain, in the fourth section, how the limited application of PSR in Kant allows him to carve out a space of absolute contingency, which has no sufficient explanation either through logical necessity or in terms of the actual causal structure of the world.

---

<sup>1</sup> KrV A42/B59. Kant, Immanuel (1998) *Critique of Pure Reason*. Edited and translated by Paul Guyer and Allen Wood. Cambridge: Cambridge University Press.

#### 4. Absolute Contingency

Kant creates a space of absolute contingency by limiting PSR to a principle of experience. If PSR is a principle of experience, then PSR need not apply to things in themselves. An unconditioned reality, occupying a space of absolute contingency that is independent of any antecedent cause, is possible at the PSR-exempt level of things in themselves.

A candidate for such an unconditioned reality is libertarian free will. Libertarians in the free will debate believe that free will exists, and that free will is incompatible with a deterministic series (Pereboom & McKenna, 2016, 31). Libertarian free will challenges the contemporary denials of free will in the natural sciences, which present it as a mere illusion<sup>1</sup>, or in politics, which view individuals in terms of social structures. Since libertarian free will is independent of a deterministic series, it has the potential to figure in the modal category of absolute contingency. Instead of being subject to external causation that forces free will towards a specific outcome, we can account for libertarian freedom solely in terms of its exercise, and not in terms of antecedent grounds. We find this sort of libertarian freedom, independent of any causal network, in Crusius. “He [Crusius] thinks that the free will is actually determined by its existence, not antecedently by grounds...” (Kant, 1996, 1:397).

In his engagement with Crusius early on, Kant rejected the idea of a self-determining libertarian freedom, that excludes alternate realizations solely in terms of its own actuality. Kant committed to the principle that “a contingent thing is never sufficiently determined, if you abandon the antecedently determining ground...” (Kant, 1996, 1:397). One of the innovations of the critical period is a rapprochement with Crusius, in the form of Kant’s commitment to a libertarian conception of freedom that arises independently of any antecedent ground. In the first *Critique*, Kant presents freedom as distinct from the causality of nature. The causality of nature is subject to laws of nature, in virtue of which events unfold from a previous state according to a deterministic rule (Kant, 1998, A444/B472). The causality of freedom is distinct from the causality of nature, insofar as the causality of freedom is “an absolute causal spontaneity beginning from itself...” (Kant, 1998, A446/B474). Causality of freedom exists through its own actuality, without any antecedent ground.

Unconditioned causality through freedom is “contrary to the causal law, and is a combination between the successive states of effective causes in accordance with which no unity of experience is possible...” (Kant, 1998, A447/B475). The unconditioned causality of freedom disrupts the causal sequence without which we cannot have any experience at all. But, PSR in the critical period is a principle of experience, not an ontological principle, and so PSR does not exclude the possibility of an unconditioned causality of freedom. Kant’s doctrine of transcendental idealism, whereby the conditions of possible experience pertaining to all categories of experience (hence,

---

<sup>1</sup> The latest installment is Robert M. Sapolsky, *Determined: A Science of Life Without Free Will*, (Penguin: 2023).

transcendental), are features of the mind (hence, idealism), permits the possibility of causality through freedom in the domain of things in themselves, apart from the categories of experience (Kant, 1998, A491/B519).

Transcendental idealism means that appearances participate in *a priori* structures of the mind, and so appearances are mind-dependent and not part of reality in itself (Kant, 1998, A491/B519).

Transcendental *realism* posits the spatiotemporal framework as an ontological structure existing in its own right, apart from experience. “The realist, in the transcendental signification, makes these modifications of our sensibility into things subsisting in themselves...” (Kant, 1998, A491/B519).

The adoption of transcendental idealism allows Kant to limit the causal law of PSR to experience. Transcendental idealism means that the causal law connecting events in space and time is only a feature of the mind, not necessarily a feature of things in themselves. Transcendental idealism is part of the epistemological modesty that prevents Kant from establishing the architectonic we find in Leibniz, which upholds PSR as a principle governing all of being. This epistemological modesty allows Kant to permit the possibility of an unconditioned libertarian freedom in the domain of things in themselves, apart from experience. If the causal law is only a function of the mind’s cognition of reality, i.e. of appearances and not necessarily of things in themselves, then reality in itself may be exempt from it. Exemption from the causal law permits the possibility of libertarian freedom that is incompatible with determinism.

Kant identifies two characters, or laws of causality (Kant, 1998, A539/B567) pertaining to the human agent. The empirical character of the human agent participates in experience, and so partakes of the causal law characteristic of experience. “...In the temporal succession all actions of natural causes are themselves in turn effects, which likewise presuppose their causes in the time-series” (Kant, 1998, A544/B572). There can be no unconditioned libertarian freedom exercised by the empirical character, because the empirical character is subject to the universality requirement of the causal law, i.e. *every* event in the empirical domain must follow by a causal law from a previous event. The empirical character forms the horizon for contemporary denials of human agency, which account for our choices entirely in terms of external forces, e.g. biology, neurochemistry, group dynamics, etc.

In order to establish freedom at the empirical level, we could develop a compatibilist account of freedom that accepts determination by an antecedent ground, but nevertheless permits freedom from the fact that one’s actions come from conscious reflection instead of impulse (Pereboom & McKenna, 2016, 50). Kant discusses a compatibilist form of freedom that accepts external determination, but maintains freedom in virtue of the fact that the marionette is at least a self-conscious one (Kant, 1998, 5:101).

But, the recognition that the proximate cause of an action has to do with what is internal to an individual’s psychology, rather than some insensate process, does not change the fact that the action is the inevitable byproduct of a prior cause independent of the agent’s control. The self-conscious

agent is only self-conscious of what he *must* do in virtue of pre-determined causal sequences over which he has no control (Kant, 1998, 5:101).

With the adoption of transcendental idealism, the empirical character subject to determinism is a mere appearance, and so is a function of mental structures, not a thing in itself. By making the empirical character a mere appearance, Kant is able to limit the ontological application of the deterministic causal law characterizing experience. As a result of this limitation, unconditioned freedom, which is incompatible with a deterministic causal law, is possible for the agent as he is in himself, independently of the way he appears.

The *intelligible* character represents that aspect of our agency which is independent of the way it appears (Kant, 1998, A538/B566). The intelligible character has potential for a completely different law of causality than what characterizes the empirical character. Since the intelligible character is not an appearance, it is not even subject to temporal succession in the way that appearances are. "This acting subject, in its intelligible character, would not stand under any conditions of time, for time is only the condition of appearances but not of things in themselves." (Kant, 1998, A540/B568). As something independent of temporal succession, the intelligible character is independent of the causal law that connects, in experience, events occurring in temporal succession. Independent of the causal law, the intelligible character is capable of unconditioned freedom: "Of it [the intelligible character] one would say quite correctly that it begins its effects in the sensible world from itself..." (Kant, 1998, A541/B569).

The unconditioned freedom of the intelligible character is absolutely contingent. It is possible for it to be different in other possible worlds, since its contrary is not a contradictory. Also, the unconditioned freedom of the intelligible character is independent of any prior causal series. The intelligible character rises above the temporal succession, subject to a causal law, that characterizes experience.

Nicholas Stang, in his work on Kant's modal metaphysics, identifies a form of possibility as *noumenal-causal possibility*, that helps us account for the absolute contingency of human freedom in its intelligible aspect. Noumenal-causal possibility refers to the possibility that the grounds of experience, themselves not possible experiences, can generate a different experience (Stang, 2016, 224). A different experience is noumenally-causally possible insofar as its noumenal ground can generate a different one. With respect to free will, an agent's actions are necessary with respect to the causal series in the empirical character. But there is a noumenal-causal possibility to change these actions, insofar as, the intelligible character, which is the ground of the empirical character, can generate a different deterministic causal series in the empirical character. So, though events characterizing the empirical character may be necessary, there is a noumenal-causal possibility not subject to this necessity, that can change the character of the necessity at the empirical level. The causal law at the empirical level does not constrain the intelligible character, which is not in

experience, but is the ground of experience. "The empirical character is...determined in the intelligible character." (Kant, 1998, A551/B579).

Kant's permission of an unconditioned power of freedom, independent of any antecedent ground, is the impetus for a new humanism that is scandalous for the absoluteness of divine sovereignty, as well as for modern ideologies that deny human agency in the name of science or politics. An absolutely contingent human will gives human agency a sovereignty over itself that is competitive with God. God might have a will for the best possible outcome, but the human will is independent, and so might undermine the achievement of this outcome. Kant limits teleology to a mere subjective guide for our investigation of nature, and so teleology need not be an objective determination of the structure of nature (Kant 2000, 20:237). In refraining from making teleology an objective structure of nature, Kant allows the teleology associated with the *human* will to become a competitor with any teleology associated with God. A teleology of the human will that is absolutely contingent, moreover, can individuate itself from natural forces or group dynamics.

In Leibniz, human choices follow from the individual concept God already has of each person. In Kant, the human agent has access, in its intelligible character, to a modality of absolute contingency by which it exercises causality in a way entirely independent of antecedent causes. The absolute contingency of the intelligible character is independent of a theological determinism in which the actions of an agent are a product of God's preliminary choice. The human will is the absolute sovereign over its own actions, and is able to remake the world according to its own judgment. Nor is the idea of the human agent a mere extension of nature or group dynamics, since the absolutely contingent intelligible character determines itself, independently of any antecedent ground.

*Human Freedom and Divine Sovereignty.* We see Kant's humanistic endorsement of a world made in our image, rather than one subject to the individual concepts chosen by divine wisdom, in his meditation on the threat posed to human freedom by God's creation of the human substance. Kant worries that, if God's creative activity, as the being of beings, accounts for the existence of human substances, then all our actions will have an ultimate ground in what is beyond our control, i.e. the "causality of a supreme being..." (Kant, 1998, 5:101).

Human agency subject to divine predetermination is what we find in Leibniz. But, the distinction between the empirical and intelligible characters means that the causal series, driven by PSR, characterizes the agent only as an appearance, and not necessarily as a thing in itself. The exemption of the intelligible character from PSR permits the agent a space independent from theological determinism. "If a human being's actions insofar as they belong to his determinations in time were not merely determinations of him as appearance but as a thing in itself, freedom could not be saved" (Kant, 1998, 5:101). If, that is, the temporal series of events connected by causal law characterized the agent in itself, and not merely as an appearance, the temporal series of events

would connect the agent in himself, not merely as an appearance, back to the all-sufficient will of God, making him an expression of God's ultimate control.

But, the deterministic causal series, characterizing the agent's empirical character, is an appearance in time, whereas the divine action, occurring in eternity occupied by an infinite and independent being outside of time, pertains to reality in itself. This distinction prevents the determinism characterizing the agent's empirical character from following necessarily from the action of God. The action of God does not control the causal succession at the empirical level, given the fact that the action of the infinite and independent God is above time, and so pertains to the intelligible level. "...It is quite easy for us to distinguish between the determination of the divine existence as independent of all temporal conditions and that of a being of the sensible world, the distinction being that between *the existence of a being in itself* and that of a *thing in appearance*" (Kant, 1998, 5:102).

God acts in the domain of things in themselves, but the fallout of this action does not flow into the causal series structured by PSR, because this causal series is an appearance characterizing only the structure of experience. If we renounce epistemological modesty, and attribute the causal series structured by PSR to reality in itself, then we connect the causal series to the divine activity at the level of things in themselves, and this creates a pervasive theological determinism. Epistemological modesty means that God's creative activity, and the deterministic appearance of the agent, occupy different domains of reality. "...Creation [of the human agent by God] has to do with their intelligible but not their sensible existence," since God's creative activity is beyond the spatiotemporal structure characterizing experience. Attributing God's creative activity to the intelligible domain beyond experience means that we cannot regard this activity as "the determining ground of appearances..." (Kant, 1998, 5:102). On the contrary, if we attribute the causal series governed by PSR to the intelligible character, i.e. to the human agent in himself, this causal series shares the same level of reality as the divine creative activity, and so becomes subject to it.

We have a noumenal-causal possibility, in our intelligible character, to ground a distinct causal series in appearances. Since we have this capacity in our intelligible characters, we are able to be independent of God's creative power. A causal series might characterize us empirically, but we can opt out of it, in virtue of our grounding of this causal series at the intelligible level. God's creative power also has access to noumenal-causal possibility, since he can ground, through his will, a different kind of world in space and time. Yet, we are not subject to the structure of the world that God creates, because we have access to noumenal-causal possibility in the intelligible aspect of our wills, by which we can change, as a ground, the empirical world in space and time. If we did not have access to noumenal-causal possibility, the causal series would characterize us absolutely, in our intelligible selves, and this causal series would link us up directly to the creative power of God, also at the intelligible level.

Exercising its power in a space of absolute contingency, the human agent in Kant is a co-creator with God, able to introduce a new causal series with freedom that is unconditioned. Neither natural causes nor group dynamics can make the human agent their passive expression. The absolute contingency associated with the human will gives it an ability to change the course of events that we cannot find in the architectonic of the best of all possible worlds in Leibniz, in which all events follow according to the pre-established individual concept God has of each creature (Leibniz, 1988, section 13).

## Conclusion

I established how Kant carved out a space of absolute contingency, which forms a loophole to the causal structure based on PSR. The modal category of absolute contingency gives to the human agent a sovereignty over its own actions that contends with, and is not a mere expression of, divine sovereignty. Leibniz used PSR to reject absolute contingency in Spinoza, who presents only necessitarian efficient causes without orientation to a purpose, and Descartes, who permits to God's will a sovereignty over truth itself. Kant's limitation of PSR to experience means that the intelligible character of the human agent has sovereignty over itself in a way that eludes the contemporary ideologies that deny human agency.

## References

- Adams, R. (1994). *Leibniz: Determinist, Theist, Idealist*. Oxford University Press.
- Allison, H. (1990). *Kant's Theory of Freedom*, Cambridge University Press.
- Aristotle. (1984). Categories. In *Complete Works of Aristotle, Volume I: Revised Oxford Translation*. Edited by J. Barnes & translated by J. L. Ackrill. Princeton University Press.
- Daniel M. W. (2002). *The Illusion of Conscious Will*, MIT Press.
- De Boer, K. (2020). *Kant's Reform of Metaphysics*, Cambridge University Press.
- De Jung, J. E. (2013). The Modesty of Kant's Metaphysics. In *Kant und Die Philosophie in Weltbürger Absicht: Akten des XI. Kant Kongresses 2010*. De Gruyter, Pgs. 553-562.
- Descartes, R. (1964-1976). *Oeuvres de Descartes*, Vol. I. Edited by P. Tannery & C. Adam. Vrin-C.N.R.S.
- Durkheim, E. (1982). *The Rules of Sociological Method and Selected Texts on Sociology and Its Method*, trans. by W.D. Halls & Edited by S. Lukes, The Free Press.
- Kant, I. (1996). A New Elucidation of the First Principles of Metaphysical Cognition. *Theoretical Philosophy: 1755-1770*. Edited and translated by D. Walford & R. Meerbote. Cambridge University Press,
- Kant, I. (1996). Critique of Practical Reason. In *Practical Philosophy*. Edited and translated by Mary J. Gregor. Cambridge University Press.
- Kant, I. (1998). *Critique of Pure Reason*. Edited and translated by P. Guyer & A. Wood. Cambridge University Press.
- Kant, I. (2000). *Critique of Judgment*. Edited and translated by P. Guyer & E. Matthews. Cambridge University Press.
- Leibniz, G. W. (1952). *Theodicy*. Edited by A. Farrer & translated by E.M. Huggard. Yale University Press.

- Leibniz, G. W. (1965). *Monadology and Other Philosophical Essays*. Translated by P. Schrecker & A. Martin Schrecker. Library of the Liberal Arts.
- Leibniz, G. W. (1969). *Philosophical Papers and Letters*. Edited by L. L. Loemker. Reidel.
- Leibniz, G. W. (1981). *New Essays on Human Understanding*. Edited & translated by P. Remnant & J. Bennett. Cambridge University Press.
- Leibniz, G. W. (1988). *Discourse on Metaphysics and Related Writings*. Edited and translated by R. N. D. Martin & S. Brown. Manchester University Press.
- Lovejoy, A. (1936). *The Great Chain of Being*. Harvard University Press.
- Nieman, S. (1994). *The Unity of Reason*, Oxford University Press.
- Pereboom, D. & McKenna, M. (2016). *Free Will: A Contemporary Introduction*, Taylor & Francis.
- Sapolsky, R. M. (2023). *Determined: A Science of Life without Free Will*, Penguin.
- Spinoza, B. (2018). *Ethics*. Edited and translated by M. J. Kisner & M. Silverthorne. Cambridge University Press.
- Stang, N. (2016). *Kant's Modal Metaphysics*, Oxford University Press.





Journal of Philosophical Investigations

## Journal of Philosophical Investigations

Print ISSN: 2251-7960 Online ISSN: 2423-4419

Homepage: <https://philosophy.tabrizu.ac.ir>



University of Tabriz

# Does Kantianism Imply Some Sort of Conceptual Creationism?

Hemmo Laiho 

Department of Philosophy, University of Turku, Finland. Email: [heanla@utu.fi](mailto:heanla@utu.fi)

### Article Info

#### Article type:

Research Article

#### Article history:

Received 02 July 2024

Received in revised form  
07 July 2024

Accepted 07 July 2024

Published online 07  
August 2024

#### Keywords:

concept, intuition, Kant,  
mind, understanding,  
sensibility, world

### ABSTRACT

I argue in the essay that the conceptualist understanding of the mind-world relation ultimately leads to the kind of view that Panayot Butchvarov calls conceptual or linguistic creationism. According to this view, “there is nothing we have not conceptualized”. In addition to being an antithesis of metaphysical realism, which maintains that there is a reality independent of us, the term refers to the kind of thinking that sees human cognitive experience (and reality itself) as thoroughly constituted according to our concepts. While it might be easy to attribute this kind of position to Kant as well, especially when read through a conceptualist lens, I argue that such a position is not in accord with Kant’s philosophical intentions. Using the Deduction and Schematism chapters of the Critique of Pure Reason as examples, I also argue that on the conceptualist understanding of the mind-world relation too much is read into Kant’s idea that sensibility and understanding must be cognitively compatible with one another.

**Cite this article:** Laiho H. (2024). Does Kantianism Imply Some Sort of Conceptual Creationism? *Journal of Philosophical Investigations*, 18(47), 49-62. <https://doi.org/10.22034/jpiut.2024.18257>



© The Author(s).

<https://doi.org/10.22034/jpiut.2024.18257>

Publisher: University of Tabriz.

## Introduction

There is already a plethora of philosophical literature proposing that Kant is a conceptualist, or that he is a nonconceptualist, or that he is something in between (see, e.g., Allais 2016; McLear 2020). I take the ultimate issue in the debate to be how the mind-world relation is constituted. Basically, the issue stands on the question of how the mind connects with the world, and what is the role of concepts and judgments in that connection, or whether we should regard *sensibility* or *understanding* as primary in establishing that connection.

In what follows, I mostly maintain a bird's eye viewpoint on the debate. My starting point is what Panayot Butchvarov calls *conceptual* or *linguistic creationism*, or “the heady view that there is nothing we have not conceptualized or verbalized” (Butchvarov, 2002, 300). In addition to being an antithesis of the metaphysically realist idea that there is a reality independent of us, conceptual/linguistic creationism refers to the kind of thinking that sees human cognitive experience—and with it, reality itself—as constituted according to our concepts and language.<sup>1</sup>

First, I will go through some reasons why such a highly anti-realist view is not, strictly speaking, a Kantian position on the mind-world relation. Yet, as I will also argue, the Kantian conceptualist or intellectualist understanding of the mind-world relation leans towards such an unwanted view. Using the Deduction and Schematism chapters of the *Critique of Pure Reason* as my examples, I will then examine the idea of the co-operation of sensibility and understanding, and whether this idea favors Kantian conceptualism or not.

### 1. Conceptual Creationism versus Commonsense versus Kantianism

Let's start by slicing up Butchvarov's description of conceptual/linguistic creationism into two separate claims:

- (1) There is nothing we have not conceptualized.
- (2) There is nothing we have not verbalized.

---

<sup>1</sup> Butchvarov (2002, 300) himself promotes moderate antirealism that aims to avoid conceptual/linguistic creationism: “Our nonrealism acknowledges the virtual tautology that nothing unconceptualized can be the content of judgments or statements and thus serve as evidence or enter in other epistemic relations. But, unlike most current versions of nonrealism, it does not deny the need for something like Kant's distinction between things-in-themselves and things-for-us. It avoids what might be called conceptual or linguistic creationism, the heady view that there is nothing we have not conceptualized or verbalized. Nor does it deny, on the side of things-for-us, the difference between what Kant called sensibility and understanding. That there is such a difference is evident, however difficult it may be to state it. We might say that understanding is up to us, while sensibility is not, but this, though in the right direction, would be misleading or at least vague. It would be better to say that we have some idea of how we may choose to conceptualize differently the things we find, but not of how we may choose to find different things.”

By adding the notions of “creation” and “reality” into the picture, we may modify these two claims as follows: (1\*) Reality is basically a product of conceptualization. (2\*) Reality is basically a product of verbalization, limited to what can be expressed in a language.<sup>1</sup>

All in all, conceptual/linguistic creationism says that reality is “up to us”—fully dependent on human concepts or language or both. If we assume that there can be no proper language use, or meaningful verbal expression, without concepts, then the claim (2\*) entails the claim (1\*). However, the opposite is not necessarily true: perhaps there can be concept use without the use of language.

The commonsensical reaction to any such creationism is that reality is *not* up to us: Whatever the role of concepts and language is in making sense of reality, most certainly the world we live in is not our own creation. We are *in* the world rather than its makers (cf. Goodman 1978). The world constantly pushes and pulls us, so to speak. Reality is as much (or more) like a constraining factor as it is an outcome of the cognitive processes of an individual. Most importantly, worldly objects exist independently of us, and this idea is not under threat due to claims such as the one that the perceptual system of an organism shapes the way the organism perceives objects. It is also phenomenologically evident that even if our descriptions of the world did depend on concepts and language, as they most certainly do, we also feel and perceive things in ways that go beyond the reach of our concepts and capacities of linguistic expression. In a word, much of reality is ineffable. Or at least some of it is, and in any case the world, and our experience of it, is “richer” or “denser” than our somewhat limited and coarse cognitive attempts at capturing it in everyday descriptions, scientific theories, religious belief, worldviews, and the like.

Kantianism appears to balance between the creationist view and commonsense realism. Creationistic looking Kantian items include space and time as “transcendentally ideal”, or as the necessary sensible features of possible experience that structurally precondition the actual perception and imagination of objects (see *CPR*, esp. Transcendental Aesthetic). The categories make another example: The Kantian concept of Substance, for example, does not pick out some sort of absolute or ultimate existents of the world as such, but, rather, indicates a cognitive rule employed in regarding spatiotemporally appearing objects as feature-bundles that exist over time (e.g., *CPR*, B6). In short, the mind-world relation is to be understood in Kantian terms “representationally”, and on one possible interpretation, this means that the worldly objects are objects only in so far as we represent them as being such objects. And in any case, one of Kant’s main points would be that our cognitive apparatus both affects and (at least partly) dictates the outputs of the mind-world relation.

---

<sup>1</sup> Of course, (1\*) and (2\*) do not directly follow from (1) and (2), since the latter could be regarded merely as characterizations of our epistemic limitations, as opposed to ontological claims about reality as such. It seems to me, however, that this distinction—or the lack of it—brings up the core problem with any such “creationistic” views: namely that they tend to run epistemological and ontological considerations together.

Still, the Kantian position does not appear to be utterly “creationist” or “irrealist” (cf. Goodman, 1978). For one thing, the things in themselves—things regarded independently of our representations—play an important, albeit negative role in the Kantian story. This is to say that the representations of things must be metaphysically grounded on something—it is just that we are not in the position to know much about this ground. Alternatively, one could take the main Kantian lesson to be that speaking of things utterly independently of our representations of them just does not make much sense in the end. Yet, we are in touch with reality all the time—surely representationalism does not need to mean that reality is purely a construct, or some sort of intra-mental collection of ideas, as the hardcore idealist would have it (see *Prol 4*: 374-5). Kant’s position would also be an antithesis of sheer subjectivism, or what might be called the pluralism of actual worlds (Goodman, 1978, 2)—after all, space-time and the categories are supposed to be the universal features constitutive of all human cognition and (at least humanly) possible experience (e.g., *CPR* A42/B59; A96-7; B148).

What’s more, the stark Kantian distinction between sensibility and understanding suggests that *conceptual* creationism would in any case be an unfitting term in the Kantian context. This would be so even if we admitted—whether for real or for argument’s sake—that Kantianism implies some sort of cognitive “creationism” or “constructionism”, because the contributions of sensibility—sensations, intuitions, perceptions, feelings, and the like—are not supposed to be conceptual (and even less verbal) items in the Kantian view. Quite the contrary, the whole point in introducing sensibility as a separate faculty from understanding or the intellect—also known as the faculty of concepts and judgments—revolves around the idea that there are elements in our experience that we have not (or need not have) conceptualized or verbalized.

Indeed, some such nonconceptual and nonverbal elements might simply be unreachable through conceptual thinking, and still be of utmost cognitive significance. Kant is quite explicit on this. Take, for example, the representations of space and time as non-discursive, the singularity of intuitions as opposed to the generality of concepts, or the distinguishing between incongruent counterparts, which cannot be based on a conceptual analysis of the objects in question, or the representation of apperception as nothing but a feeling of existence in the end:

Time is no discursive or, as one calls it, general concept, but a pure form of sensible intuition. Different times are only parts of one and the same time. That representation, however, which can only be given through a single object, is an intuition. (*CPR*, A32-2/B47)

What indeed can be more similar to, and in all parts more equal to, my hand or my ear than its image in the mirror? And yet I cannot put such a hand as is seen in the mirror in the place of its original; for if the one was a right hand, then the other in the mirror is a left, and the image of the right

ear is a left one, which can never take the place of the other. Now there are no inner differences here that any understanding could merely think and yet the differences are inner as far as the senses teach[.] (*Prol 4*, 286)

If the representation of apperception, the *I*, were a concept through which anything might be thought, it could then be used as a predicate for other things, or contain such predicates in itself. But it is nothing more than a feeling of an existence without the least concept, and is only a representation of that to which all thinking stands in relation (*relatione accidentis*). (*Prol 4*, 334n)

One could also stress that the worldly objects, understood in Kantian terms as *apparentia*, are at least partly strictly tied to the laws of sensibility (see, e.g., *Prol 4*, 346; *CPR*, B273; *CPR*, A441/B469). If nothing else, the objects obeying these laws are practically forced upon us: we do not make them appear, nor do we make them appear the way they do (at least not in any proper sense of the word ‘make’). The appearances simply define our *perceptual* or *sensory consciousness*, as one could put it. Take for example a ball flying suddenly through a window: there is no escape from the fact that something simply happens—a thing moves and breaks another thing—and makes us see and hear these things and react primitively in a certain way, and all this independently of how we may describe or interpret the situation, or how we may react to it in a more sophisticated manner after the initial shock. There is little we can do to contribute to reality through our concepts and language in that basic sense. I believe Kant would totally agree on this, and I also believe that his further idea that understanding prescribes the laws of nature (see, e.g., *CPR*, B164; *Prol 4*, 319-20) does not change any of that. After all, the same Kant would also maintain that:

The conditions under which alone the objects of human cognition are given precede those under which those objects are thought (*CPR*, A16/B30).

Or at least all of the above is true if the Kantian nonconceptualist or sensiblist or the “sensibility first” view (Hanna, 2015) of the mind-world relation is true. On the Kantian conceptualist or intellectualist understanding of the mind-world relation, however, the cognitive roles of sensibility and understanding are to be understood not only as strictly dependent and intertwined, but understanding is seen as primary. As it is sometimes put, the contributions of sensibility must be governed or guided by concepts, especially by the pure concepts of understanding also known as the categories (e.g., Williams, 2012). On one possible interpretation, like the one advocated by McDowell, there is no such thing as *nonconceptual sensory consciousness* or “bare getting of an extra-conceptual Given” (McDowell, 1996, 9) at all—not even when you primitively perceive the ball hit the window before you realize what just happened. This is because the conceptual capacities are supposed to be “already operative in the deliverances of sensibility themselves” (McDowell

1996, 39), which is also why the contents of perceptual experience must be conceptual contents according to the conceptualist. This opens the door for conceptual/linguistic creationism: if even sensations, intuitions, perceptions, feelings, and the like, are “concept-infused” from the start, then it looks like there is indeed nothing we have not conceptualized and/or verbalized.<sup>1</sup>

## 2. Conceptualism Implies Judgmentalism Implies Verbalism Implies Creationism

A way to defend the conceptualist position—but which does not go very far if one wants to altogether avoid the view that there is nothing we have not conceptualized or verbalized—is to insist that in so far as we aim to make sense of reality, we need to express the contents of perceptual experience in judgments. And everybody agrees that making judgments means the employment of conceptual capacities. This much also seems true: if I want to understand what happened in the episode of the ball hitting the window, and communicate the details of the event to you, I surely do have to express myself using concepts that constitute my judgments—I have to think and say things like: “Did you see the ball break the window? Is that a baseball or tennis ball? Who’s the culprit?”

At the same time, the “judgmentalist”—or should I say the “verbalist”—kind of conceptualism basically shrinks perceptual experience to *sentences*.<sup>2</sup> If this much is true, and in addition, if the notion of nonconceptual sensory given is excluded from the equation, it is difficult to see how to avoid the slide towards the position according to which concepts and language dictate the mind-world relation. Stress the language-dependence enough and you are left without some of the core reasons to debunk conceptual/linguistic creationism: You cannot anymore say that reality cannot be fully captured by judgments or expressed through linguistic means. You cannot say that reality is to some extent ineffable. You cannot say that we are in the grip of something independent from us, directly evidenced by simple perceptual situations, for example. You cannot legitimately do such things because you are married to conceptual content, or more generally, to the kind of position that sees even a simplest perception propositionally, that is, in terms of *things being thus and so* (see McDowell, 1996, 9). As already said, such a theoretical move basically reduces the contents of perceptual experience to sentences, or to be more precise, to propositions necessarily *expressed* in sentences. And so, reality more or less becomes a product of verbalization, limited to what can be verbalized or otherwise linguistically expressed. As Wittgenstein famously put it:

---

<sup>1</sup> Alternatively, one might worry in this context about idealism (a worry explicitly brought up by Sellars, for example: see Williams, 2012, 63-4). In the Kantian context, in particular, pushing the conceptualist interpretation of Kant’s theory of intuition too far can be seen as leading to a similar outcome: “In my view, the dominant [conceptualist] reading of intuition [...] is likely to lead to understanding transcendental idealism as centrally a kind of conceptual idealism—a position which is about seeing empirical reality as carved up by or dependent on our concepts.” (Allais, 2015, 148)

<sup>2</sup> The Sellarsian claim-talk (e.g., McDowell, 1998, *passim*) does not change any of this: ultimately, the “claims” contained in perceptual experience are items expressed in sentences.

*The limits of my language mean the limits of my world* (Wittgenstein, 1922, 5.6.).

Surely there are conceptualists who do not want to “deplete” reality that much. Perhaps McDowell is one of them, though that is easier said than done from someone who also thinks that “receptivity does not make an even notionally separable contribution to the co-operation [between receptivity and spontaneity]” (McDowell, 1996, 9). Be that as it may, the kind of conceptualist who thinks that the “deliverances” of sensibility must be guided or “put under” concepts to play any cognitive role whatsoever, can still avoid the “verbalist” or “judgmentalist” version of conceptualism. All they need to do is show that concepts ought to be understood in some other manner than merely as the constituents of judgments that can only find real use by being expressed in the sentences of some language or another.

It seems to me, however, that nobody has ever *concretely* shown what such a concept-use would really mean, or how exactly it is supposed to take place, and particularly so in the Kantian context (see, e.g., Longuenesse, 2000, 45, 50 and *passim*; Allison, 2004, 79 and *passim*). Would it mean that when I saw the ball hit the window, I could only make sense of the situation by *implicitly subsuming*—just like that—the ball, the window, and their clash, under the concepts of Substance and Cause? Then again, what would this even mean if the “judgmentalist” or “verbalist” view is deemed unacceptable—as *too* intellectualistic, say? Would it mean that the implicit idea of an object that exists over time (and has got some sort of powers to affect other objects) is somehow implanted in our sense perceptions, already at the level of sensory intakes perhaps? But how could that be and on what grounds?

Be that as it may, instead of arguing for the de facto conceptuality of sensory inputs or the like, or the conceptuality of appearances as such, let alone the details of the cognitive processes involved, the main conceptualist argument would nevertheless be that for anything originated in sensibility to be judgeable for us, it must be *intrinsically compatible* with concepts—most notably with the categories in the Kantian context. And if that is so, it must then mean (says the conceptualist) that the ball hitting the window taken seemingly merely phenomenally must share the same cognitive structure as the ball hitting the window taken propositionally as an explication of *things being and behaving thus and so*. Therefore, they say, every perceptual experience must possess conceptual content, to which they are eager to add that otherwise the contents of perceptual experience could not enter in rational relations of justification or “the space of reasons” (see, e.g., McDowell, 1996, 4-5).

### 3. Can the Compatibility between Sensibility and Understanding Be Fully Explained?

Let’s look at the intrinsic compatibility view closer. It is based on a kind of transcendental argument that basically says that if the contents of perceptions were not isomorphic with the contents of judgments, then cognitive experience as we know it would not be possible. It is telling that such an

argument is not specifically an explanation of how cognitive experience takes place, and it effectively overlooks the abovementioned reasons to believe that perceptual contact with reality is much more direct and uniquely different from judgments. This makes the argument rather weak and uninformative and certainly not a definitive argument against nonconceptualism. As a matter of fact, the conceptualist compatibility argument reminds us of the refutation of *substance dualism* on the grounds that the mind-body causation would be very hard to understand anyway. While that is certainly true, that alone does not make substance dualism untrue in the slightest.

How about Kant? What kind of an argument he gives for the compatibility view? A good place to look for is the notorious Transcendental Deduction, especially the section 26 of the B-version, where Kant makes a notable move as he explicates his previous claim that “all synthesis, through which even perception itself becomes possible, stands under the categories” (CPR, B161). Kant gives two examples of this “standing under” relation. The first example concerns the perceptual grasping of the spatial form of a house in outer intuition and its abstracted counterpart, namely the category of Quality, which suggests the attribution of size to the house, and, moreover, the “agreement” of these two “synthetic unities” (CPR, B162). The second example concerns the perceptual recognition of water going from a fluid state to a solid state. This time the crucial element is the sequence between the two states, either considered in terms of their relation in time in inner intuition or as the abstraction of that relation in terms of effects and causes (CPR, B162-3). In a word, the category-application is a process of *abstraction* that presumably requires something *concrete* to be abstracted *from*. What else would this concrete thing be but an appearance given in intuition? In addition, take notice of the following: In the first example, the category-application concerns the quantification of the house, but *not* having the intuition or appearance of this particular object as such. Similarly, in the second example, the appearances of water turning from liquid to ice are already available for the perceiver, whereas the category-application concerns how to causally determine the event as a whole.

Kant’s view here looks to me exactly as the one *opposed* by the likes of McDowell, according to whom there cannot be anything like nonconceptual sensory consciousness, or “getting of an extra-conceptual Given”. But doesn’t the appearance—elsewhere explicated by Kant as the “undetermined object of an empirical intuition” (CPR, A20/B34)—play the role of the nonconceptually given? I would say it does. Some of Kant’s conditionals are quite revealing too—he says: “*if I abstract* from the constant form of my inner intuition” and “*if I apply* [the category of cause] to my sensibility” (CPR, B163; my italics). This does not at all suggest that the intuition or appearance itself would have to be “infused” by the category.

Of course, the abstraction process can only succeed if there *is* some kind of compatibility between understanding and sensibility. However, it seems that the Deduction remains rather silent on the question of how the abstraction process exactly takes place and what kind of compatibility



(that guarantees the success of that process) is precisely at issue here (see also Laiho, 2019, 45-6). In a letter to Herz, Kant himself suggests that this is something we cannot really explain:

we are absolutely unable to explain further how it is that a sensible intuition (such as space and time), the form of our sensibility, or such functions of the understanding as those out of which logic develops are possible; nor can we explain why it is that one form agrees with another in forming a possible cognition (C 11, 51).

Given this, the case for Kantian conceptualism remains weak: there is the idea of the necessary compatibility between sensibility and understanding, but no specific explication of their link. Besides, couldn't one suggest nonconceptualism *all the way up* on the same grounds (cf. Stalnaker, 2003, 105-6)? Say, by admitting that understanding and sensibility must operate together at some point of the cognitive process, while emphasizing that the main Kantian point would nevertheless be that their co-operation must be *grounded* in nonconceptual intuitions for that co-operation of theirs to have any significance at *any* point of the cognitive process? Such a strategy would in fact fully avoid the slide towards the conceptual creationist view that there is nothing we have not conceptualized, and the ill consequence that without a constant link to the nonconceptual ground it becomes unclear how always to distinguish the reality of concepts, that is, whether they "actually relate to objects or are mere beings of thought" (*Prol 4*, 295).

A little later in the Deduction, Kant makes another interesting move as he explicates the two examples further by referring to the laws of nature in general, as opposed to particular laws of nature that are always partly empirical (*CPR*, B163-5). These laws of possible experience, as Kant calls them in the *Prolegomena*, are "prescribed" by space and time "in combination with" the categories (*Prol 4*, 375). Yet, and this I find worthy of emphasis, these laws are *very general* in character. The category of Cause, for example, implies "the law, that if an event is perceived then it is always referred to something preceding from which it follows according to a universal rule" (*Prol 4*, 296). This in turn implies the more general view that nature as a whole must comply with such principles, since otherwise it would not be possible for us to know anything necessary and *a priori* about it. But what this more general view does not imply is that the appearances themselves must be governed or "infused" by the categories to count as appearing objects in the first place.

Of course, one might still argue that in order to count as fully-fledged objects, understood as epistemic items self-consciously available to the cognitive subject, the appearances need to be combined with the categories (see, e.g., Allais, 2009, 405). A textually based variation of this argument would go that mere representations need to be "related to an object", understood as "that in the concept of which the manifold of given intuition is united" (*CPR*, B137). This, in turn, points towards the categories, and "the unity of consciousness" (*CPR*, B136-7) constituted by them, which is supposed to secure how every intuition, and the manifold contained in it, is thinkable (see *CPR*,

B138). To my mind, however, the introduction of the one consciousness acts as just another compatibility argument, which leaves unanswered the question of how exactly the unity of the categories in one consciousness is related to the appearances in actual cognition. In support of this reading, I would like to pinpoint Kant's remark earlier in the Deduction (§21):

In the above proof, however, I still could not abstract from one point, namely, from the fact that the manifold for intuition must already be given prior to the synthesis of the understanding and independently from it; how, however, is here left undetermined (*CPR*, B145).

Put in general terms, while Kant clearly wants to keep intact the independency of intuition from the operations of understanding, he also leaves much unexplained in the Deduction. Nor do I see any further proof forthcoming in that piece of text. Moreover, it appears that it is the Schematism chapter that follows the Deduction—not the Deduction itself—that is supposed to give us insight into the question of *how* sensibility and understanding come together.

#### 4. What Is Schematism Supposed to Explain?

We already touched upon the idea of the implicit subsumption of appearances under concepts, where the concepts are to be understood more like rules of synthesizing sense perceptions as constituents of actual judgments. Such a view would have more explanatory power regarding the co-operation between sensibility and understanding, in comparison to the explanatorily weak compatibility view. Such a view would also avoid the “verbalist” or “judgmentalist” version of conceptualism. At the same time, it is difficult to see what exactly would actual, albeit implicit, concept-use that already takes place at the level of appearances amount to, and how it would relate to explicit concept-use in actual judgments. What's more, the idea appears to stem from contemporary psychological studies, such as those that focus on the continuity of objects. The results of these studies suggest, among other things, that even very young children do not expect that a perceived object that has moved behind an obstacle has all of a sudden disappeared; instead, they expect the same object to soon reappear (see, e.g., Bertenthal & al. 2007). I think one could use the concept of substance in this or similar context to refer to a built-in cognitive mechanism thanks to which the test subjects anticipate perceived objects to exist over time even unseen.

But can we really find a similar idea in Kant? A good place to look for is another obscure piece of text: The Schematism chapter, which is supposed to explain how the heterogeneous elements of sensibility and understanding are brought together, that is, “how [...] the application of the category to appearances [is] possible” (*CPR*, A137/B176). More specifically, the Schematism chapter should clarify the *subsumption* of intuitions under the categories (*CPR*, A137/B176). That is to say, it should illuminate the process of regarding the intuited objects in accordance with one or the other of the categories, or as belonging under them, in which process the schemata are supposed to function as mediating representations (*CPR*, A138/B177). Unfortunately, Kant is sparse on details,

and gives better examples of geometrical and empirical concepts, and their respective schemata, though his ultimate point supposedly concerns the categories and their transcendental schemata. It almost seems as if he was in a rush when writing that section, or that providing a clear-cut answer to the question about the exact relationship between sensibility and understanding was never his primary concern (see esp. *CPR*, A142/B182). As we have seen, the same might be said about the Deduction.

What should be clear, though, is that the appearances and the categories never enter into a containment relation—he says:

No one would say that the category [...] is contained in the appearance”  
(*CPR*, A137-8/B176-7).

What’s more, the categories appear in the Schematism chapter as some sort of rules of determination that are *laid onto* the intuited appearances with the help of the schemata, as opposed to something that are *put into* the intuitions themselves. Read this way, the strict distinction between sensibility and understanding can be kept intact, and one can also continue to stress that the appearances are given independently of the intellect. In addition, one can see why the schemata are needed in the first place: namely, because the intuited objects or appearances never become conceptual as such.

It is noteworthy that with the categories and their schemata, Kant emphasizes *time*. Here’s a rough sketch of what I think he means by this: The application of the categories depends on temporal conditions—in the case of Substance, on “the persistence of the real in time”, which is also its schema (*CPR*, A144/B183). Such a rule can be laid onto an appearance if there indeed is something that persists over time. The mere category cannot do the trick because it is supposed to be totally different in kind from the appearance given through sensibility. In addition, as Kant puts it,

if one leaves out the sensible determination of persistence, substance would signify nothing more than a something that can be thought as a subject (without being a predicate of something else)” (*CPR*, A147/B186).<sup>1</sup>

In other words, the concept would be limited to a merely logical use. With the schema in place, on the other hand, the concept of Substance can be extended on the appearance. For this, imagination is required, since imagination apparently can do the trick because in its time-

---

<sup>1</sup> In the Schematism, Kant also reminds the reader that the goal of the Deduction was to establish that the categories “relate *a priori* solely to appearances” and “cannot pertain to things in themselves” (*CPR*, A139/B178). We see here again how Kant’s approach, here described in his own words, appears more general than the kind of approach that would aim to explain how the appearances become “infused” by the categories.

determinations it can employ both sources: time, understood as the intuition of inner sense, and the rules of synthesis expressed by the categories, which always have a temporal dimension to them, thus sharing the same *a priori* ground as their basis (*CPR*, A138-9/B177-8).

Kant's explication of the role of imagination leaves much to be desired. Basically, he appears to make just another indirect compatibility argument when he postulates a mediating faculty between sensibility and understanding—as he already did when he introduced the schemata as mediating representations. Yet he succeeds, I think, in showing how only under suitable temporal conditions something that appears can be *declared* to be a substance, for example.<sup>1</sup> If this so, maybe the Schematism should not be read as an analysis of implicit concept-use at the level of intuitions as such, but as a more general or higher level explication of how appearances can be regarded as instances for category-application, *when* a category is applied to sensibility (cf. *CPR*, B163). This, however, appears to be a matter of judgment. Not that this would be surprising given that the Schematism is a part of the section titled Transcendental Doctrine of the Power of Judgment.

## Conclusion

I suggested in the last two sections that the Deduction is not exactly an explanation of how the co-operation of sensibility and understanding takes place, and that the Schematism too leaves much to be desired in this respect. What the Deduction shows at best is that sensibility and understanding, though separate and distinct, must be compatible in their operations. As such, however, the compatibility argument does not show preference for Kantian conceptualism over Kantian nonconceptualism. Earlier, I suggested that not only does Kantian nonconceptualism better appreciate the uniqueness of the two faculties and their different cognitive contributions, but helps us keep with Kant's partly realist tendencies that disfavor any such position that might be called conceptual or linguistic creationism. Taken together, these points support, even if only indirectly, the kind of Kantianism that emphasizes the primacy of sensibility. In fact, we could here follow the arch-conceptualist McDowell himself and insist that we do not even “[need] to give an account of how concepts and intuitions are brought into alignment” (McDowell, 2002, 457). Indeed, why not accept that we do not really need to—perhaps because we are unable to—fully understand the co-operation between sensibility and understanding? It is just that this argument works in favor of nonconceptualism too.<sup>2</sup>

---

<sup>1</sup> The term “declare” is borrowed from Kant himself: “If I declare [*erkläre*] a thing to be a substance in appearance, predicates of its intuition must be given to me previously” (*CPR*, A399).

<sup>2</sup> This essay is a slightly modified and extended version of my presentation at the *6th Annual Workshop of the Contemporary Kantian Philosophy Project*, “Kant and the Primacy of Sensibility”, held at the Washington University in St. Louis, May 19-20, 2023. I thank the participants for fruitful philosophical discussion and helpful comments.

## References

- Allais, L. (2009). Kant, Non-Conceptual Content and the Representation of Space. *Journal of the History of Philosophy* 47 (3), 383-413. <http://dx.doi.org/10.1353/hph.0.0134>
- Allais, L. (2015). *Manifest Reality. Kant's Idealism and His Realism*. Oxford University Press.
- Allais, L. (2016). Conceptualism and Nonconceptualism in Kant: A Survey of the Recent Debate. In D. *Kantian Nonconceptualism*. pp. 1-25, Edited by Schulting, Palgrave.
- Allison, H. E. (2004). *Kant's Transcendental Idealism. An Interpretation and Defense*. Revised and Enlarged Edition. Yale University Press.
- Bertenthal, B & al. (2007). Phenomenal permanence and the development of predictive tracking in infancy. *Child Dev.* 78 (1), 350-63. <http://dx.doi.org/10.1111/j.1467-8624.2007.01002.x>
- Butchvarov, P. (2002). Metaphysical Realism and Logical Nonrealism. In *The Blackwell Guide to Metaphysics*. pp. 282-302, Edited by R. M. Gale, Blackwell,
- Goodman, N. (1978). *Ways of Worldmaking*. Hackett.
- Hanna, R. (2015). *Cognition, Content, and the A Priori: A Study in the Philosophy of Mind and Knowledge*. Oxford University Press.
- Kant, I. (1900-). *Kant's gesammelte Schriften*. Edited by D. Wissenschaften, Reimer/de Gruyter.
- Kant, I. (1999). *Correspondence*, translated & Edited by A. Zweig, Cambridge University Press.
- Kant, I. (2000). *Critique of Pure Reason*. translated & Edited by P. Guyer & A. Wood, Cambridge University Press. (CPR A/B)
- Kant, I. (2002). Prolegomena to Any Future Metaphysics That Will Be Able to Come Forward as Science. In *Theoretical Philosophy after 1781*, translated by G. Hatfield, Edited by H. Allison & P. Heath, Cambridge University Press.
- Laiho, H. (2019). Types of Representational Content in Kant." *Kantian Journal* 38 (1), 31-54.
- Longuenesse, B. (2000). *Kant and the Capacity to Judge. Sensibility and Discursivity in the Transcendental Analytic of the Critique of Pure Reason*. Charles, translated by T. Wolfe, Princeton University Press. (Original publication 1993)
- McDowell, J. (1996). *Mind and World*. With a New Introduction by the Author. Harvard University Press. (Original publication 1994)
- McDowell, J. (1998). The Woodbridge Lectures 1997: Having the World in View: Sellars, Kant and Intentionality. *The Journal of Philosophy*, 95 (9), 431-491.
- McDowell, J. (2002). The Content of Perceptual Experience. In *Vision and Mind. Selected Readings in the Philosophy of Perception*, pp. 443-458, Edited by A. Noë & E. Thompson, The MIT Press.
- McLear, C. (2020). Kantian Conceptualism/Nonconceptualism. In *The Stanford Encyclopedia of Philosophy* Edited by E. N. Zalta. <https://plato.stanford.edu/archives/fall2021/entries/kant-onconceptualism>
- Stalnaker, R. (2003). What Might Nonconceptual Content Be? In *Essays on Nonconceptual Content*. pp. 95-106, Edited by Y. H. Gunther, The MIT Press. (Original publication 1998)

- Williams, J. J. (2012). How Conceptually Guided Are Kantian Intuitions? *History of Philosophy Quarterly* 29 (1), 57-78. <https://doi.org/10.36311/2318-0501.2021.v9n1.p97>
- Wittgenstein, L. (1922). *Tractatus Logico-Philosophicus*. Edited by C. K. Ogden & F. P. Ramsey, Kegan Paul, Trench, Trubner & Co.

## Kant, Niels Bohr and Quantum Spontaneity

Willem McCloud 

Independent Scholar, South Africa. Email: [wmcloud@yahoo.com](mailto:wmcloud@yahoo.com)

| Article Info                                                                                                                                                                                                                                                                                    | ABSTRACT                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p><b>Article type:</b><br/>Research Article</p> <p><b>Article history:</b><br/>Received 15 July 2024<br/>Received in revised form 18 July 2024<br/>Accepted 18 July 2024<br/>Published online 07 August 2024</p> <p><b>Keywords:</b><br/>quantum physics,<br/>spontaneity, Bohr &amp; Kant</p> | <p>This paper focuses on absolute spontaneity, first postulated by Immanuel Kant. In the early twentieth century spontaneity entered the domain of quantum physics when Niels Bohr included it as part of his quantum postulate. Later on, David Bohm developed the concept of a quantum potential in his description and interpretation of quantum physics, a concept that can also be understood in terms of spontaneity. A discussion of Kant's influence on the interpretation of quantum physics is followed by a consideration of the inclusion of spontaneity in Niels Bohr's epistemological approach to quantum mechanics and David Bohm's quantum potential as part of his ontological approach to quantum physics. Kant's influence on both Bohr and Bohm is examined as well as the applicability of his critical metaphysics to quantum theory. Critical metaphysics read together with Bohr and Bohm's interpretations of quantum physics is then utilized to make some proposals with regard to the problems plaguing our best theories of physics.</p> |
| <p><b>Cite this article:</b> McCloud, W. (2024). Kant, Niels Bohr and Quantum Spontaneity. <i>Journal of Philosophical Investigations</i>, 18(47), 63-88. <a href="https://doi.org/10.22034/jpiut.2024.62515.3820">https://doi.org/10.22034/jpiut.2024.62515.3820</a></p>                       |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|                                                                                                                                                                                                              | <p>© The Author(s). Publisher: University of Tabriz.</p> <p><a href="https://doi.org/10.22034/jpiut.2024.62515.3820">https://doi.org/10.22034/jpiut.2024.62515.3820</a></p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |

## Introduction

The 300-year anniversary of Kant's birthday in 1724 coincides with the commemoration of an important discovery made by Niels Bohr 100 years ago in 1924 in his efforts to understand quantum mechanics, namely that the causality which governs all reactions in the classical world, breaks down in quantum physics. This discovery was a direct result of the earlier discovery in 1900 by Max Planck that energy comes in packets called quanta, which also gave its name to quantum physics. These discoveries for the first-time confirmed Kant's idea that absolute spontaneity could very well be part and parcel of our world.

The implications of spontaneity being an essential part of our world is profound and perhaps one of the greatest and most unexpected discoveries of all time in the natural sciences. It went directly against the generally accepted view of modern times that all interactions in the world are deterministically determined in causal terms. This discovery also underlies one of the great problems on which physicists to this day have not reached consensus, namely the measurement problem.

In this paper I take a closer look at spontaneity as featuring in quantum physics, descriptions thereof in quantum theories as well as the possible application of Kant's critical metaphysics in which spontaneity plays an important part to quantum physics. I also look into the origins of spontaneous reactions, a problem that goes right down to the most basic and fundamental forces known to us. In this analysis Kant's critical metaphysics is utilized in an effort to provide insights and even clues for solving other problems in physics, problems like the apparent incompatibility between quantum physics and general relativity and even those pertaining to consciousness.

### 1. The Inclusion of Spontaneity in Quantum Theory

In 1924 Bohr discovered that in quantum theory a *causal* connection between electronic motion and radiation cannot be exhibited in *space and time* (Pringe, 2009). As a result of this insight, Bohr later on in 1928 published his so-called quantum postulate, a theoretical generalization based on an empirical assumption, in which this discontinuous or spontaneous process resulting from the indivisibility of the quantum is a central feature:

[Quantum theory's] essence may be expressed in the so-called quantum postulate, which attributes to any atomic process an essential discontinuity, or rather individuality, completely foreign to the classical theories and symbolised by Planck's quantum of action. (Bohr, 1928, 580)

The quantum postulate "means that in the case of all atomic processes every energy change result from an indivisible – and because of that discontinuous – transition between different states that cannot be continuously connected" (Valente, 2010, 1). As a consequence of this discontinuous



nature of quantum physics, all quantum outcomes are always described in probabilistic terms. Quantum physics is intrinsically probabilistic, that is, non-causal (Bitbol and Osnaghi, 2013,160).

Also in 1924, Louis *de Broglie* proposed a new speculative hypothesis that electrons and other particles of matter can behave like waves, which was confirmed in 1926 by the electron diffraction experiments of G. P. Thomson. During that time, physicists were able to produce the first formulation of the basic principles of quantum physics. In 1925 Werner Heisenberg, Max Born and Pascual Jordan formulated a description of quantum mechanics called matrix mechanics and in 1926 Erwin Schrödinger invented wave mechanics with his formulation of the non-relativistic Schrödinger equation. Schrödinger afterwards showed that these two approaches are equivalent. Then, in 1927, Heisenberg formulated the uncertainty principle.

## 2. Niels Bohr and Kantian Philosophy

When Bohr made this discovery that causality breaks down in quantum systems, an immediate concern was that this may be in conflict with Kant's transcendental philosophy, which has provided the epistemological grounding for Newtonian physics. As Léna Soler writes:

At the time, it was more and more said, here and there in philosophical and scientific circles, that the (then just born) quantum mechanics refuted Kantian philosophy, especially the Kantian table of categories and its concept of causality. (Soler, 2009, 330)

In Niels Bohr's formulation of quantum theory, he, however, did use Kantian philosophy to ground the epistemology of this new field of studies. In general, philosophers of science are in agreement that Bohr made use of Kantian philosophy but the extent and the manner in which he did so are not agreed upon. Already in the nineteen thirties did Neo-Kantian scholars like E. Cassirer, G. Hermann and C. F. von Weizsäcker recognized a Kantian influence in Bohr's approach and many other philosophers of science also discuss different parallels, sometimes holding opposing views in this regard (Kauark-Leite 2017; Bitbol, 2017, 2013; Cuffaro, 2010; Pringe, 2009; Chevalley 1994; Kaiser, 1992; Faye 1991; Honner, 1987, 1982; Fano, 1988 & Folse 1985). Whereas most of them base their views on Kant's *Critique of Pure Reason*, Pringe uses the *Critique of the Power of Judgment* in his approach.

Michael Cuffaro (2010, 1, 25) holds that any proper "interpretation of Bohr should start with Kant" and "a Kantian (who does not deny the validity of the uncertainty relations), starting from the principles of Kantian philosophy, would be led to many of the same conclusions as Bohr" (Cuffaro, 2010, 25). Kauark-Leite (2010, 250) holds the view that Bohr "does nothing but extend the Kantian analysis to a totally different epistemic situation." Pringe (2023, 249) emphasizes that the discontinuity of atomic processes enables us to establish a remarkable connection between transcendental philosophy and Bohr's interpretation of quantum theory. And he shows (Pringe,

2007, 2023) how Bohr's view can not only be understood in terms of classical Kantian (critical) philosophy but also provides a detailed exposé as to how quantum observations can be viewed in objective epistemic terms in the framework of that philosophy.

### **3. The Copenhagen Interpretation and Its Demise**

#### **3.1. Bohr's Interpretation of Quantum Mechanics**

As a consequence of the spontaneous collapse of quantum states to reduced states, the observer has no way to gain direct empirical access to quantum objects. They only have access to quantum phenomena and that through their measurement apparatuses. In Bohr's view, quantum objects fall outside the possibility of empirical access and we can thus not obtain any real knowledge about them, except for the fact that they exist. He did not attribute any intrinsic and measurement-independent state properties to atomic objects in addition to the classical ones being manifested in measurement (Faye, 1991).

In Bohr's view the quantum mechanical formulation is thus not true in the sense that it gives a literal representation of the atomic world; it only provides a symbolic representation (waves and particles serve as symbols pertaining to quantum objects). Bohr is therefore called an "entity realist" who opposes "theory realism" (Folse, 1986; Faye 1991).

Bohr's interpretation of quantum physics, with the added insights and contributions of Heisenberg, Born and others, later on became known as the Copenhagen interpretation, despite the fact that Bohr and Heisenberg never fully agreed on the way to understand the mathematical formulation of quantum mechanics. "Today the Copenhagen interpretation is mostly regarded as synonymous with indeterminism, Bohr's correspondence principle, Born's statistical interpretation of the wave function, and Bohr's complementarity interpretation of certain atomic phenomena" (Faye 2019). The heyday of the Copenhagen interpretation spanned approximately the period from 1927 to 1952.

#### **3.2. The Measurement Problem**

Some interpreters of Bohr, like Don Howard (1994, 2004, 2005), are of the opinion that he believed that the superposition of quantum states is entangled with the measuring apparatus. This view is not generally shared and most interpreters seem to disagree.

Mario Valente (2010, 7), for example, emphasizes that, according to Bohr, not all of the experimental arrangement has to be considered in direct interaction with the quantum system, only the "significant parts of the experimental arrangement" (Bohr, 1962, 92). This explains why the experimental arrangement with its entirely classically described "fixed measuring rods and synchronized clocks" (Bohr, 1955, 90) stands "outside and independent of the object under consideration" (Bohr, 1985, 369). Valente writes: "This reading of Bohr's ideas implies considering the part of the experimental arrangement not directly in interaction with the quantum system as

describable by classical physics.” After a discussion of the views of Paul Teller, Henrik Zinkernagel and Simon Sanders, Valente comes to the conclusion that they all understand

Bohr as implying the need for a classical physics account of (at least) part of the experimental arrangement (the one not directly in interaction with the quantum system), stressing in particular the need for a classical account of the reference frame. This view is clearly at odds with Howard’s. Howard’s reconstruction implies an all-quantum description of the entangled pair instrument & object, giving just a classical description (in Howard’s sense), for both the instrument and object, of the property being measured. (Valente, 2010, 13)

In 1932 John von Neumann produced a formulation of quantum physics in his *Mathematical Foundations of Quantum Mechanics* that has often been taken as for the most part being in agreement with the Copenhagen interpretation. In this formulation Von Neumann argues that quantum mechanics can only be described in terms of the Schrödinger equation; in contrast, the measurement process cannot be described by quantum physics. He coined the term “projection postulate” for the reduction of the wave function when a measurement is taken.

In Von Neumann’s formulation the quantum object and the measurement apparatus are in an entangled state. When a measurement is taken, the superposition of states collapses to determinate states. As already shown, this view does not agree with Bohr’s view (as it is usually understood) that the quantum object is not in a superposition with the measurement apparatus; all atomic processes are discontinuous (spontaneous) and the resulting quantum phenomena stand in relation with the quantum apparatus (as described in the quantum postulate; see Valente 2010, 5 for a detailed discussion). A “cut” is made between the object-system and the experimental context (Bitbol and Osnaghi, 2013, 154). In this case no question arises as to why the superposition collapses.

In Von Neumann’s formulation, he stands in need of finding a cause for the collapse which is intrinsically linked with the act of measurement, a problem that does not occur in the Bohr view. In fact, various scholars like Nancy Cartwright (1983) and R. I. G. Hughes (1989) have argued that the reduction of the wave packet should not be understood in terms of measurement; it also happens in other situations such as atomic decay. Von Neumann then postulates that the superposition of states extends to include the actual observer, whose perception (consciousness) somehow causes the collapse.

The extending of the superposition to include macro structures in the classical world was criticized by Schrödinger in his famous thought experiment about the cat being both alive and dead

until an observation is made. This image is often presented in all seriousness as being Schrödinger's view whereas he actually uses it to show how Von Neumann misinterpret the theory to include cat-sized objects like the human observer (the cat is either dead or alive but not both). No evidence exists that quantum features apply to the classical world on a macro level (quantum features do appear on a cellular level as shown in quantum biology but that more or less defines the limits of its application).

Even though Bohr's interpretation is not saddled with Von Neumann's measurement problem, it is nonetheless true that in quantum theory the operational notion of continuous time described by the Schrödinger equation stands in glaring contrast with the discontinuity described in Bohr's quantum postulate and the corresponding statistics of outcomes given by the Born Rule.

As a consequence, the measurement problem is usually defined slightly differently, namely that the two kinds of evolution associated with the quantum mechanical description, that is 1) the deterministic evolution in accordance with the Schrödinger equation (that should be generally valid) and 2) the projection postulate (that should be deductible from the dynamic description), are flatly in contradiction with each other (Albert, 1992, 37). The theory of quantum mechanics can apparently not be fully described in one unified mathematical description.

### **3.3. The Gap Between the Classical and Quantum Worlds**

Since the time when quantum theory was first formulated physicists were struck by the strange difference between the classical and quantum descriptions of the world, with quantum entities demonstrating behavior that has never before been observed in the classical world (like superpositions of states, entanglement and non-locality). This difference was in part captured by Bohr's quantum postulate that describes the spontaneous nature of atomic events as a consequence of the “indivisibility of the quantum of action” (i.e. quanta; Bohr, 1934, 5) that stands in dramatic contrast with the classical world in which only deterministic causality is known.

Many authors have discussed this “gap”. For Bohr, the difference is self-evident and differentiates where predicates can be legitimately used and where not. Bitbol writes that “we are faced with a persistent dialectic between two irreducible domains of discourse (objectified and situated)” (Bitbol, 2007, 258; see Hughes, 1989, 312, 316) In the Copenhagen description of quantum mechanics this primarily concerns the difference between the superposition of states and reduced states (a transition process described by the projection postulate). Various other interpretations have tried to overcome this divide by rejecting the projection postulate (believing that the collapse of the wave packet is not a real event).

Among these views are David Bohm's (and De Broglie's) mechanics that uses a guiding equation to define the positions of the particles (or configurations of fields) described by the wave function, Hugh Everett's “relative state” formulation according to which subsystems “branch off”

from the state vector of the universe, and Bas van Fraassen's modal formulation (called constructive empiricism) according to which the quantum state delimits what is possible whereas measured properties say what is actual (for discussions of these views, see Healey, 2009, 274, Hughes, 1989, 311, and Earman, 1986, 223). Another theory is that of Giancarlo Ghirardi, Alberto Rimini and Tullio Weber, which modifies the Schrödinger equation by adding stochastic and non-linear effects. Different variations of these theories have been proposed through the years.

These different views introduce new problems of their own, for example “action-at-a-distance” (Bohmian mechanics), many worlds (Everett’s view) and an instrumentalist view (Van Fraassen’s approach; for those who find it problematic). The Ghirardi-Rimini-Weber (G-R-W) theory turns the Schrödinger equation into a scholastic dynamical law in which the instantaneous (often called “spontaneous”) collapses of the system happen as a consequence of background perturbations (Allori et al., 2008, 357; with the “shifty split” being done away with, Bell, 1990).

This introduces the question as to the nature of the spontaneity ascribed to the quantum world. In contrast with the G-R-W theory, Cartwright (1983) argues that Bohr holds to the view that the transition between states is absolutely spontaneous (i.e. no stochastic process in keeping with deterministic causality underlies such motions). When a quantum system makes a spontaneous transition from one state to another, some conserved quantity like energy, momentum and angular momentum will be emitted or absorbed; the exchange of energy activates the detector and some observable quantity is measured (Cartwright, 1983, 179). Michael Redhead (1987) has indeed shown on mathematical grounds based on Bell’s theorem that in the case of the Aspect experiment, the collapse is truly spontaneous without any possible stochastic form of determinism involved. And recently the Majorana Demonstrator experiment performed at the Sanford Underground Research Facility has effectively shown the G-R-W theory to be wrong (Donadi et al. 2020).

All the aforementioned efforts try to obtain a unified (mathematical) description of quantum physics. What is important to note is that the shift away from the Copenhagen interpretation at the same time implies a rejection of Kant’s transcendental approach that underlies Bohr’s view (Bitbol and Osnaghi 2013, 154). From this it seems to follow that the utility of a Kantian approach to quantum physics and even science more generally is closely linked to the prospects of Bohr’s view. We will return to this assessment presently.

#### **4. The Metaphysics of Quantum Physics**

At this point the inevitable question needs to be asked, namely why Bohr’s view and the Copenhagen interpretation of quantum mechanics lost its appeal. This is a very important question that all Kantians have to consider seriously because it also engages with the question about the usefulness and benefit of Kantian philosophy of science within current scientific debate, especially concerning quantum physics.

Even though this is not an easy question to answer, I would like to make some suggestions. There are mainly two reasons for the demise of the Copenhagen interpretation, namely that the physics community is not only interested in quantum mechanics and quantum theory more generally, as a successful theory, they are also interested in the bigger questions about the nature of our universe. They look for metaphysical insights that can *explain* things. As a consequence, it is not strange to find discussions centering on the “metaphysics of quantum mechanics”, with the different views basically exploring different metaphysical options pertaining to the true nature of the world.

Another reason is that physicists put a lot of trust in mathematics and often express a belief in mathematical theories that have a certain appeal (such as its “beauty” or simplicity) despite the fact that no substantial empirical evidence has been found to support them. A good example is string theory or the multiverse. As for quantum mechanics, the impulse is to take the Schrödinger equation seriously as a description of what really happens. And this introduces the main problem: the Schrödinger equation describes deterministic motion, with the projection postulate simply an add-on that is not mathematically supported except for the Born Rule. The natural conclusion then seems to be that mathematics cannot be wrong and our world is really deterministic despite Bohr’s quantum postulate and his refusal to accept a literal view of the Schrödinger equation.

When all is considered, it seems that the primary problem comes down to what we make of the gap between the classical and quantum descriptions of the world. A central question concerns the ontological status of this gap: is it real or is it simply a gap in our understanding of the workings of the universe. And from a Kantian perspective it needs to be asked what contribution Kant’s philosophy can make towards solving this problem.

The question about the ontological implications of Bohr’s view has indeed been asked. Mauro Dorato (2017) argues that Bohr’s view does imply some form of ontological distinction between the classical and quantum realms. Scholars who regard Bohr as believing that the measuring apparatus and quantum object are in a state of entanglement (and that this entanglement involves other large objects from the classical world) can in this way overcome this distinction. But nothing prohibits those who do not hold this latter view from considering the possibility of such an ontological distinction in all seriousness and thus endeavor to define the cut between these realms in a non-ambiguous way.

#### **4.1. Towards a Kantian Solution**

As Kant already introduces absolute spontaneity in his critical metaphysics in the *Critique of Pure Reason* (First *Critique*; edition A 1781, edition B 1789), we can also take a closer look at his formulation thereof. In accordance with Kant’s epistemology having been applied to the quantum system, as follows from Bohr’s interpretation and other detailed formulations of quantum epistemology in keeping with the Bohrian approach (Pringe, 2007, 2023), it could very well be that

Kant's broader critical metaphysics also finds an application in the world described by quantum physics.

Even though Kant establishes the limits of pure reason and undoes all "proud ontologies" (A247; and many Kantian philosophers as a consequence stay clear of metaphysics), nothing in his philosophy prohibits the correct use of metaphysics. Contra Bitbol (2008), philosophers like Willem McLoud (2018), Stephen Palmquist (2013) and Robert Hanna (2006) argue that Kant's realm of regulative ideas allows for the construction of a systematic critical metaphysics that includes ontological distinctions (see also Allais, 2004, Langton, 1998 and Ameriks, 1992 for noumenal/intrinsic properties) which can serve as a hypothesis pertaining to the natural workings of nature. It is especially in the *Critique of the Power of Judgment* (Third Critique; KdU 1790) that Kant produces a detailed and systematic metaphysics when he considers the natural products of nature.

Let us for a moment consider a broad outline of Kant's critical (regular) metaphysics and how quantum physics can be accommodated within it. This will also show how ontological distinctions enter his metaphysical picture. Needless to say, even though this branch of Kant's philosophy that he first develops in the second part of the First Critique follows logically and systematically from the epistemology formulated earlier in this Critique and which applies to objects of experience (and experiment) in the classical word (for the most part described by Newton's laws), it nonetheless also stands apart and is quite distinct from that.

Kant introduces the basic features of his critical metaphysics in the context of the third antinomy (conflict of laws), one of two dynamical antimonies concerned with the *consideration of existence alone* (i.e. no magnitudes of the series of conditions are considered; A536/B564). The transcendental and empirical use of reason (i.e. the thesis and anti-thesis positions) pertaining to the unconditioned in a series of conditions can be extended *to conceptualize* both an intelligible (noumenal) world and a sensible world, with the latter not taken as an accumulative world-whole but as "nature". As these two ideas concern different modes of existence (Allison, 2012, 17) it follows immediately that the Kantian concepts of "nature" (to be distinguished from our current understanding of nature) and the noumenal realm are ontologically distinct from each other.

The idea/concept of "nature" that Kant introduces does not refer to material nature (as an aggregate) but rather to nature as a systemic whole (A419/B447). All regulative ideas pertaining to world-concepts (i.e. systemic nature and the noumenal world) are "transcendent" to our experience (we can never experience them), with systemic nature taken as a total system of existence (A420/B448) that refers to the world of appearances. As a system, nature is comprised of the totality of causal relations, called mechanism (see Bxxvii-xxx, A419/B447 n., and KdU 5, 379). Mechanism is a regulative concept that belongs to the concept of systemic nature and should be clearly distinguished from the causality of the second analogy that is a rule of the understanding

which applies to phenomena (Allison, 2012, 202-2033). Both are, however, deterministic concepts of causality.

This brings us to the noumenal world. As follows directly from the fact that the third antinomy is concerned with modes of existence and as Kant also explicitly mentions (A420/B448; McLoud 2018), the realm of noumena “oversteps” the sensible world (nature) *in kind*, that is, it refers to another kind of existence apart from nature. What is more is that the transcendental idea of freedom, which follows directly from the thesis position of the third antimony (the *thought* of the unconditioned suggests absolute spontaneity as an effective cause; Allison, 1990, 24), can without contradiction be ascribed to the noumenal realm (which lies “outside” nature where mechanism rules).

The noumenal realm thus becomes the realm of freedom, often contrasted by Kant with nature. Kant views this absolute spontaneity as the only alternative kind of causality (A532/B560). McLoud (2018) argues that Kant’s motivation for introducing noumena in his metaphysical schema is primarily to “save freedom” (see A537/B565). Whereas nature is a system of existence regulated by mechanism, the noumenal realm is a system of existence regulated by transcendental freedom.

In the dynamical antinomies, the anti-thesis position *thinks* the unconditioned in the framework of sensible conditions whereas the thesis position does not merely think the unconditional; it also *thinks it outside sensible conditions in an intelligible (noumenal) world*. Instead of creating a conflict of reason, this represents a *real (logical) possibility*. The outcome of the third antinomy, when viewed in this manner (with the thesis position taken outside possible sensible conditions), is that *existence in two different worlds can be brought into interaction with each other*, i.e., an intelligible (spontaneous) cause can produce phenomenal effects.

Absolute spontaneity as an effective cause of phenomena *has no previous causal links in the structure of causal relations* governed by the second analogy (in the world of phenomena) or even in nature (governed by mechanism); this is why it is viewed as beginning “a series of occurrences *entirely from itself*” (A534/B562, italics in the original). Kant says:

Accordingly, a causality must be assumed through which something happens without its cause being further determined by another previous cause, i.e., an absolute causal spontaneity beginning from itself. (A446/B474, boldfacing in the original)

As such the grounds for transcendental freedom (that is, noumena) stand forever outside the causal chain of events that constitute nature; it can, however, without contradiction produce outcomes which interact with the causal chains of the phenomenal world. As such it is not “incompatible” with nature (A558/B586).



We can now apply this metaphysical system to the world described by science. The Kantian concept of “nature” (regulated by mechanism) allows for the *concepts* of space, time and causality to be used in mathematical theories pertaining to systemic nature (see also Kauark-Leite 2010, 248). In this regard Mauro Dorato (2002) shows that time and causality as conditions of experience can be related to concepts thereof in the framework of space-time theories like Einstein’s theories of relativity.

When it comes to the noumenal realm, numerous scholars have made use of this idea in their efforts to describe quantum mechanics in Kantian terms. Using mere methodological approaches, these scholars typically describe quantum objects in terms of noumena (or “things-in-themselves”) because they are not objects of possible experience (Cuffaro, 2010, 16; Pringe, 2007, 157, n. 31; Fano, 1988). This means that the quantum world is being associated with Kant’s noumenal world. Some authors have moreover proposed that the reduction of the wave packet should be viewed in causal terms (see Cartwright, 1983, 182; 1989, 249; Bartels 1999, S170 and Pringe 2007), with Pringe even using the Kantian concept of an absolute spontaneous cause (as already described) to describe what he calls quantum causality.

Using a weak ontological approach, McLoud (2018) argues in his monograph, titled “Kant, Noumena and Quantum Physics,” that all the conditions for the Kantian noumenal realm (corresponding with Kant’s characterization of that realm in the First and Third *Critiques*, which he views as consistent with each other) are satisfied in the quantum realm as described in quantum field theory (which unites quantum mechanics with special relativity), i.e. with quantum entities existing “outside” proper space, time and causality. As a consequence, he applies the ontological distinction that Kant makes between entities belonging to nature and the noumenal realm to the classical and quantum realms respectively.

McLoud (2018) moreover argues that the Kantian conditions for the possibility of an effective absolute spontaneous causality are also satisfied. He identifies this with quantum causality.

When the gap between the classical and quantum worlds is considered in terms of Kant’s critical metaphysics, it immediately follows why these can be regarded as an ontological distinction. Bitbol and Osnaghi (2013), however, do not agree with this assessment:

Bohr's prescription in no way presupposes or implies an ontological distinction between macroscopic and microscopic systems. There is nothing in the physical nature of macroscopic objects that distinguishes them from the microscopic ones, and which rules out the possibility of describing them as quantum systems. (Bitbol and Osnaghi, 2013, 152)

It seems that Bitbol and Osnaghi (2013) argues that since everything in nature is built from elementary particles, they share the same “physical nature”. The problem is, however, that quantum

entities in the pre-measurement stage cannot be observed as matter. As Bitbol (2007) shows in another paper, it is only quantum phenomena manifest by impacts, bubble chamber tracks and clicks on counters that appear in space-time that can be described as matter. Quantum objects cannot be observed and does not appear in proper space-time.

When the entities in the quantum field description are considered, they cannot even be described as “particles” as noted by Richard Healey: “For a quantum field theory removes even the basic particle ontology, while leaving it quite unclear what is to replace it.” (Healey, 2009, 221)

If an ontological gap between the classical and quantum realms really exists, all efforts to ignore and overcome this gap and to negate absolute spontaneity as a real causality found in quantum mechanics will come to nothing.

#### **4.2. Kant’s “Productive Cause”**

In the Third *Critique*, Kant goes a step further in presenting his critical metaphysics, this time in the framework of his philosophy of science. The noumenal realm, now described as the supersensible substratum of nature, forms a crucial part of Kant’s philosophy of science in this *Critique*, especially in the part called Critique of the Teleological Power of Judgment. As a consequence, we can safely assume that Kant regarded this concept as consistent with science. We cannot eliminate it from Kant’s philosophy of science without seriously damaging his arguments (McLoud, 2018, 50).

In the Third *Critique*, Kant is especially concerned with the products of nature. In the place of mechanism and spontaneity (introduced in the First *Critique*), he now introduces two similar but more sophisticated concepts, namely mechanism (differently constituted) and teleology (with “natural purpose” as the main feature).

With regard to the internal possibility that things have in themselves to produce their external form, Kant then introduces the relation between the “whole” and the “parts” as well as the particular causal relation which governs them. He distinguishes two conceivable ways in which this could be possible, namely through the “mechanism of nature”, when the material whole is explained by the causal relation between the component parts (the parts determine the whole), and through a “natural purpose”, when the idea of the whole (located in the supersensible ground of nature) serves as ground and condition for the parts and their internal arrangement (the whole determines the parts) (Allison, 2012, 203; McLaughlin, 1990, 129).

In the Third *Critique* the idea of the “whole”, understood as non-aggregated wholes-and-parts that contain the ground and possibility for the production of the products of nature (of the material whole with its parts), effectively supplants the concept of noumena that Kant uses in the First *Critique* (McLoud, 2018, 46-48). What is more is that Kant now introduces a certain potentiality, a “productive cause” (*KdU* 5, 370, 379), also called “the spontaneity of a cause” (*KdU* 5, 411), that

the products of nature have in themselves to produce their material forms, an idea based on the more basic one of absolute spontaneity. Kant conceptualizes *the capacity* of non-extended wholes-and-parts in the supersensible substratum of nature to produce extended parts and aggregated wholes in nature (and space/time) in terms of a “formative force” (Kant speaks of “self-organization” in this regard.) (McLoud, 2018, 46-49, 68)

As before, these Kantian ideas can be applied to quantum physics. Even though Kant formulated his ideas for the products of nature, his extensive use of the supersensible realm in the formulation of his view (and his explicit mention of fundamental entities in this regard in the *Metaphysical Foundations of Natural Science*; MAN, 507) allows for the application to the quantum world when that realm is understood in such terms. As a consequence, we can take superimposed states as agreeing with Kant’s non-aggregated wholes-and-parts, the parts constituting the whole, with the parts referring to fundamental entities. At bottom, in multi-particle systems, it is, in fact, these fundamental entities that are coupled (through their states) into wholes-and-parts as Kant proposed.

The outcomes of measurement can then also be understood in terms of parts and wholes. In contrast with superpositions of states conceptualized in terms of non-aggregated wholes-and-parts, the outcomes of measurement involve aggregated wholes defined in terms of probabilities in accordance with the Born postulate (the individual outcomes form an aggregated whole as in all probability formulations). Since definite values cannot be assigned to pre-measurement states in superposition, the outcomes are accordingly not determined by such values; they are given probabilistically.

As found in the Kantian system, the idea that quantum systems can be described in terms of potentialities goes back to Heisenberg’s later writings. He invoked the Aristotelian idea of potentiality, namely that all change consists in the actualization of potentialities, to describe the relation between the quantum state and its outcomes. Henry Margenau also thought in terms of propensities or latent quantities, i.e. that the measurement of an observable converts latent values into possessed values (Redhead, 1987, 48).

Hughes (1989) presents an event interpretation of quantum mechanics in which quantum properties are replaced by “latencies”. When a particular latency is ascribed to a quantum system, probabilities are assigned to the values of a family of observables which would be realized in events (Hughes, 1989, 309). Teller (1995), who also understands superimposed (i.e., quantum) properties in terms of propensities, applies such concepts to quantum fields.

#### 4.3. The Quantum Potential

David Bohm’s quantum potential also needs to be mentioned, with various authors having observed close similarities between Kantian and Bohmian ideas (Najmabadi et al. 2020; McLoud 2018; Palmquist, 2015). In contrast with Bohr, Bohm was interested in what determines the behavior of

an individual quantum system, i.e. the nature of the system's ontology (Van Strien, 2024, 3). This eventually led to the discovery of the quantum potential and the development of Bohm's theory (to be distinguished from Bohmian mechanics in that this is a second order theory, not a first order one; Romano, 2020, 12).

According to Basil Hiley (2002), the quantum potential can be regarded as an "internal" potential energy that belongs to particles in superpositions of states (it has no equivalent in classical mechanics and has no external source). Insofar as multi-particle systems are concerned, the quantum potential is understood in terms of a "whole" that determines the properties of the individual particles and their relationship (and not the other way round). As such it is a non-local energy, different from kinetic and (classical) potential energy, necessary for energy conservation which involves a spontaneous self-organizing process involving a basic underlying field (see Hiley, 1999, 7). As this terminology is remarkably similar to that of Kant, with both Kant and Bohm having been inspired in their views by the processes they observed in nature, a Kantian influence on Bohm can perhaps not be excluded given the exact same terminology.

Quantum field theory also needs to be brought into the discussion. What is quite clear from Auyang (1995), who follows a Kantian approach, is that quantum fields closely mirror the Kantian conception of the supersensible substratum of nature (McLoud, 2018, 62-7). In this framework pertaining to matter fields, the "basic entities or individuals of the physical world" are "extensionless in all four dimensions" (that is, they have no space or time parts) even though they could be indexed in the framework of a "primitive" space-time manifold *M* (Auyang, 1995, 123, 129).

These quantum entities are called "events" to express the fact that their *indexing* in *M* involves both space *and* time (as such they have *the potentiality* to be realized in space-time). These "events" should not be confused with what is normally understood by events in space-time. Only once events are actualized in some manner, are they represented mathematically as timelike or spacelike curves which are generated by mapping some part of a real number system onto the manifold *M* (Auyang, 1995, 171). As far as time is concerned, the manifold *M* is

too primitive to confer special meaning on the time dimension... *M* is...  
*the condition for the possibility of introducing the time parameter and the  
notion of being 'in time'...* [It] is independent of temporal concepts. It  
contains the time dimension as one aspect and makes possible the  
introduction of the time parameter, but *is itself beyond time and change.*  
(Auyang, 1995, 170, my italics)

Auyang's description of quantum fields is quite similar to Bohm's implicate order as can be seen from Hiley's comment in this regard, "[T]here is a deep underlying process from which not

only particles and fields emerge, but this process is the source of space-time itself.” (Hiley, 2010, 14) In this implicit order the quantum particles can be seen as vibrations of a global field existing on the fundamental level (Fiscaletti, 2018, 14).

When the quantum realm described by quantum fields (usually taken as the basic ontology of the world; Auyang, 1995, 45) are contrasted with the classical world with its objects of experience, it follows that quantum mechanics in actually brings these two worlds together. The temporal parameter of the Schrödinger equation allows for the classical description of quantum phenomena (Valente, 2010), whereas the mathematical space that describes the quantum entities are complex vector spaces (Cartwright, 1999, 217) that make their observation in proper space impossible (similar to entities in quantum fields not being observable in space-time). What is more is that the amplitudes associated with these entities are also complex quantities while our instruments can only measure real numbers (Auyang, 1995, 73).

#### 4.4. The Intermediary Role of Quantum Mechanics

Using Kant’s critical metaphysics, the quantum field description can be viewed as describing another ontological mode of being than the one found in the classical world, with these two worlds being brought together in quantum mechanics (a possibility not foreseen by Kant but consistent with his metaphysics). Whereas the quantum mode is ruled by spontaneity, the classical world is ruled by determinism. As such, the time parameter actually introduces classical characteristics to the quantum system: the time *evolution* of the superposition of states is governed by deterministic laws which subject the state vector (the inseparability of wholes-and-parts) to given forces and constraints in a manner similar to classical equations of motion (Albert, 1992, 34). One can view the Schrödinger equation as extending the classical time framework into quantum physics.

We can moreover read the time parameter as placing a (classical) constraint on the quantum mode (which is then in some manner constrained to be in time but not in proper space). The dynamic evolution of the quantum system, made possible through the time framework, constrains the superposition of states to evolve in a certain manner in time. When the system is further constrained (disturbed) in accordance with different experimental configurations, the superposition of states collapses to reduced states.

Another way in which this can perhaps be conceived is in terms of Bohm’s theory (Bohm and Hiley, 1993) in which the particles are being acted upon by two quantum fields which generate the quantum potential as well as an “Aristotelian” (classical) potential. Instead of Bohm’s use of a configuration space, the equations can be reformulated in ordinary three-dimensional space in which these fields become multi-fields that are dynamically coupled (Hubert and Romano 2018). When the quantum potential is negligible, the particles will move according to Newtonian trajectories. As for the quantum potential, it produces a non-local force that acts on the superposition of states simultaneously (Romano, 2020, 21).

When a two-particle configuration is considered (like in the Aspect experiment) the quantum field that generates the quantum potential will operate simultaneously on both particles despite their distance from each other. Even though the “collapse” of the wave packet does not feature in Bohmian theory, it seems reasonable to assume that the non-local collapse of the superposition of states to reduced states needs to be attributed to the quantum potential (given its non-local and spontaneous character). And as measurement of a property on one particle leads to the instantaneous and (absolutely) spontaneous reduction of the state of the other particle, this spontaneous “collapse” can only be attributed to the quantum potential acting non-locally.

It can thus be proposed that whereas in Bohr’s interpretation of quantum mechanics the quantum phenomena (properties) are observed without any consideration of the internal dynamics of the quantum system, the Bohm interpretation provides answers as to why the simultaneous collapse takes place: the quantum potential is responsible for the spontaneous collapse. It may be suggested that the way in which the potentiality for collapse is included in the Schrödinger equation (or rather its decomposed formulation) is through the quantum potential.

As the quantum potential describes the non-local entanglement of the particles, namely its potentiality for self-organization, it stands to reason that it also contains the potentiality for actualization though a spontaneous collapse. Even though this does not follow from Bohmian mechanics it does seem to be consistent with Bohm’s theory of the implicate order bringing forth the explicate order.

The strange combination of classical characteristics (described by the time parameter in the Schrödinger equation and the “Aristotelian” potential in Bohm’s theory) together with the quantum mode of existence (superpositions of states and the non-local quantum potential) in one system in quantum mechanics is reflected in the way in which quantum particles differ from their classical counterparts. They are not situated in space-time like classical particles; they do, however, retain some definite classical characteristics (deterministic evolution) due to their continuous time evolution which disappears in the quantum field description.

This metaphysical picture explains the measurement problem (the need for two different formulations, namely the Schrödinger equation and the projection postulate). The Schrödinger description can now be viewed as bringing two different modes of existence together in one formulation, i.e. the deterministic time evolution that constraints the superposition of states (“wholes-and-parts”) from collapsing (despite them having the potential to do so). When the system is further constrained in accordance with different experimental configurations, the superposition of states collapses to reduced states. In the case of the quantum field description, no similar constraints are placed on the quantum system, allowing it to make absolute spontaneous transitions to reduced states (for a detailed discussion see McLoud, 2018, 67-74, 78-81).

This explanation is supported by the work of Mohammad Jamali et al. (2019) who show that the quantum potential (which appears as part of an extra term in a modified Schrödinger equation) can indeed be associated with the collapse of the wave function in standard quantum mechanics. They propose that “because of this extra term, which leads to a nonlinear modified Schrödinger equation, there may be a possibility for the solution of measurement problem in QM (due to its nonlinearity and non-unitary evolution).” (Jamali et al. 2019, 7)

## 5. Implications for Contemporary Conversations in Physics

When we take the Bohmian quantum potential as the scientific equivalent of the Kantian potentiality or “productive cause” (*KdU* 5, 370, 379), we can apply Kantian metaphysics to an even broader range and array of issues pertaining to contemporary physics and the challenges associated with obtaining a comprehensive understanding of the workings of nature. Even though philosophers of science usually “follow the science”, there is no reason why they cannot make proposals guiding the direction of science. And when nature is interpreted through the lens of Kantian metaphysics it is indeed possible to make some proposals with regard to the problems plaguing our best theories.

### 5.1. Quantum Field Theory vs General Relativity

Perhaps the greatest challenge for physicists is reconciling quantum field theory with general relativity. This problem is about 100 years old and no convincing unifying theory has yet been found. Scientists are divided in their opinions as to where the problem lies, with some thinking quantum theory has to be modified (to include the quantization of gravity), others that general relativity is an incomplete theory (given the singularities in the theory as well as observational discrepancies like the Hubble tension) whereas still others have proposed that a more general theory (like string theory) would be the answer. None of these proposals have thus far provided real solutions.

Can Kant’s critical metaphysics provide some clues to better understanding the issue? In the Kantian conception of things, as presented already in the First *Critique*, systemic nature ruled by mechanism has reference to the classical world, with general relativity being a theory that describes the dynamics of “nature”. In contrast, the noumenal world ruled by absolute spontaneity can be viewed as having reference to the quantum world described by quantum field theory (McLoud, 2018). I have argued that in the Kantian system, these two worlds are ontologically distinct, being ruled by mechanism (determinism) and spontaneity respectively. This immediately suggests that they are sufficiently different that theories describing these worlds would be incompatible with each other (as we indeed find).

In theoretical descriptions of the quantum world, spontaneity has indeed become an essential element of the theory, especially when the quantum potential is included (also in quantum field

theory; Fisaletti 2018; Licata and Fisaletti 2014). The most distinguishing feature of the quantum potential is that it is non-local, which means that it operates outside the Lorentzian space-time manifold and can thus (exactly like quantum entities) not be directly observed (Earman 1986, 188). Authors have argued, however, that we have good reasons to believe that the quantum potential is real (Romano, 2020, 17) and many applications to natural phenomena seem to confirm that.

In his discussion of different mathematical approaches to the quantum potential, Davide Fisaletti (2018) discusses the work of different scholars. Among them counts Valeriy Sbitnev who regards the quantum potential as

[A]n information channel which emerges from the zero-point fluctuations of a physical vacuum acting as a special superfluid medium... the physical vacuum consists of an enormous number of virtual pairs of particles anti-particles with opposite orientations of spins (thus constituting a Bose ensemble) ... (Fisaletti, 2018, 29)

It may be proposed that the fluctuations of the electromagnetic field in the vacuum (described by quantum electrodynamics (QED), the quantum field theory of electrons and electromagnetic fields) are the source of the spontaneous character of the quantum potential. Jamali et al. (2019) actually shows that extending Bohmian theory to quantum field theory (in a second quantization) produces a modified quantum potential that can explain phenomena like creation and annihilation of particles.

What is quite fascinating is that the same process featuring in Sbitnev's work, namely the constant emergence and interaction of pairs of particles and their antiparticles such as electrons and positrons, has recently been identified as the source of dark energy in the cosmos. According to Alexandre Tkatchenko (2023), his team obtained results that agree well with the measured values for the cosmological constant (in general relativity). It will be interesting to discover what relation exists between the quantum potential and dark energy and if they perhaps refer to the same feature.

Various different theoretical models describing the coupling of general relativity with quantum field theory have been proposed through the years, most recently by Jonathan Oppenheim (2023). From the standpoint of Kantian metaphysics such interaction is possible because of the intermediate role that quantum mechanics plays between quantum fields and general relativity.

Any real unifying theory will have to take the ontological differences between the classical and quantum worlds into account, namely the spontaneous character of quantum fields (described by the quantum potential) and the space-time description of gravity. Oppenheim did in fact use classical gravity (with no need for quantum gravity). For his part, Fisaletti (2018, 92, 124) writes that Bohm's quantum potential introduces interesting unifying perspectives also as regards the treatment of gravitation.



Kantian metaphysics suggests that the ontological difference between quantum fields and the classical world (general relativity/gravity) implies that the quantization of gravity is not achievable.

## 5.2. Consciousness

As already mentioned, in the Third *Critique* Kant reworks and refines his concepts of mechanism and spontaneity within the framework of the description of the products of nature, with a potentiality replacing the simple notion of absolute spontaneity. It is this potentiality that agrees so closely with the quantum potential. It is therefore to be expected (in accordance with the Kantian insights) that Bohm's theory would be applied to organic life and its self-organizational features.

Quantum coherence consistent with self-organization is observed in many biological contexts, from the selectivity and transport of ion channels through the membranes into cells (Seifi et al. 2022) to the quantum coherence in phase differences of dipole-bound electron oscillations in the brain (Poznanski et al. 2018). Quantum coherence (micro vibrations) is experimentally verified by a bundle of microtubule (EEG waves), with genetic experiments also having uncovered important roles for the quantum aspects of microtubules together with kinesin in the regulation of physiological processes (Levi, 2020, 74).

This brings us to consciousness. What is quite remarkable is that Kant describes apperception (primitive self-awareness or self-consciousness; see Hanna and Thompson 2013) in quite similar terms as the products of nature, namely as being spontaneously produced from the underlying noumenal reality. Robert Peppin (1987) shows that Kant attributes its spontaneity to apperception having its origins in the noumenal self situated in the noumenal realm. Kant for example says that

The soul in transcendental apperception is *substantia noumenon*, hence it has no permanence in time, since this belongs only to objects in space.”  
(Reflexion, 6001, AA, vol. XVIII, 420-1)

In this way Kant ascribes the spontaneity that characterizes apperceiving as having its origin in the noumenal self (soul). It has been suggested that in this Kant contravenes his view on Noumenal Ignorance, but this is not necessarily the case as one can regard Kant's ascription of spontaneity to the noumenal realm not as constituting any real knowledge but simply the only possible way that he was able to accommodate spontaneity within his philosophy (see Peppin, 1987, 272).

The Kantian idea of spontaneous mental activities has also been reworked in contemporary philosophical models of consciousness, such as the Essential Embodiment Theory of Robert Hanna and Michelle Maiese (2009), who view such activities as characterizing complex dynamic living organisms, steering clear of both dualism and materialism. In Essential Embodiment Theory the mind is taken as deeply and inextricably interwoven as well as emergent from an interactive system consisting of brain, body and world.

In accordance with Kant's conception of a spontaneous consciousness (and my identifying his spontaneous potentiality with the quantum potential), theories of consciousness using Bohmian theory have indeed been proposed. In one case, the well-known model of Roger Penrose and Stuart Hameroff is fused with Bohm's metaphysical concepts to develop a Bohm-Penrose-Hameroff model for consciousness and free will (Gallego, 2011).

A mathematical model in which the de Broglie-Bohm wave mechanics is extended to explain macro-quantum phenomena is proposed by R. R. Poznanski et al. (2018) in a paper titled Spontaneous Potentiality as Formative Cause of Thermo-quantum Consciousness. They show that macroscopic quantum coherent states (associated with consciousness) can exist in the brain when the thermo-quantum fluctuations of dipole-bound electron oscillations go through their phase differences, "guided" by the macro-quantum potential energy through "long-range" correlations of phase differences:

It is the spontaneous phase differences of dipolar-bound electron oscillations that govern 'long-range order' (i.e., high degree of coherence), leading to stability and unitary binding of consciousness. (Poznanski et al, 2018, 372)

The important role of information within quantum potentials is emphasized by David Bohm and his co-worker Basil Hiley. They introduced the idea of "active information" to account for the properties of the quantum potential (Bohm and Hiley, 1993). For many-body systems, the self-organization at the heart of the quantum process involves a non-local correlation of the motion of all the bodies in the entangled state, which are all being simultaneously organized by a collective field, implying that they have access to "a common pool of information encoded in the entangled wave function" (Hiley, 1999). When some of these states/packets collapse, they "transport" the information to other states/packets which means that the information is conserved by the quantum potential. This explains how information is conserved and stored in consciousness and memory.

An interesting feature of critical metaphysics is that the "wholes" embedded in the supersensible ground of nature which underlies the self-organizational processes that give form to biological formations correlate with the human noumenon (soul), which implies that the human noumenon is simply a more complex and sophisticated "whole" that evolved in nature. When it comes to consciousness, lower order biological organisms thus share this feature even though that of humans reaches the more advanced level of self-awareness (apperception).

It may be asked what happens to the information encoded in consciousness when biological organisms die. In quantum physics information is regarded as being conserved but that need not imply that the "wholes" (taken as ensembles of quantum potentials closely integrated with their bodies) continue to be preserved as *single united entities* apart from bodies. For humans, however,

an ancient view found all over the world (also in Platonic philosophy) insists that the soul survives death as an immortal entity. Is it plausible that this possibility can be established in science?

When the link between the soul and apperception in critical metaphysics is taken as point of departure, it may be suggested that the coherent, complex and integrated nature of the information stored in the framework of human self-consciousness in the brain (serving as the guiding operative for the collective human ensemble of quantum potentials in the body), can provide it with a permanence that lower order beings lack. This agrees to some extent with a proposal made by Stuart Hameroff (2012), namely “[When] the patient dies, it is possible that this quantum information can exist outside the body, perhaps indefinitely, as the soul”. Even though this proposal takes us beyond the reach of current science, this is to be expected. In the words of Roger Penrose (2022):

[Consciousness] must be beyond computable physics... a theory that we do not know yet.

The Kantian ontological distinction between nature and the noumenal realm may again be relevant in this regard. As for the noumenal realm, it can be proposed that Bohm’s idea of a non-local (beyond space-time) and thus unobservable “quantum field” consisting of a universal network of quantum potentials intersecting with each other (and forming the implicate order), provides for an underlying manifold enabling existence in an invisible world to which souls and other spiritual beings might belong. This invisible world stands in contrast with but not entirely apart from our space-time manifold in which all material things and beings exist.

Such a view is very similar to the Platonic conception of a realm of forms in which souls exist (*Phaedo*, 78b-79) that stands in contrast with the world of becoming and material existence, with the realm of forms also later on in the *Timaeus* serving as the substratum of nature. Both these worlds thus belong to one cosmos.

## Conclusion

In this paper the role of spontaneity in our understanding of the world is taken under consideration, since the time when Immanuel Kant first included absolute spontaneity in his critical metaphysics and also after its discovery in nature in the form of quantum spontaneity. On the one hand I give an overview of the way that spontaneity is included in the formulation of quantum physics, from Niels Bohr’s quantum postulate (and his epistemological approach) to David Bohm’s quantum potential (and his ontological approach). On the other hand, I engage with Kant’s critical metaphysics in search for clues as to how quantum physics need to be interpreted within the broader framework of our understanding of the workings of nature.

I then use critical metaphysics together with our current understanding of science to make some proposals with regard to the problems plaguing our best theories, including the problem of reconciling quantum field theory with general relativity and those concerning the essence of consciousness. The state of our current knowledge makes proposals possible that would have been impossible a few years ago, proposals that align with Kantian metaphysics. In the final instance, I think it is reasonable, on this anniversary of Kant's birthday, to come to the conclusion that his critical metaphysics is remarkably encompassing, relevant and accurate in the light of our current knowledge of science. As such it may be expected that other aspects of his metaphysics which are not yet within the reach of science, will also in the progress of science turn out to be correct.

## References

- Albert, D. Z. (1992). *Quantum Mechanics and Experience*, Harvard University Press.
- Allais, L. (2004). Kant's One World: Interpreting Transcendental Idealism, *British Journal for the History of Philosophy*, 12(4), 655-684. <http://dx.doi.org/10.1080/0960878042000279314>
- Allison, H. E. (2012). *Essays on Kant*, Oxford University Press.
- Allori, V. & et al. (2008). On the Common Structure of Bohmian Mechanics and the Ghirardi-Rimini-Weber Theory, *The British Journal for the Philosophy of Science*, 59(3), 353-389. <http://dx.doi.org/10.1093/bjps/axn012>
- Ameriks, K. (1992). Kantian Idealism Today, *History of Philosophy Quarterly*, 9(3), 329-34.
- Auyang, S. Y. (1995). *How is Quantum Field Theory Possible?* Oxford University Press.
- Bartels, A. (1999). Objects or Events? Towards an Ontology for Quantum Field Theory, *Philosophy of Science*, 66(3), 170-184.
- Bell, J. (1990). Against 'measurement', *Phys. World*, 3(8), 33. <http://dx.doi.org/10.1088/2058-7058/3/8/26>
- Bitbol, M. (2007). Materialism, Stances, and Open-Mindedness', in *Images of Empiricism*, Edited by B. Monton, Oxford University Press.
- Bitbol, M. (2008). Reflective Metaphysics: Understanding Quantum Mechanics from a Kantian Standpoint', *Philosophica*, 83(1), 53-83. <http://dx.doi.org/10.21825/philosophica.82161>
- Bitbol, M. (2013). Bohr's complementarity and Kant's epistemology, in *Séminaire Poincaré*, 17, 145-166. [http://dx.doi.org/10.1007/978-3-319-14316-3\\_8](http://dx.doi.org/10.1007/978-3-319-14316-3_8)
- Bitbol, M. (2017). On Bohr's Transcendental Research Program, in *Niels Bohr and the Philosophy of Physics*, pp. 47-66. Edited by Faye & Folse, Cambridge University Press.
- Bitbol, M. & Osnaghi, S. (2013). Bohr's Complementarity and Kant's Epistemology, Bohr, 1913-2013, *Séminaire Poincaré*, XVII, 145-166.
- Bohm, D. & Hiley, B. J. (1993). *The Undivided Universe*, Routledge.
- Bohr, N. (1928). The *Quantum Postulate* and the Recent Development of Atomic Theory, *Nature*, 121, 580-590.

- Bohr, N. (1934). *Atomic Theory and the Description of Nature*. Cambridge University Press.
- Bohr, N. (1948). On the notions of causality and complementarity, *Dialectica*, 2(3-4), 312-319. <http://dx.doi.org/10.1111/j.1746-8361.1948.tb00703.x>
- Bohr, N. (1955). Atoms and Human Knowledge. In, *Atomic Physics and Human Knowledge*, pp. 83-93, Wiley.
- Bohr, N. (1962). The Solvay meeting and the development of quantum physics. In, *Essays 1958-1962 on atomic Physics and Human Knowledge*, pp. 79-100, Interscience Publishers.
- Cartwright, N. (1983). *How the Laws of Physics Lie?* Oxford University Press.
- Cartwright, N. (1989). *Nature's capacities and their Measurement*, Clarendon/Oxford University Press.
- Chevalley, C. (1994). Niels Bohr's Words and the Atlantis of Kantianism, in *Niels Bohr and Contemporary Philosophy*, pp. 33–55, Edited by J. Faye & H. Folse, Kluwer. Cambridge University Press.
- Cuffaro, M. (2010). The Kantian framework of complementarity, in *Studies in History and Philosophy of Science*, 41, 309–317. <http://dx.doi.org/10.1016/j.shpsb.2010.04.003>
- Donadi, S. & et al. (2021). Underground test of gravity-related wave function collapse. *Nature Physics*, 17(1), 74–78. <https://www.nature.com/articles/s41567-020-1008-4>
- Dorato, M. (2002). Kant, Gödel and Relativity', in *Proceedings of the invited papers for the 11th International Congress of the Logic Methodology and Philosophy of Science*, Edited by P. Gardenfors, K. Kijania-Placek, & J. Wolenski, Kluwer.
- Dorato, M. (2017). Bohr's Relational Holism and the Classical-Quantum Interaction, in *Niels Bohr and the Philosophy of Physics*, pp. 133–154, Edited by J. Faye & H. Folse, Bloomsbury Academic.
- Earman, J. (1986). *A Primer on Determinism*, D. Reidel.
- Fano, V. (1988). How Italian Philosophy Reacted to the Advent of Quantum Mechanics in the Thirties', in *The Nature of Quantum Mechanics*, pp. 385-401, Edited by G Tarozzi & A. Merwe, Kluwer, Cambridge University Press.
- Faye, J. (1991). *Niels Bohr: His Heritage and Legacy. An Antirealist View of Quantum Mechanics*, Kluwer Academic Publisher.
- Faye, J. (2019). Copenhagen Interpretation of Quantum Mechanics, *Stanford Encyclopedia*: <https://plato.stanford.edu/ENTRIES/qm-copenhagen/>.
- Fiscaletti, D. (2018). *The Geometry of Quantum Potential Entropic Information of the Vacuum*, SpaceLife Institute. <https://doi.org/10.1142/10653>
- Folse, H. (1985). *The Philosophy of Niels Bohr: The Framework of Complementarity*, North Holland.
- Folse, H. (1986). Niels Bohr, Complementarity, and Realism, in *PSA 1986: Proceedings of the Biennial Meeting of the Philosophy of Science Association Vol. I*, Edited by A. Fine & P. Machamer East Lansing: PSA, pp. 96–104, Cambridge University Press.

- Gallego, M. B. (2011). The Bohm-Penrose-Hameroff model for consciousness and free will theoretical foundations and empirical evidences, *Pensamiento*, 67 (254), 661-674.
- Hameroff, S. (2012). American science documentary television series *Through the Wormhole* narrated by. Morgan Freeman and aired in 2012 on the Science Channel.
- Hanna, R. (2006). *Kant, Science and Human Nature*, Oxford University.
- Hanna, R. & Maiese, M. (2009). *Embodied Minds in Action*, Oxford University.
- Hanna, R. & Thonpson, E. (2013). Neurophenomenology and the Spontaneity of Consciousness, *Canadian Journal of Philosophy*, 33: sup.1: 133-162.
- Healey, R. (2009). *Gauging What's Real. The Foundations of Contemporary Gauge Theories*, Oxford University.
- Hiley, B. J. (1999). Active Information and Teleportation, in *Epistemological and Experimental Perspectives on Quantum Physics*, Edited by D. Greenberger & et. al, Kluwer. Cambridge University Press.
- Hiley, B. J. (2002). From the Heisenberg Picture to Bohm: A New Perspective on Active Information and its relation to Shannon Information', in *Proc. Conf. Quantum Theory: Reconsidering of Foundations*, Växjö University.
- Hiley, B. J. (2010). *Some Remarks on the Evolution of Bohm's Proposals for an Alternative to Standard Quantum Mechanics*. Preprint.
- Honner, J. (1982). The transcendental philosophy of Niels Bohr, *Studies in History and Philosophy of Modern Physics*, 13: 1–29.
- Honner, J. (1987). *The Description of Nature: Niels Bohr and The Philosophy of Quantum Physics*, Clarendon Press.
- Howard, D. (1994). What makes a classical concept classical? Toward a reconstruction of Niels Bohr's philosophy of physics. In *Niels Bohr and contemporary philosophy*, Edited by J. Faye & H. Folse, Kluwer.
- Howard, D. (2004). Who invented the "Copenhagen interpretation?" A study in mythology. *Philosophy of Science* 71:669-682. <http://dx.doi.org/10.1086/425941>
- Howard, D. (2005). Revisiting the Einstein-Bohr dialogue. Cal Tech, October 2005; Jerusalem, Bar-Hillel Lecture, December 2005; Special issue of Iyyun in honor of Mara Beller).
- Hubert, M. & Romano, D. (2018). The wave-function as a multi-field, *European Journal for Philosophy of Science*, 8(3), 521–537. <https://doi.org/10.1007/s13194-017-0198-9>
- Hughes, R. I. G. (1989). *The Structure and Interpretation of Quantum Mechanics*, Harvard University Press.
- Jamali, M.; Golshani, M. & Jamali, Y. (2019). A proposed mechanism for mind-brain interaction using extended Bohmian quantum mechanics in Avicenna's monotheistic perspective, *Heliyon*. <https://doi.org/10.1016/j.heliyon.2019.e02130>

- Kaiser, D. (1992). More Roots of Complementarity: Kantian Aspects and Influences, *Studies in History and Philosophy of Science*, 23, 213–239.
- Kauark-Leite, P. (2010). Transcendental philosophy and quantum physics, *Manuscripto – Rev. Int. Fil.* 33, 243-267.
- Kauark-Leite, P. (2017). Transcendental versus Quantitative Meaning of Bohr's Complementarity Principle, in *Niels Bohr and the Philosophy of Physics*, pp. 67–90, Edited by J. Faye & H. Folse, Bloomsbury Academic.
- Langton, R. (1998). *Kantian Humility. Our Ignorance of Things in Themselves*, Oxford University Press.
- Levi, P. (2020). Basic Quantum Field Model of the Self-Organization of Microtubules in Eukaryotic Cells, *European Journal of Biophysics*, 8(2), 60-75. <https://doi.org/10.11648/j.ejb.20200802.17>
- Licata, I. & Fiscaletti, D. (2014). *Quantum Potential: Physics, Geometry, and Algebra*, Springer-Verlag.
- McLaughlin, P. (1990). *Kant's Critique of Teleology in Biological Explanation*, The Edwin Mellen Press.
- McLoud, W. (2018). Kant, Noumena and Quantum Physics, *Contemporary Studies in Kantian Philosophy*, 3.
- Najmabadi, I.; Tahmasebie, S. & Dehbashi, M. (2020). A Fresh Look at the Position of the Thing-in-Itself According to Kant and Hegel Based on the Metaphysical Foundations of David Bohm's Quantum Physics. *Journal of Philosophical Theological Research*, 22(84), 49-72. <https://doi.org/10.22091/jptr.2020.4954.2229>
- Penrose, R. (2022). Consciousness must be beyond computable physics. *New Scientist*, YouTube Channel. [https://www.youtube.com/watch?v=TfouEFuB-co&ab\\_channel=NewScientist](https://www.youtube.com/watch?v=TfouEFuB-co&ab_channel=NewScientist)
- Peppin, R. B. (1987). Kant on the Spontaneity of Mind, *Canadian Journal of Philosophy*, 17(2), 449-476.
- Poznanski, R. R. & et al. (2018). Spontaneous potentiality as formative cause of thermo-quantum consciousness', *J. Integr. Neurosci.* 17(4). <https://doi.org/10.31083/j.jin.2018.04.0418>
- Pringe, H. (2007). *Critique of the Quantum Power of Judgment. A Transcendental Foundation of Quantum Objectivity*, Walter de Gruyter.
- Pringe, H. (2009). A Transcendental Account of Correspondence and Complementarity, in *Constituting Objectivity: Transcendental Perspectives on Modern Physics*, Edited by M. Bitbol, P. Kerszberg & J. Petitot, Springer.
- Pringe, H. (2023). Kant and Bohr on Quantum Objectivity, *Roum. Philosophie*, 67(2), 247–265.
- Redhead, M. (1987). *Incompleteness, Nonlocality, and Realism: A Prolegomenon to the Philosophy of Quantum Mechanics*, Oxford University Press.
- Romano, D. (2020). Multi-field and Bohm's theory, *Synthese* (11), 29. <https://link.springer.com/article/10.1007/s11229-020-02737-6>
- Seifi, M.; Soltanmanesh, A. & Shafiee, A. (2022). Quantum coherence on selectivity and transport of ion channels, *Sci Rep.* 12(1):9237. <https://doi.org/10.1038/s41598-022-13323-w>



- Soler, L. (2009). The Convergence of Transcendental Philosophy and Quantum Physics: Grete Henry-Hermann's 1935 Pioneering Proposal', in *Constituting Objectivity: Transcendental Perspectives on Modern Physics*, Edited by M. Bitbol, P. Kerszberg & J. Petitot, Springer.
- Teller, P. (1995). *An Interpretive Introduction to Quantum Field Theory*, Princeton University Press.
- Tkatchenko, A. & et al. (2023). Casimir Self-Interaction Energy Density of Quantum Electrodynamical Fields, *Physical Review Letters*. <https://doi.org/10.1103/PhysRevLett.130.041601>
- Valente, M. B. (2010). Bohr's quantum postulate and time in quantum mechanics, [philsci-archive.pitt.edu/8335/1/Bohr%27s\\_quantum\\_postulate\\_and\\_time\\_in\\_quantum\\_mechanics.pdf](https://philsci-archive.pitt.edu/8335/1/Bohr%27s_quantum_postulate_and_time_in_quantum_mechanics.pdf)
- Van Strien, M. (2024). Why Bohm was never a determinist, in *Guiding Waves in Quantum Mechanics: 100 Years of de Broglie-Bohm Pilot-Wave Theory*, Edited by A. Oldofredi, Oxford University.



## The Kantian Self versus Pattern Theory of the Self

Mohamad Mahdi Moghadas<sup>1</sup>  | Ali Fath Taheri<sup>2</sup> 

<sup>1</sup> Corresponding Author, Ph.D. Candidate, Department of Philosophy, Imam Khomeini International University, Iran. Email: [mehmoghadas@gmail.com](mailto:mehmoghadas@gmail.com)

<sup>2</sup> Assistant Professor, Department of Philosophy, Imam Khomeini International University, Iran. Email: [fathtaheri@hum.ikiu.ac.ir](mailto:fathtaheri@hum.ikiu.ac.ir)

---

### Article Info

#### Article type:

Research Article

#### Article history:

Received 09 June 2024

Received in revised form  
28 June 2024

Accepted 11 July 2024

Published online 07  
August 2024

#### Keywords:

self, Kant, pattern theory  
of the self, noumenal self,  
phenomenal self

### ABSTRACT

In the history of philosophy, the concept of the self has long been a subject of intense debate and scrutiny. Within Kant's critical philosophy, the self holds a significant position and is deemed essential for the very notion of experience. This article aims to clarify Kant's viewpoint on the concept of self. Kant posits the existence of an inner sense faculty, through which he introduces the empirical or phenomenal self, a concept that aligns with Hume's stance on the self. Furthermore, Kant introduces the idea of a noumenal self through the faculty of understanding and apperception, suggesting that this noumenal self is the foundation upon which all our experiences are made possible. Our primary objective is to clarify the distinctions between the two forms of self, which will be succeeded by an assessment of each. Although numerous scholars tend to view Kant's perspective on the self as a negative concept, an alternative viewpoint emerges, proposing a potential positive interpretation of Kant's concept of the self. Subsequently, we present Gallagher's Pattern theory of the self and conduct a comparative analysis between the Kantian self and the components that constitute the self.

---

**Cite this article:** Moghadas, M. M. & Fath Taheri, A. (2024). The Kantian Self versus Pattern Theory of the Self. *Journal of Philosophical Investigations*, 18(47), 89-110. <https://doi.org/10.22034/JPIUT.2024.62276.3801>



© The Author(s).

Publisher: University of Tabriz.

<https://doi.org/10.22034/JPIUT.2024.62276.3801>

---

## Introduction

Immanuel Kant's philosophy has had a profound impact on Western thought, and his views on the self are no exception. However, understanding Kant's self is complex, as it lies at the intersection of his epistemology, metaphysics, and ethics. Central to Kant's project is the distinction between phenomena and noumena. We can only access the world through our *a priori* forms of intuition, space and time, and categories of understanding. This distinction, leaves the self-in-itself unknowable. However, Kant argues for a *transcendental self*, a unifying principle that synthesizes our diverse experiences into one cognition. This self is not a substance but an activity – the "I think" that accompanies all representations (B131).

Early interpretations of Kant's self-theory were heavily influenced by the prevailing philosophical movements. Fichte and Schelling, two prominent Post-Kantian Idealists, placed great emphasis on the active involvement of the self in the construction of reality. They viewed the transcendental self as a dynamic and innovative force, which propelled Kant's concept of self towards a more comprehensive and substantial understanding. Conversely, Realists such as Maimonides directed their attention towards the constraints inherent in Kant's project. They contended that the differentiation between the phenomenal and noumenal self ultimately erodes the possibility of attaining genuine knowledge about the self. In contrast, Neo-Kantians like Hermann Cohen highlighted the ethical aspect of the self, underscoring its significance in moral considerations. They saw the self as constituted through the pursuit of the Categorical Imperative, highlighting the role of moral agency in shaping selfhood. Phenomenologists like Edmund Husserl offered a different perspective. He argues that Kant neglected the lived experience of the self, focusing too heavily on the *a priori* structures of consciousness. Scholars like Henry Allison emphasize this active, unifying role of the self in constituting experience. He suggests the self is a 'regulative idea' that allows us to make sense of the unity of experience over time (Allison, 2001). Patricia Kitcher argues that the self is a "concept under which we subsume all our mental states" (Kitcher, 1982). According to Günter Zöller's review of the literature, "differences of method and philosophical approach aside, the interpretations ... show a remarkable agreement in their understanding of Kant's thinking self as a form or structure that eludes any attempt at reification" (Zöller 1993, 460).

On the other hand, Kant critiques the 'rational psychologist' tradition, which seeks to prove the self as a permanent, immaterial soul. He argues that such arguments, like the "cogito ergo sum", rely on Paralogisms that confuse the transcendental self with a psychological entity (B407-411). Wilfrid Sellars (1956) highlights this critique as a significant contribution to modern philosophy. Several scholars interpret the self as a schema, a unifying principle that organizes experience. These interpretations highlight the self's role in unifying consciousness without necessarily positing a permanent soul. While some may seek a positive account of the self ("the referent of 'I'") within

Kant's philosophy, the prevailing interpretation casts him as skeptical of such endeavors. Scholars often argue that Kant warns against attempts to define the self's essence and instead constructs his theories without relying on metaphysical assumptions. This dominant interpretive lens draws parallels between Kant's project and Ludwig Wittgenstein's assertion that the self "doesn't belong to the world, but is a limit of the world," suggesting a potential prefiguration of Wittgenstein's ideas (Marshall, 2010, 1).

Kitcher states that the problem with Kant's views about the self is that he has too many of them because the self has too many roles to play in his system (Kitcher, 1982, 41). Accordingly, Kant talks about the metaphysics of the I (*das Ich*), the mind (*das Gemüt*), the thinking subject (*das denkende Subjekt*), and the soul (*die Seele*). In addition, he uses both the substantive form of 'self' (*'Selbst'*) and the simple reflexive form (*'selbst'*). The specific interrelations between these terms remain unarticulated in Kant's work; however, he consistently deploys them in an indistinguishable fashion (Marshall, 2010, 2).

While Kant's conception of the self is multifaceted and intricate, it can be contextualized within a historical framework as a reaction to the ideas put forth by earlier philosophers such as René Descartes, John Locke, and David Hume. Kant's critique of Descartes' concept of the self and his rebuttal of Hume's bundle theory are especially noteworthy. This article aims to elucidate the relationship between Kant's responses to these two philosophers. Consequently, we will delve into Kant's perspective on the self, culminating in a critical examination of his views. Finally, we compare Kant's theory of the self with the 'Pattern Theory of the Self' proposing by Shaun Gallagher.

### **1. Kant, Descartes, Hume, and problem of the self**

Descartes suggests that the mind, or the self, which engages in thinking and knowing, is a substantial entity. Its essential nature is clearly understood through feelings and reflection, but can also be confused by the awareness of external ideas that are mistakenly perceived as qualities of physical bodies. The thinking mind is a non-physical substance, separate from the physical body. This thinking mind constitutes the self – an individual, immaterial entity. Descartes believed we can access the self through introspection, gaining clear and distinct knowledge of its essential nature as a thinking thing. In Descartes' view, the thinking mind is the self. Kant, it seems, provides a list of attributes for this self. These attributes are assigned to what we call "rational psychology":

The topics of the rational doctrine of the soul, from which everything else that it may contain has to be derived, are therefore the following: 1. The soul is substance 2. In its quality simple 3. In the different times in which it exists, numerically identical ..., and 4. In relation to possible objects in space (A344/B402).

Based on these four characteristics, Kant introduces four fallacies or Paralogisms in the section ‘Transcendental Dialectics’ of his *Critique of Pure Reason*. Kant challenged the notion of a purely individual, substantial self. He argued for a distinction between the phenomenal self (the self we experience) and the noumenal self (the "thing-in-itself" beyond experience). The phenomenal self is constituted by the transcendental unity of apperception, a unifying principle that allows us to synthesize experiences and maintain a sense of selfhood over time. This self is not a substance but a necessary condition for experience itself.

Hume argues that we have no direct experience of a self as a substance. Instead, the self is merely a ‘bundle’ or ‘collection’ of our perceptions – impressions and ideas. These perceptions constantly change, leading Hume to declare famously, “There is no abiding or identical impression, to which a train of thought can be referred, and which can be regarded as the foundation of personality.” For Hume, personal identity is a matter of memory and imagination, creating a sense of continuity across these fleeting perceptions. In his famous statement, Hume rejects that we can catch any perception of the self:

For my part, when I enter most intimately into what I call myself, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never can catch myself at any time without a perception, and never can observe anything but the perception. When my perceptions are removed for any time, as by sound sleep; so long am I insensible of myself, and may truly be said not to exist. And were all my perceptions removed by death, and coued I neither think, nor feel, nor see, nor love, nor hate after the dissolution of my body, I should be entirely annihilated, nor do I conceive what is farther requisite to make me a perfect non-entity. If anyone, upon serious and unprejudiced reflection thinks he has a different notion of himself, I must confess I call reason no longer with him. All I can allow him is, that he may be in the right as well as I, and that we are essentially different in this particular. He may, perhaps, perceive something simple and continued, which he calls himself; though I am certain there is no such principle in me. But setting aside some metaphysicians of this kind, I may venture to affirm of the rest of mankind, that they are nothing but a bundle or collection of different perceptions, which succeed each other with an inconceivable rapidity, and are in a perpetual flux and movement... There is properly no simplicity in it at one time, nor identity in different; whatever natural propension we may have to imagine that simplicity and identity... They are the successive perceptions only, that constitute the mind; nor have we the most distant notion of the place, where

these scenes are represented, or of the materials, of which it is composed. (Hume, 2009, 395-396).

Hume argues that the evidence of experience provides no basis for believing in a distinct entity within the individual responsible for thought. Consequently, a singular, unified thinking self becomes even less tenable. In Hume's view, the mind lacks a singular, enduring substance. Instead, it is a collection of diverse perceptions, or mental states interconnected through specific relations. These relations, however, mislead us into mistaking a sequence of distinct experiences for a unified and lasting self. Hume contested the existence of inherent, necessary connections between mental states, suggesting that our perception of such relationships might be illusory. This is the point Kant conflicts with him. Kant aimed to establish essential connections among mental states by asserting that these are necessary conditions for the feasibility of experience. Kant's consistent focus on a necessary connection or necessary synthesis of mental states has virtually nothing to do with theoretical psychology. This is a typical Kantian reaction to the predicament that Hume posed. In both editions of the Deduction, Kant continually adverts to a relation among representations, the relation of synthesis. He offers an explicit definition of this relation:

By synthesis in the most general sense, however, I understand the action of putting different representations together with each other and comprehending their manifoldness in one cognition (A77/B103).

For Kant, 'representation' signifies both 'mental state' and the 'content of a mental state'. The last clause refers to the content (manifold) of different mental states being combined in a further mental state. As Kant realized, token mental states are episodic, occurring at different times, so they themselves cannot literally be put together (Kitcher, 1982, 53).

Kant establishes a clear link between transcendental synthesis and the concept of self-identity by directly equating transcendental synthesis with the transcendental unity of apperception. In B135, Kant states that "I am therefore conscious of the identical self in regard to the manifold of the representations that are given to me in an intuition because I call them all together my representations, which constitute one. But that is as much as to say that I am conscious a priori of their necessary synthesis, which is called the original synthetic unity of apperception..."

Despite these considerations, Kant, in his examination of the problem of the self, distinguishes at least two types of self: 'the empirical self' and 'the transcendental self'. In the following, we will strive to present a precise and clear outline of Kant's perspective on the self.

## **2. The Empirical Self and Inner Sense**

While the transcendental self is necessary for experience, Kant also acknowledges the empirical self. In this regard, we consider the self as an object of experience. We have an inner sense, a faculty that allows us to introspect our mental states like desires, thoughts, and feelings. This

empirical self is located in time and is constantly changing. Katharina Kraus (2020) argues that inner sense provides the raw material for self-knowledge, which is an ongoing process of self-reflection and moral development.

Before delving into an examination of Kant's perspective, two crucial considerations merit attention. According to Marshall (2010, 2), the two point is that (1) Kant recognizes a broad use of 'I' that refers to the "whole man," involving body and soul. However, Kant's use of 'I' in the *Critique of Pure Reason* is meant in a stricter sense: "I, as thinking, am an object of inner sense, and am called 'soul'" (A342/B400). (2) Kant holds that having a self requires more than merely possessing representations — it requires having certain faculties.

Kant presents the inner sense as an empirical apperception. Here, Kant admits Hume approach about the self that there is no abiding self. In A107 he states:

The consciousness of oneself in accordance with the determinations of our state in internal perception is merely empirical, forever variable; it can provide no standing or abiding self in this stream of inner appearances, and is customarily called inner sense or empirical apperception.

Kant distinguishes between two modes of 'I'. The first, the "I as subject of thinking," corresponds to pure apperception. Pure apperception represents the intellectual, a priori foundation for synthesizing our thoughts. This "I" is not an object but the unifying principle that allows us to experience ourselves as thinking beings. In contrast, the second mode, the "I as the object of perception," is linked to the inner sense. The inner sense is the subjective counterpart of the outer senses, providing us with a sensory consciousness of ourselves and our thoughts, including the various "determinations" that shape our inner experience. Kant, explicitly states that what we intuit in the inner sense is not the kind of things-in-itself (Forgione, 2019, 26-27):

We must order the determinations of inner sense as appearances in time in just the same way as we order those of outer sense in space; hence if we admit about the latter that we cognize objects by their means only insofar as we are externally affected, then we must also concede that through inner sense we intuit ourselves only as we are internally affected by our selves, i.e., as far as inner intuition is concerned we cognize our own subject only as appearance but not in accordance with what it is in itself (B156)

The crux of the matter lies in the subject's engagement with the appearances through the inner sense. This engagement leads to a self-perception that is no more than an appearance, mirroring how the subject interacts with appearances of external objects. In both scenarios, the subject fails to access the 'thing-in-itself.'

According to Kant, time is the a priori form of intuition which is the condition of all outer and inner perception.

Time is the a priori formal condition of all appearances in general. Space, as the pure form of all outer intuitions, is limited as an a priori condition merely to outer intuitions. But since, on the contrary, all representations, whether or not they have outer things as their object, nevertheless as determinations of the mind themselves belong to the inner state, while this inner state belongs under the formal condition of inner intuition, and thus of time, so time is an a priori condition of all appearance in general, and indeed the immediate condition of the inner intuition (of our souls), and thereby also the mediate condition of outer appearances (A34/B50).

Kant argues that time, as the a priori formal condition, allows for the revelation of all possible representations. Through time, which functions as the form of the inner sense, the subject recognizes these representations as belonging to itself. Similar to its role as the condition for all appearances in general, time is the immediate condition for inner intuition (of our souls) and the mediate condition for outer appearances. However, in A33/B50, it seems that Kant gives another role to inner sense. In this passage, Kant states “Time is nothing other than the form of inner sense, i.e., of the intuition of our self and our inner state”. Again, in B55, he says that the object of inner sense is the object of myself and my state. Kant distinguishes two critical roles for the time in this regard. The first concerns the subject's direct intuition of itself through the inner sense. Time, as the form of the inner sense, enables this immediate self-intuition. The second role involves the subject intuiting its own representations or mental states. Here, the subject's capacity to be affected by the specific determinations of these inner states allows for their recognition as belonging to the self.

In general, the inner sense has no separate manifold of its own. The matter of inner intuitions consists of no sensory impressions produced by or grounded in the self, much in the same way that the matter of outer intuitions consists of impressions produced by the affection from outer objects. According to Kant, there is no empirical intuition of the self as an external object, since the self cannot be encountered as a spatio-temporal object. For this reason, the manifold of the inner sense properly consists of *outer sense* representations: “the representations of outer sense make up the proper material with which we occupy our mind” (B67). As a result, the inner sense regarded as the limiting condition through which the empirical manifold given in outer sense may be attained: “I exist as an intelligence . . . which, in regard to the manifold that it is to combine, is subject to a limiting condition that it calls inner sense” (B158) (Forgione, 2019, 28).

As we have seen, there are two functions for the inner sense: 1- the intuition of the self, and 2- the intuition of the representations. Hence, if the subject intuites itself solely through the inner sense,

without any self-object that originates or causes the sensory impression, then the appearance of the self can only be produced by the empirical manifold presented in the inner sense. However, since the manifold of the inner sense is composed of representations from the outer sense, the cognition of the self as appearance, as argued by Kant in various passages, can only be realized through the introspectively available representations, essentially derived from the outer sense.

Kant makes a clear distinction between the 'I' as 'I think', 'intelligence', or 'thinking subject' in transcendental apperception, and the 'I' that intuits itself as two separate aspects of the same mind or subject. The 'I' as intelligence recognizes itself as an object of thought, since the subject is also presented to itself in intuition, along with other phenomena, "not as I am for the understanding but rather as I appear to myself". However, the subject presented to itself as appearance in intuition should be seen as the outcome of the combination or determination of the inner sense by understanding (Forgione, 2019, 29).

The faculty of inner sense is characterized as passive, whereas understanding is regarded as an active faculty. In B153, Kant warns that we, like rational psychologists, should not confuse these two faculties with different natures:

Here is now the place to make intelligible the paradox that must have struck everyone in the exposition of the form of inner sense (§ 6): namely how this presents even ourselves to consciousness only as we appear to ourselves, not as we are in ourselves, since we intuit ourselves only as we are internally affected, which seems to be contradictory, since we would have to relate to ourselves passively; for this reason it is customary in the systems of psychology to treat inner sense as the same as the faculty of apperception (which we carefully distinguish).

According to Kant, the mind is perceived through our inner sense or empirical apperception, which allows us to be conscious of the continuous flow of inner appearances related to the state of the self. According to both Hume and Kant, there is no abiding self in this process. The self that is spoken in psychology, known as the 'phenomenal self', empirically recognized as a series of mental states unfolding in time, with time being the inherent form of our inner sense as Kant stated. This, distinct from the 'noumenal or transcendental self', which is the enduring self that truly exists.

The phenomenal self comprises exclusively of the empirically observed self states within classical psychology, which are uncovered through introspective examination of internal states or experiences. On the other hand, the noumenal self represents a nonempirical "limiting concept" that reason guides us towards through an analysis of our phenomenal self (Chessick, 1980, 461).

However, Kant had a plan to introduce noumenal self, as an abiding self. In the next section, we probe the Kant's view about this abiding self.



### 3. Noumenal Self and Apperception

There is a crucial distinction in Kant, which separates two realms ontologically: phenomena and noumena. According to each realm, there is a particular idea of the self. As we have seen, the phenomenal self is an empirical self that associate with the faculty of inner sense and intuition, but, in Kant perspective of the self, there is another concept of the self that has different properties and associate with the other faculties. This, called ‘noumenal or transcendental self’ that has a clear link to the faculty of understanding.

Kant posits that the mind is comprised of two essential faculties, specifically 'receptivity' and 'spontaneity':

If we will call the receptivity of our mind to receive representations insofar as it is affected in some way sensibility, then on the contrary the faculty for bringing forth representations itself, or the spontaneity of cognition, is the understanding. It comes along with our nature that intuition can never be other than sensible, i.e., that it contains only the way in which we are affected by objects. The faculty for thinking of objects of sensible intuition, on the contrary, is the understanding. Neither of these properties is to be preferred to the other. Without sensibility no object would be given to us, and without understanding none would be thought (A51/B75).

Kant proceeds to introduce three distinct faculties that further elaborate on the general aspects of the mind. These faculties are ‘sensibility’, ‘understanding’, and ‘reason’. By delineating these three specifications of receptivity and spontaneity, Kant provides a more comprehensive understanding of the workings of the human mind:

The genus is representation in general (*repraesentatio*). Under it stands the representation with consciousness (*perceptio*). A perception! that refers to the subject as a modification of its state is a sensation (*sensatio*); an objective perception is a cognition (*cognitio*). The latter is either an intuition or a concept (*intuitus vel conceptus*). The former is immediately related to the object and is singular; the latter is mediate, by means of a mark, which can be common to several things. A concept is either an empirical or a pure concept, and the pure concept, insofar as it has its origin solely in the understanding (not in a pure image of sensibility), is called *notio*. A concept made up of notions, which goes beyond the possibility of experience, is an idea or a concept of reason (A320/B377)

Representation can be categorized into three main types: ‘sensations’, ‘intuitions’, and ‘concepts’. Sensibility, which refers to the faculty of sensory representations, encompasses

sensations and intuitions. These sensory representations are influenced either by external entities or by the subject itself. On the other hand, the faculty of understanding is responsible for dealing with conceptual representations spontaneously. Reason, another spontaneous faculty, concerns concepts known as 'ideas'. It is important to note that these ideas, such as God and the soul, cannot be encountered through direct experience (Forgione, 2019, 23).

In the prominent passage, Kant states that what he calls 'I think' must accompany all my representation:

The I think must be able to accompany all my representations; for otherwise something would be represented in me that could not be thought at all, which is as much as to say that the representation would either be impossible or else at least would be nothing for me. That representation that can be given prior to all thinking is called intuition. Thus, all manifold of intuition has a necessary relation to the I think in the same subject in which this manifold is to be encountered. But this representation is an act of spontaneity, i.e., it cannot be regarded as belonging to sensibility. I call it the pure apperception, in order to distinguish it from the empirical one, or also the original apperception, since it is that self-consciousness which, because it produces the representation I think, which must be able to accompany all others and which in all consciousness is one and the same, cannot be accompanied by any further representation. I also call its unity the transcendental unity of self-consciousness in order to designate the possibility of a priori cognition from it (B131-132).

In this context, Kant establishes a connection between the 'I think' and the concepts of spontaneity, as well as the pure or original apperception or self-consciousness. Kant, then, presents a principle under that all manifolds can be united in one consciousness:

I can combine a manifold of given representations in one consciousness that it is possible for me to represent the identity of the consciousness in these representations itself, i.e., the analytical unity of apperception is only possible under the presupposition of some synthetic one... Synthetic unity of the manifold of intuitions, as given a priori, is thus the ground of the identity of apperception itself, which precedes a priori all my determinate thinking... only an operation of the understanding, which is itself nothing further than the faculty of combining *a priori* and bringing the manifold of given representations under unity of apperception, which principle is the supreme one in the whole of human cognition... Now this principle of the necessary unity of apperception is, to be sure, itself identical, thus an analytical proposition, yet it declares as necessary a

synthesis of the manifold given in an intuition, without which that thoroughgoing identity of self-consciousness could not be thought (B134-135).

In (A77/B103) Kant talks about synthesis that “By synthesis in the most general sense, however, I understand the action of putting different representations together with each other and comprehending their manifoldness in one cognition”. According to Kant, the term 'representation' encompasses two meanings: it refers to both the 'mental state' itself and the 'content' of that mental state. The latter aspect pertains to the combination of various mental states' contents into a subsequent mental state. Kant recognized that individual mental states occur episodically at different times, making it impossible to assemble them together.

Kant also firms about these facts that we are unable to know the things-in-themselves:

...in any case we would still completely cognize only our own way of intuiting, i.e., our sensibility, and this always only under the conditions originally depending on the subject, space and time; what the objects may be in themselves would still never be known through the most enlightened cognition of their appearance, which alone is given to us (A43).

The faculty of understanding, according to Kant, generates the synthetic unity of thought, which is an aspect of the mind that cannot be directly known. This synthetic unity of thought is closely linked to the unity of consciousness, the conscious sense of self, or what Kant refers to as the 'I am I.' This unity of consciousness is an integral component of all thinking processes and is necessary for the synthetic unity of the manifold, as described by Kant. Consequently, our comprehension of objects does not arise from a disjointed series of sensory representations, but rather is presented to us in a coherent and interconnected way. The faculty of understanding has the responsibility of presenting the synthetic unity of representation, in addition, these experiences must be perceived as belonging to a singular consciousness, commonly referred to as the 'I' or the self. Kant refers to this as the *synthetic unity of apperception*. The central argument in Kant's immensely challenging yet essential 'transcendental deduction of the pure concepts of the understanding' that stated in Deduction, relies on the premise of the transcendental unity of self-consciousness, the awareness of 'I am I', the cohesive perception of a continuous and unified self. This enduring core of self-consciousness is undeniably necessary to differentiate one's self-boundaries and self-experiences from those originating from the external world (Chessick, 1980, 459).

Kant posits that we can contemplate the external reality and the knowing self, but we can never directly comprehend or describe them. The unknown self, also known as the noumenal or intelligible self by Kant, influences the mind internally and contributes to our introspective empirical sense of self experienced over time. This unknown self is believed to be the cause of our introspective experiences of self states, forming the phenomenal self. Therefore, Kant's

fundamental unproven assumption is that there is an underlying, unknowable, enduring 'I' that underpins our mental activities.

The concept of the noumenal self, acting as a constraining concept, is closely related to our experiences, as it is directly inferred by our reason through our experiences. It serves as a *regulative* principle according to Kant, aiding reason in organizing and categorizing our experiences of the phenomenal self. Since it is solely a product of rational thought, no further elaboration can be provided. Kant refers to this notion of the noumenal self as the *negative* noumenal self.

Although Kant, in *Critique of Pure Reason*, describes the self as a negative concept, however, in *Critique of Practical Reason*, he ignored entirely it and gives a positive one. For Kant, the self achieves its highest form in the realm of morality. The Categorical Imperative, the cornerstone of his ethics, demands that we act on principles that could be universally willed. This, requires us to see ourselves not just as self-interested individuals but as rational beings capable of acting on reason alone. Allen Wood (1999) expresses that Kant's emphasis on the self as a free and autonomous being, bound by the moral law, is a defining feature of his ethical philosophy.

In Kant's Moral philosophy, the noumenal self is utilized as an autonomous entity, with many assumptions made about its nature. This transition is commonly characterized as a progression by Kant, moving from the negative connotation of the noumenal self to a more positive interpretation. This evolution involves a shift from a simplistic understanding of the noumenal self, derived from direct empirical observations to a more intricate and abstract concept, a transition that cannot be adequately supported within Kant's own philosophical framework outlined in the *Critique of Pure Reason* (Chessick, 1980, 461).

In the next section, we aim to evaluate and compare Kant's view on the self.

## **4. Evaluation and Comparison**

### **4.1. Evaluation**

Kant's view of the self has been subject to various criticisms. Some argue that the transcendental self is too abstract and provides little insight into the lived experience of the self. Others criticize the strict divide between the phenomenal and noumenal self, questioning whether the self-in-itself is truly unknowable. Patricia Kitcher (1982) critiques Kant's account of personal identity, arguing that it fails to capture the diachronic unity of the self. Kant's conception of the self is a complex and debated topic. Despite the challenges, his emphasis on the active, unifying role of the self in constituting experience, the distinction between the transcendental and empirical self, and the centrality of the self in morality, continue to influence contemporary discussions on selfhood.

As we have seen, Kant encountered the philosophical dilemma that arose from Hume's work, which ultimately led to an insurmountable obstacle, leaving only room for absolute skepticism. In an attempt to overcome this unsatisfactory conclusion, Kant redirected the focus of philosophy and psychology away from the empirical aspects of 'reality' towards the exploration of what could be

known or unknown beyond this perceived reality. In conjunction with Hume, Kant asserts that the phenomenal or empirical self, which arises from inner sense, cannot be considered enduring. However, to refute Hume's perspective, Kant introduces a noumenal or transcendental self that is indispensable for any form of experience. Without this transcendental self, the very possibility of experience is rendered unattainable.

According to Kant, the noumenal self has an internal impact on us, generating inner sense information. This inner sense data is then processed by the understanding, specifically through the "transcendental synthesis of the imagination," resulting in a consciousness of the self and a perception of how we appear to ourselves. There are two types of empirical experiences of the self: "1) 'I am I', which is required for all thought-the synthetic unity of apperception, and 2) self states appearing in a unified manifold in time" (Chessick, 1980, 460).

Kant employs the concept of the noumenal self as a regulatory notion, an 'idea of reason' that serves as a constraint. Kant dismisses "pure rational psychology" as a spurious field of study, as it attempts to examine the noumenal self as some form of 'substance'. According to Kant, the only viable approach is an empirical psychology, which entails an introspective examination of one's internal states as they manifest in the phenomena of our introspection, in contrast a purely rational psychology concerning the noumenal self underlying these experiences is deemed unattainable.

Accordingly, if the rational psychologist's argument regarding the thinking subject were accurate, we would possess a metaphysical understanding of the noumenal or things-in-themselves, like simple substance. This manner challenges Kant's limitation of metaphysics to experience in the *Critique of Pure Reason*. In the Paralogisms, Kant not only refutes the psychologist's argument, but also establishes what can be appropriately inferred about the thinking self. This affirmative inference is believed to align with his prohibition against knowledge of the noumenal. Peter Strawson believes that we can separate Kant's doctrine, transcendental idealism, and however reject the rational psychology: ". . . it is an important point that the force of Kant's exposure of the illusion of rational psychology can be carried out without any reference to these doctrines" (Strawson, 2007, 170).

Kant grappled with the concept of the inner self and endeavored to make a distinction between what can be known and what cannot be known about the self. According to Kant, there exists an unknowable self at the core of each individual that has the power to impact both the inner and outer experiences significantly. This noumenal self not only triggers these experiences but also shapes how they perceived. Due to the presence of this unique noumenal self or unconscious in every person, individuals may have vastly different reactions to themselves and the world around them, despite all being exposed to the same external stimuli on a biological or physical level.

Some philosopher probes the subject's separability from its materiality, namely from its body. Kant in the fourth Paralogism says that:

I distinguish my own existence, that of a thinking being, from other things outside me (to which my body also belongs) - this is equally an analytic proposition; for **other** things are those that I think of as **distinguished** from me. But I do not thereby know at all whether this consciousness of myself would even be possible without things outside me through which representations are given to me, and thus whether I could exist merely as a thinking being (without being a human being) ... It would be a great, or indeed the only stumbling block to our entire critique, if it were possible to prove a priori that all thinking beings are 'in themselves simple substances, thus (as a consequence of the same ground of proof) that personality is inseparable from them, and that they are conscious of their existence as detached from all matter (B409).

According to Kant, in terms of what can be inferred from the cogito, would be evident. Descartes argued that through reflection on the cogito, he distinctly perceived that as a thinking entity, he could exist independently from his body.

Discussion about bodily identity takes place in the B edition of *Refutation of Idealism* (B275-279). Kant argues that my existence is inherently tied to my ability to be influenced by outer factors, as well as to the experience of being influenced by outer factors. Contrary to proponents of psychological continuity views, Kant argues that I have a capacity for being a continuous and genuine holder of inner states. Additionally, Kant rejects the notion of bodily or organism identity, which posits that a person's identity is tied to the identity of their organized body, which possesses the ability to have an inner perspective. Unlike this perspective, Kant maintains that while embodiment is essential for me, bodily identity in any form does not contribute to my identity.

Arthur Melnick (2009) puts Wilfred Sellars and P.F. Strawson in one side, and John McDowell in the other side about bodily identity. According to Melnick, one perspective posited by Wilfrid Sellars and Strawson argues that Kant must fully embrace bodily identity, rather than just embodiment. Conversely, another viewpoint advocated by John McDowell contends that Kant falls short of even reaching embodiment. Sellars argues that Kant is compelled to adopt the position that I exist as a phenomenal (material) substance due to the implications of the First Analogy. According to Sellars, this analogy demonstrates that any alteration in state, including inner states, must be attributed to the transformation of an enduring intuitable substance (Sellars, 1970, 12). Peter Strawson (2007) argues that attributing states to oneself is only possible when contrasted with an objective spatial world. According to Strawson, our states form a pathway within this world, and for this to occur, it is necessary for us to be a material body that can be identified among other bodies.

McDowell has different ideas on materiality. Accordingly, once we allow that I am an ongoing capacity for inner attending and having thus a point of view, we must go on to hold that I am not essentially embodied at all:

If we insist on supplying this self-awareness with an object, we can locate the object in the world only geometrically, as a point of view...if we start from a putative sense of self as a most geometrically in the world, how can we work up from there to the sense of self we actually have, as a bodily presence in the world? (When I say that is what self-awareness is, I am not implying that one's bodily presence in the world is always borne in on one in self-awareness.) (McDowell, 2000, 104).

However, according to Kant, having a body for receptivity is essential, but it could not be part of my identity as a bodily one. Nevertheless, here we have a tension between B edition of Refutation of Idealism and B edition of forth Paralogism.

Kant in B409, is obviously referring to the B edition Refutation of Idealism for an answer to the question of separable existence (the question, recall, not fully answered in the A edition Fourth Paralogism). His answer there, contrary to McDowell, is that I (the capacity for unifying inner attending or for having a geometric point of view) must be an embodied perceiver. We have then, on the one side, Sellars and Strawson contending that Kant has to go all the way to my being identical to a body (of a particular kind 'person') and McDowell on the other side contending that Kant can't get as far as being embodied at all (whether or not such embodiment is part of my identity) (Melnick, 2009, 152-153).

So far, all the perspectives we have analyzed have regarded the notion of self in Kant as negative. However, scholars have also endeavored to portray a favorable outlook for self, based on Kant's perspective, employing the metaphysical definition. Collin Marshall (2010) attempts to give a positive metaphysics according to its definition in contemporary analytic philosophy. Accordingly, "... metaphysics of the self would be some account of how selves are individuated, what sort of entities they are, and how they fit into our broadest conception of reality" (Marshall, 2010, 2).

To summarize, it seems that Kant's analysis of the self indicates his belief that, if it exists, it must transcend the realm of appearances and belong to the realm of things in themselves. Given that this realm can only be understood a priori, a metaphysical exploration of the self would necessitate an a priori investigation of a thing in itself. However, Kant's fundamental tenet asserts that we lack any knowledge of things in themselves. Kant in B158 explicitly states that "I therefore have no cognition of myself as I am, but only as I appear to myself." Marshall gives a more and less radical version of this argument. Based on Kant's assertions regarding the boundaries of the categories, one can argue that our understanding of the self is limited. In this particular rendition of the argument, not only do we lack knowledge about the metaphysics of the self, but we are also

incapable of representing any factual information about the self as it truly exists. A less extreme version of this argument centers on Kant's ideas concerning the limitations of our cognition. This version is most effective when we assume that Kant's central proposition, which states that we have no cognitive understanding of our true selves, implies that we possess no knowledge about any aspects of our selves beyond their outward appearances (Marshall, 2010, 4).

Marshall highlights the grounds that form a basis for questioning these assumptions. The first ground is that, Kant's argument that we lack knowledge of ourselves beyond our appearances is undermined by his own assertions regarding our existence extending beyond mere appearances. For example, he argues that our sensations are caused by external objects impacting our minds (see A19/B33) and that the synthesis of representations can only be carried out by the subject itself (see B130). Kant's assertions regarding cognition (*Erkenntnis*) provide a second basis for doubting the assumption, he consistently employs this term to convey specific limitations. Kant explicitly states that 'cognition' is a technical term in the Deduction, where he defines it as the "determinate relation of given representations to an object" (B137, also B146–74). Additionally, in the Introduction to the B edition, he elucidates that the cognition production entails comparing various sensory representations. These passages present compelling grounds to question any hasty assimilation of Kant's statements about cognition with our conventional knowledge in our sense. Kant's third ground pertains to natural language. He does not assert that we lack knowledge or cognition of any facts regarding things in themselves; rather, he posits that we lack knowledge/cognition of the things in themselves. This distinction is significant because in ordinary discourse (in both German and English), the phrase "S does not know X" is commonly employed in situations where S does possess some knowledge about X (Marshall, 2010, 5).

Referring to the view of some commentators (e.g., Waxman, 1991, 276–77, Ameriks, 2000, 5), Marshall states that Kant's revolutionary 'Copernican' method in philosophy is rooted in the idea of a determinate mind with specific faculties. In the A edition, Kant asserts that "Thus we ourselves bring into the appearances that order and regularity in them that we call nature, and moreover we would not be able to find it there if we, or the nature of our mind, had not originally put it there" (A125). The project's validity hinges on our recognition of the mind as a genuine entity with its own inherent nature, which is more fundamental than the world of appearances. Furthermore, it is crucial that we must possess a priori knowledge of certain aspects of this nature. Kant's elucidation of the "logical" characteristics of the self or self-consciousness (e.g., A350, B413) effectively captures this notion. He describes 'transcendental logic' as the discipline that aims to "isolate the understanding ... and elevate from our cognition merely the part of our thought that has its origin solely in the understanding" (A62/B87). In the *Critique*, Kant investigates the self or mind as the object of his inquiry, indicating that he must have possessed some affirmative and positive concept of what he was delving into.



One of the crucial notions in Deduction, according to Marshall, is the “unity of apperception”. Kant depicts it as “the synthetic unity of apperception is the highest point to which one must affix all use of the understanding, even the whole of logic and, after it, transcendental philosophy” (B134n). Kant, somewhat more intelligibly, also describes it as the “identity of self-consciousness.” Kant argues that this unity is deemed as “necessary” in a certain manner (for example, A107, B142), and it is intrinsically linked to our basic notion of an object (for instance, A109, B139). He also posits that synthesis guided by concepts and/or judgments are the mechanisms through which cognitions are unified (such as A111-12, B141), and that we possess an a priori conception of this unity when applied to the manifold of sensible intuition (e.g., A118, B150). Accordingly, it is crucial to note that unless Kant is introducing a new philosophical concept or attributing self-consciousness to an external entity, it is difficult to comprehend the notion of “identity of self-consciousness” across various representations without interpreting it as a unified conscious self, present in all representations (Marshall, 2010, 7-8).

Therefore, as we have already seen, Marshall proposes a positive metaphysics of the self. However, this problem can be more complex, according to the Kant description about distinction between noumena and phenomena. According to Broad (1978, 234), it is argued that Kant's explanation of the human self and its understanding of itself is highly intricate, and there is uncertainty regarding the possibility of extracting a unified and coherent doctrine from his diverse statements. Kemp Smith emphasizes that the self is the only origin of all unity: “As a pure and original unity it precedes experience; to its synthetic activities all conceptual unity is due; and by reflection upon the constancy of these activities it comes to consciousness of its own identity” (Kemp Smith, 1962, 208).

A selection to Kant could involve an intellectual intuition faculty capable of directly perceiving the noumenal self and noumenal objects, similar to Husserl's (1913) ‘eidetic intuition’. This would suggest a favorable interpretation of noumena, referring to transcendent objects and a transcendent self that can be directly understood. Therefore, when noumena is used positively, it indicates that the reality or entities beyond our experiential world can be defined, structured, and assigned functions (Chessick, 1980, 460).

After assessing Kant's perspective on the concept of self, we proceed to compare it with the theory referred to as the “Pattern Theory of The Self.”

#### **4.2. Comparing Kantian View and Pattern Theory of the Self**

Shaun Gallagher (2013) proposes the new theory of the self that depicts the self as a pattern. Accordingly, he notices that we should not confuse it with a mathematics one: “I do not mean to associate a pattern theory of self with “Pattern Theory” in mathematics” (Gallagher, 2013, 2). Gallagher utilizes the pattern theory of emotion to elucidate his own pattern theory. He argues that this particular pattern is an appropriate framework for explaining his theory due to its inherent

characteristics: “(1) it reflects a commensurable concept of the pattern (i.e., it refers to the same kind of pattern that I think is relevant to the notion of self, and (2) it may contribute directly to a pattern theory of self since, as Gallagher suggests, affect is one aspect that forms part of the pattern of self” (Gallagher, 2013, 2). The pattern theory of emotion posits that emotions consist of intricate patterns involving bodily processes, experiences, expressions, behaviors, and actions, thereby defining them as “individuated in patterns of characteristic features” (Izard, 1972). Pattern theory suggests that the concept of “emotion” is a *cluster* concept that encompasses a multitude of characteristic attributes. These attributes, when combined, give rise to a specific pattern that represents an emotion. However, it is crucial to acknowledge that no single attribute in isolation is sufficient to define an emotion. Instead, it is the collective pattern of characteristic features that constitutes an emotion. For the self, also, we can make it similarly. Gallagher wants to create a pattern consist of multi features, together makes a pattern:

In a way similar to the construction of a pattern theory of emotion, I want to suggest that we can develop a pattern theory of self. On such a view, what we call self consists of a complex and sufficient pattern of certain contributories, none of which on their own is necessary or essential to any particular self. This is not a pattern theory of “*the* self.” Rather, what we call “self” is a cluster concept which includes a sufficient number of characteristic features (Gallagher, 2013, 3).

Gallagher (2013, 3-4) proposes the eight kinds of features that make a pattern of the self. These are an aspect or set of aspect that constitute the pattern theory of the self:

- (1) Minimal embodied aspects: include here core biological, ecological aspects, which allow the system to distinguish between itself and what is not itself (Gallagher, 2013, 3).
- (2) Minimal experiential aspects: to the extent that the bodily system can be conscious, it will pre-reflectively experience, from a first-person perspective, the self/non-self distinction in the various sensory-motor modalities available to it... such aspects contribute to an experiential and embodied sense of ownership... and a sense of agency for one’s actions (Gallagher, 2013, 3).
- (3) Affective aspects: the fact that someone manifests a certain temperament may reflect a particular mix of affective factors that range from very basic and mostly covert or tacit bodily affects to what may be for her a typical emotional pattern or mood, for example, someone may be a typical extrovert who enthusiastically engages in outwardly directed actions (Gallagher, 2013, 4).

- (4) Intersubjective aspects: human and possibly some nonhuman animals are born with a capacity for attuning to intersubjective existence (Neisser, 1988); this may take the form of being aware that someone else is present and possibly gazing at you (Gallagher, 2013, 4).
- (5) Psychological/cognitive aspects: traditional theories of the self focus on these aspects, which may range from explicit self-consciousness to conceptual understanding of self as self, to personality traits of which one may not be self-conscious at all (Gallagher, 2013, 4).
- (6) Narrative aspects: although there are many variations of this idea, the basic claim is that selves are inherently narrative entities (Schechtman, 2011), and for some theorists, narratives are constitutive for selves (Gallagher, 2013, 4).
- (7) Extended aspects: James (1890) suggested that what we call self may include physical pieces of property, such as clothes, homes, and various things that we own. We identify ourselves with stuff we own, and perhaps with the technologies we use, the institutions we work in, or the nation states that we inhabit (Gallagher, 2013, 4).
- (8) Situated aspects: these are aspects that play some ... role in shaping who we are. They include the kind of family structure and environment where we grew up; cultural and normative practices that define our way of living, and so on (see Gergen, 2011) (Gallagher, 2013, 4).

According to the different aspects of the pattern theory of the self, Gallagher states that some of these may be more important than other, some may be, in a particular theory of the self, completely rules out, and there can be some different aspect of self constitution that we have not mentioned. Therefore, here, the self is not an entity that exists without worldly attributes. Most of these aspects are related to our everyday life and the world we inhabit. More important, the bodily features have a crucial one that constitutes the self. Now, if we compare this pattern theory of self with Kantian theory of self, how many of these aspects can be attributed to the Kantian self?

If we start from the phenomenal self, we can assign some of the most important aspects to it. Accordingly, the phenomenal self is the self that originates from inner sense and intuition, which is related to time and space. For this type of self, the intuition and faculty of sensibility are crucial. Therefore, this phenomenal self is inevitably dependent on the body and its capacity of receptivity. The configuration of our physical body, in conjunction with external stimuli, dictates the types of sensations that we are able to perceive and those that we are unable to. Accordingly, for the phenomenal self, we can ascribe the minimal embodied aspect. Conversely, this particular aspect cannot be attributed to the noumenal self, as it is a state that is not formed by the representation and intuition that are closely linked to bodily characteristics. Since, the minimal experiential aspects are related to what we experience in the world, and sensibility and intuition are related to the world we inhabit and experience, so that we can ascribe the second set of aspects to the phenomenal self; and in contrast, for this reason, this could not be attributed to the noumenal self.

Despite the potential connection between bodily condition and affective aspects, Kant did not extensively address these aspects when exploring the concept of the self. Thus, we could not find a considered relation between phenomenal self and affective aspects. It has a same situation when we probe this aspect and the noumenal self. Descartes posits that the mind or the self has the capacity to exist, thrive, and maintain its identity independently of its presence in the world. Kant, on the other hand, appears to overlook this particular dimension of worldly existence. According to Kant, the phenomenal self is solely concerned with physical entities and does not encompass the intersubjective realm. This notion can be extended to the noumenal self, which is unrelated to these particular facets. However, if we probe this aspect with the self that Kant provided in moral philosophy, this aspect can be entered to the set of features of the self. Through examining psychological and cognitive aspects, such as self-consciousness, it becomes evident that this particular characteristic plays a vital role in defining and shaping the concept of self in both the phenomenal and noumenal realms. Kant's ability to elucidate and assign meaning to the notion of self hinges greatly on these features. If we compare the Kantian phenomenal self to Hume's bundle theory in terms of narrative aspects, we can conclude that the self in Kant's perspective possesses a narrative feature due to its constitution solely of representations and intuitions. However, Kant argues that the noumenal self cannot be attributed in this way. According to Kant, the noumenal self is considered to be the things-in-themselves and serves as a necessary condition for all of our experiences. It is not constituted from a collection of intuitions. Ultimately, taking into account the extended and situated aspects, it is incapable of being attributed to both the phenomenal and noumenal self. However, it can be readily associated with the self that Kant portrays and elucidates in his moral and legal philosophy.

## **Conclusion**

The concept of the self, is one of the most complex concepts in the history of philosophy. In Kant's philosophy, the concept of self holds significant importance amidst the numerous complex ideas presented. As observed, Kant discusses two distinct kinds of the self, corresponding to the two faculties of the mind. The self can be divided into two aspects: the empirical or phenomenal self, which is associated with sensibility and inner sense, and the noumenal or transcendental self, which is connected to understanding and apperception. One can gain self-awareness and awareness of one's states only through inner sense, as they manifest within the subjective framework of time in the phenomenal world. The timeless noumenal self, conversely, exists outside of one's knowledge, together with its distinctive functions. The noumenal self is accountable for pure apperception, leading to a coherent self-awareness and a combined consciousness when encountering one's self-experience. The department of the noumenal self, also known as the productive imagination, carries out the synthesis processes. These processes are guided by certain innate concepts that are part of understanding faculty. Through this method, all experiences, whether they pertain to external

objects or our internal states, are systematically and cohesively presented. This enables a distinct separation between the "I" experiencing the phenomena and the external reality. Kant does not considerate the phenomenal self as a true one. Instead, he attempts, by introducing phenomenal self, to depict Hume's position on the self. According to Kant, the noumenal self is the prerequisite for enabling all experiences. However, due to the differentiation between phenomena and noumena, we are unable to comprehend any knowledge regarding the noumenal self. Accessing this realm is impossible for beings like us, as it pertains to the things-in-themselves. Nevertheless, Collin Marshall endeavors to portray a positive perspective on the self as he delves into the potentiality of the metaphysics surrounding the concept of the self. We have introduced the theory of the self named 'Pattern Theory of the self'. Accordingly, we investigate features of Kantian self in terms of features that Gallagher posited as a constituting feature of the self. Finally, we have examined the constituent characteristics of the Kantian self, encompassing the phenomenal, noumenal, and potentially moral self. As observed, certain aspects of these features are harmonious, while others are not.

## References

- Allison, H. E. (2001). *Kant's Transcendental Idealism: An Interpretation and Defense*. Yale University Press.
- Ameriks, K. (2000). *Kant's Theory of Mind*, 2<sup>nd</sup> Edition, Oxford University Press.
- Broad, C. (1978). *Kant: An Introduction*. Cambridge University Press
- Chessick, R. (1980). The Problematical Self in Kant and Kohut. *The Psychoanalytic Quarterly*, 49 (3), 456-473. <https://doi.org/10.1080/21674086.1980.11926923>
- Forgione, L. (2019). *Kant and the problem of self-knowledge*. Routledge, Taylor & Francis
- Gallagher, S. (2013). A pattern theory of self. *Frontiers in Human Neuroscience*, 7. <https://doi.org/10.3389/fnhum.2013.00443>
- Gergen, K. J. (2011). The social construction of self, in *The Oxford Handbook of the Self*, pp. 633–653, Edited by S. Gallagher, Oxford University Press.
- Hume, D. (2009). *A Treatise of Human Nature*, The Floating Press.
- Husserl, E. (1913). *Ideas. General Introduction to Pure Phenomenology*. Translated by W. R. B. Gibson. Macmillan.
- Izard, C. E. (1972). *Patterns of Emotions. A New Analysis of Anxiety and Depression*. Academic Press.
- James, W. (1890). *The Principles of Psychology*. Dover, 1950.
- Kant, I. (1998). *Critique of Pure Reason*. Edited & translated by P. Guyer, & A. W. Wood. Cambridge University Press.
- Kemp Smith, N. (1962). *A Commentary to Kant's Critique of Pure Reason*. Humanities Press.
- Kitcher, P. (1982). Kant on self-identity. *The Philosophical Review*, 91(1), 41-72. <http://www.jstor.org/stable/2184668>

- Kraus, K. T. (2020). *Kant on Self-Knowledge and Self-Formation: The Nature of Inner Experience*. Cambridge University Press.
- Marshall, C. (2010). Kant's Metaphysics of the Self. *Philosophers' Imprint*, No.10, pp. 1-21.
- McDowell, J. (2000). *Mind and World*. Harvard University Press.
- Melnick, A. (2009). *Kant's Theory of the Self*. Translated by Taylor & Francis, Routledge.
- Neisser, U. (1988). Five kinds of self-knowledge. *Philos. Psychol.* 1, 35–59.  
<https://doi.org/10.1080/09515088808572924>.
- Schechtman, M. (2011). The narrative self, in *The Oxford Handbook of the Self*, pp: 394–416, Edited by S. Gallagher, Oxford University Press,
- Sellars, W. (1956). *Empiricism and the philosophy of mind*.  
[https://conservancy.umn.edu/bitstream/11299/184083/1/1\\_11\\_Sellars.pdf](https://conservancy.umn.edu/bitstream/11299/184083/1/1_11_Sellars.pdf)
- Sellars, W. (1970). ...this I or He or It (The thing) which thinks... *Proceedings and Addresses of the American Philosophical Association*, 44, 5–31. <https://doi.org/10.2307/3129681>
- Strawson, P. (2007). *The bounds of sense: an essay on Kant's critique of pure reason*. Routledge.
- Waxman, W. (1991). *Kant's Model of the Mind: A new interpretation of transcendental idealism*. Oxford University Press.
- Wood, A. (1999). *Kant's ethical thought*. Cambridge University Press
- Zöller, G. (1993). Main Developments in Recent Scholarship on the Critique of Pure Reason. *Philosophy and Phenomenological Research*. 53(2), 445–466.



Journal of Philosophical Investigations

## Journal of Philosophical Investigations

Print ISSN: 2251-7960 Online ISSN: 2423-4419

Homepage: <https://philosophy.tabrizu.ac.ir>



University of Tabriz

# Grasping the Grounds of Thought: The Thing-in-Itself, Actancy and Ecology

Otto Paans 

Independent Philosopher, The Netherlands. Email: [ocpaans@gmail.com](mailto:ocpaans@gmail.com)

### Article Info

#### Article type:

Research Article

#### Article history:

Received 27 July 2024

Received in revised form

31 July 2024

Accepted 31 July 2024

Published online 07

August 2024

#### Keywords:

Noumena; Kantian  
Philosophy; Metaphysics;  
Anthropology; Actancy

### ABSTRACT

The Thing-in-Itself has been contentious issue within Kantian philosophy. Initially, it seems like an unfortunate side-effect of Kant's distinction between phenomena and noumena. This article deals with this issue in a different manner, attempting to re-situate the Thing-in-Itself within Kantian philosophy, albeit from an anthropological rather than a critical angle. The anthropological works of Kant fully recognize that subjectivity and lived experience, as well as a thoroughgoing cognitive gradualism are necessary to "orient ourselves in thinking". By reading the importance of the Thing-in-Itself from the anthropological viewpoint of Otto Friedrich Bollnow and the Kyoto School philosophy of Ueda Shizuteru, I argue that in all its negativity, the Thing-in-Itself constitutes the outer expanse of thought. Connecting this exposition with contemporary thinking on actancy and ecology, and following the Romantic tradition represented by Schopenhauer and Schelling, I argue that the Thing-in-Itself can be grasped indirectly and non-conceptually. As such, it constitutes the ground of thought. This insight makes Kant's initially problematic concept directly relevant for our current ecological predicament, through which we realize the necessity for epistemic humility and embracing the unknown or the noumenal dimension that we cannot conceptually represent.

**Cite this article:** Paans, O. (2024). Grasping the Grounds of Thought: The Thing-in-Itself, Actancy and Ecology. *Journal of Philosophical Investigations*, 18(47), 111-138. <https://doi.org/10.22034/jpiut.2024.62685.3833>



© The Author(s).

<https://doi.org/10.22034/jpiut.2024.62685.3833>

Publisher: University of Tabriz.

*[W]isdom too, true insight into life, the proper view and the apt judgment, stem from the way human beings grasp the intuitive world, and not from mere knowledge, i.e. not from abstract concepts (Schopenhauer 2018, 83).*

*The names, the shapes and forms we give Quality depend only partly on the Quality. (...) People differ about Quality, not because Quality is different, but because people are different in terms of experience (Pirsig 2006, 318–319).*

*Thus, profound metaphysics is rooted in an implicit geometry which whether we will or not—confers spatiality upon thought; if a metaphysician could not draw, what would he think? (Bachelard 1994, 212)*

## Introduction

Arguably one of the thorniest issues that Immanuel Kant bequeathed to modern philosophy is the nature and role of the Thing-in-Itself (*Ding an Sich*). It sparked already a row in the so-called 1782 “Göttingen review” of the first *Critique* and hasn’t stopped to be a philosophical stumble block ever since. Although the review by Feder and Garve largely mistook Kant’s transcendental idealism for Berkeleyan subjective idealism, some of their insistent criticism lingered.

During Kant’s later years and in the decades immediately following his death, the first generation of Neo-Kantians found themselves forced to either defend or reject the Thing-in-Itself. Conversely, the rising tradition of speculative idealism (represented by Fichte, Schelling and Hegel) tried to collapse the distinction between appearance and Thing-in-Itself, in favor of access to the Absolute or the Unconditioned. Schopenhauer took a different route and universalized it as the Will-to-Live, or the immanent drive behind the cosmic play of appearances. Later generations of Neo-Kantians either tried to integrate or politely ignore this issue, although it stubbornly lingered around in the margins of their philosophy.<sup>1</sup> For 20<sup>th</sup> century analytic and continental philosophy alike, the issue held limited importance outside Kant scholarship, where its reception ranged from trying to integrate the theory in the larger Kantian corpus to methodologically eliminating it. Recently, the rise of Object-Oriented Ontology, Onticology and Correlationism (Bryant 2011; Graham 2013, 2018; Meillassoux 2018) pulled it back into the center of philosophical attention, although the real issue is sidelined, and Kant is held to be just a philosopher of “human access to the world”.

Within Kant scholarship, we encounter so-called “one-world” readings, as well as “two-world” or “two-aspect” readings if we survey the available literature.<sup>2</sup> The one-world interpretation stresses that appearances and Thing-in-Themselves are in fact identical; the two-world reading

---

<sup>1</sup> Beiser (2014, 88) points out that Jakob Friedrich Fries’ affective approach relegated the access to the Thing-in-Itself to the realm of feeling or approximation (*Ahndung*). Likewise, Friedrich Albert Lange’s broadly materialist conception held that Things-in-Themselves formed a kind of “common denominator” that grounds cognition (2014, 380). By contrast, Alois Riehl’s realist conception stressed that Things-in-Themselves are dual-aspect entities that appear partially to the perceiving subject, but that can be considered apart from their form of appearance (2014, 565).

<sup>2</sup> See Allison 2001 and Marshall 2013 for extensive discussions of this literature.



postulates two different worlds, in which each phenomenal appearance *here* has a noumenal correlate over *there*. The two-aspect view advocates that *phenomena* and *noumena* are simply two aspects of a single object, disagreeing about the degree to which we can cognize or otherwise detect them. These efforts seem stuck in a loop: the crux lies in how one treats the phenomena/noumena binary. Viewed as the opposition of two radically opposed domains that can in principle not causally touch, we saddle ourselves with the irresolvable problem of maintaining the distinction while still finding ways to render these domains causally connected.

Amid all the philosophical noise that has arisen around this topic, a new possibility can be found – one that makes Kant’s troublesome inheritance directly critically relevant to what has been recently called “ecological thought” and “rational anthropology” alike. Moreover, I avail myself of the opportunity to align Kantian philosophy and the so-called “spatial turn” in philosophy and sociology. This approach has tangibly Kantian roots, as Kant himself lectured on physical geography from 1756 until 1796, on anthropology from 1772 until 1796, and wrote the essays *On the Ultimate Ground of the Differentiation of Regions in Space* (1768) and *What Does it Mean to Orient Oneself in Thinking?* (1786). Consequently, Kant’s thinking is not just concerned with formal distinctions and logical propositions as developed in his theoretical philosophy, but equally involves the study of the human being in space, or, embodied in an environment as espoused in his practical philosophy (Clewis, 2018).

This geographical undercurrent is especially present in the anthropological writings and in the lectures on physical geography. Although it is correct to attribute straightforwardly ecological views to Kant, his philosophy provides ample opportunity to build bridges between lived experience, ecology and “the movement of thinking”, or to construct a “rational anthropology” (Hanna, 2017a) for the Anthropocene or Symbiocene.

I examine the phenomenon/noumenon binary from a geographical and spatial – and by implication – anthropological perspective, moving away from readings that seemingly try to square a circle. What if we repositioned the issue, inscribing it in the practical philosophy of Kant, rather than in his theoretical works? We would acquire a fully Kantian account from our access to the world, this time not from a purely critical, but an anthropological angle.

### **1. The Anthropological Angle: An Argument**

preliminarily, I summarily state my overall approach. First, we should make a careful distinction between two different aspects of Kant’s overall philosophical project. Whereas notably the first *Critique* is concerned with delineating the “limits of sense and reason” as Kant said in his 1772 letter to Markus Herz (C 10, 129), the practical and anthropological works of Kant deal with how knowledge acquisition and growth develops in everyday life, notably in the social, moral and cultural domains. Put differently, the critical side of Kant’s philosophy deals with the conditions under which we can state we know X or Y, while the practical side deals with the process of how the gradually come to know or understand X or Y. Both sides are complementary: the process of

acquiring knowledge unfolds within the rules set out in the critical works. For good reason, the question “what can I know?”, and the question “what is the human being?” appear in the same series of basic questions that Kant regards as the mission of all philosophy (JL 9, 25) To understand the human being in its environment (*Welterkenntnis*) is to investigate how knowledge develops and how everyday practices support and direct this process.

To answer the question as to what the human being is, we require a shift in perspective. Instead of regarding the process of cognition from the viewpoint of criteria, we should – following the anthropological tradition – regard cognition as a living, organic process of real-time sensemaking. That this process has rules is clear. However, to understand how it unfolds we must invoke a first-person perspective instead of a non-perspectivist, critical-methodological one. While Kant’s theoretical philosophy deals with the *formal criteria* of objectivity, his practical philosophy deals with the *gradual process of arriving* at such objectivity (Paans, 2023).

The issue of access to the Thing-in-Itself can be framed as a strictly logical problem, if we follow the critical approach set out in the first *Critique* and the *Prolegomena*. But once we adopt *that* approach, we end up in the situation described above, attempting to circumvent a self-imposed problem. But once we frame it as an anthropological problem, we find that the issue of access to ultimate reality takes on a pragmatic and practical hue: the issue is not *what* we can know, but *how* we are able to access or apprehend a domain currently beyond our grasp, but that still entices us to venture further. Kant was aware of reason’s propensity to explore beyond its own limits:

It is true: we cannot provide, beyond all possible experience, any determinate concept of what things in themselves may be. But we are nevertheless not free to hold back entirely in the face of inquiries about those things; for experience never fully satisfies reason. (Pro 4, 351)

Thus, the representation of noumena as beings of the understanding (Verstandeswesen) is “not merely permitted but also inevitable”. (Pro 4, 315)

The lived space of experience is anthropological in nature, as Marc Augé (2006) already noted. It possesses experiential depth, as personal and collective identities, worldviews and thought-shapers (Hanna & Panns, 2021) are all forged within its broad, intimately environmental and cultural frame of references. By situating the issue of access to the Thing-in-Itself within anthropological space, we accomplish what A.N. Whitehead described as a “concrete analysis of intuitive experience”:

The point before us is that this scientific field of thought is now, in the twentieth century, too narrow for the concrete facts which are before it for analysis (...) Thus, in order to understand the difficulties of modern scientific thought and also its reactions on the modern world, we should have in our minds some conception

of a wider field of abstraction, a more concrete analysis, which shall stand nearer to the complete concreteness of our intuitive experience. (Whitehead 1960, 65)

Conceived as a form of concrete, lived, first-person and organically developing analysis, we solve the following by addressing the status of the Thing-in-Itself in an anthropological manner:

First, we liberate ourselves of a few strange and questionable historical assumptions about what can be and can't be thought and how this *limit-of-thinking* is to be conceptualized from the perspective of lived experience.

Second, we can bring the *geographical orientation* in Kant's thinking into sharper relief so that we can usefully apply to current world problems, in particular our ecological predicament.

Third, it offers an opportunity to extend Kant's *cognitive gradualism* in a new direction – one that has direct moral implications for individual experience, and indeed for the practice of creative piety.

I re-interpret Kant's initial and problematic notion from two unfamiliar vantage points, respectively the philosophical anthropology of Otto Friedrich Bollnow (1903–1991), as elaborated in his 1963 book *Mensch und Raum* and the Kyoto School philosophy of Ueda Shizuteru (1926–2019).<sup>1</sup> Before doing so, however, I examine (a) a selection of relevant textual evidence within the Kantian corpus that formulates the characteristics of the Thing-in-Itself and (b) demarcate a number of specific themes that surface later on.

Then, I provide a reformulation and position of the Thing-in-Itself with a recourse to the works of Bollnow, and Shizuteru, establishing bridges between the venturing nature of reason and Kantian cognitive gradualism. After doing so, I extend its range of application towards contemporary ecological thought. In a nutshell, I aim to show how a conceptual stumble-block in Kantian philosophy can be reinvigorated for its critical impetus and contemporary relevance.

## 2. Four Fundamental Notions

The fundamental Kantian distinction between things as they appear to us (*phenomena*) and things as they are in themselves (*noumena*) can be interpreted in various ways. However, a fundamental way in which the entire problematic can be framed is found in the very first line of the *Critique*:

---

<sup>1</sup> I used the Dutch translation of Bollnow's work (titled *Mens en Ruimte*). The English translation titled *Human Space* seems to miss the finer points of Bollnow's German, while the Dutch translation captures its nuances better due to the close relationship between the two languages. Regarding Japanese philosophy in general and the Kyoto School in particular, I fully admit that my treatment of Shizuteru's thought is limited to the translations available to me. Much work of the Kyoto School has not been translated, providing us with a limited picture of what these philosophers intended and achieved. Yet, there seems to be enough material to construct a relatively reliable picture. I cite Bollnow's work parenthetically, using the abbreviation MR, followed by the page number. Shizuteru's essay is cited parenthetically, using the abbreviation HOH, followed by the page number.

Reason has the peculiar fate in one species of its cognition that it is burdened with questions that it cannot dismiss, since they are given to it as problems by the very nature of reason itself, but which also it cannot answer, since they transcend every capacity of human reason. (CPR Avii)

Apart from pre-empting the Existentialist notion that the problems of humanity are partially caused by the very fact that it has the capacity to problematize its own existence, Kant raises a point about the nature of reason itself. Simply put, our reasoning strategies lead into certain unavoidable problems. Problems, moreover, that are so fundamental that dismissing them topples the entire edifice of thought built upon them. To avoid those problems, philosophers and scientists alike often revert to “wishful-and-fearful thinking”. (Griffin 1998, 11–14) The issue of the Thing-in-Itself is a splendid case study.

### 2.1. Kantian Cognitive Gradualism (KCG)

Kant’s simple formula is that for each appearance X, there must be some cause Y that causes it. But since the limits of our sense and reason demarcate what we can cognize (and thus claim to know), there is nothing left for us but to postulate some causally efficacious entity beyond our cognition, which furnishes our sensibility with *phenomena* or appearances.

That mysterious Y is the Thing-in-Itself (*Ding an Sich*) or *noumenon* that we cannot even truly know exists, save by inference to the principle that “for every effect, there must be cause.” (R, 18.353) So, we can postulate it as the logical termination of a casual chain.

Kant defines the position of appearances in his transcendental idealism as follows:

Everything intuited in space or in time, hence all objects of an experience possible for us, are nothing but appearances, i.e., mere representations, which, as they are represented, as extended beings or series of alterations, have outside our thoughts no existence grounded in itself. This doctrine I call transcendental idealism. (CPR A490–91/B518–519).

But Kant introduces quite some ambiguities as well as suggestive remarks here. First, appearances (*Erscheinungen*) are not the same as experiences (*Erfahrungen*). Kant seems to regard the latter as a more developed stage of the former. This is fully consistent with his cognitive gradualism, whereby judgements are gradually constructed through reflection. While such proto-judgements are not necessarily illusions, they are rudimentary, maturing over time. (AF 25:480; CPR B69–71; Pro 4, 293) This tells already something about Kant’s remarkably process-based (and anthropological) theory of cognition.

We have experiences, but these are not given as ready-made or finished impressions. We can relate to them in various ways and revise how we conceptualize or value them. This implies a certain critical distance: we have a choice to reflectively judge a certain appearance even if it is directly given to us, grounding our freedom and guaranteeing the space for moral action. (Pro

4:290; R, 17:635) To insist otherwise would rob agents of moral capacity and would lead straight into hard determinism. A familiar error in reasoning consists in taking initial appearances or proto-judgements for fully developed experiences. This may lead into illusions – or the faulty application of the categorial structures of understanding to intuitions. (A, 7:146) Importantly, Kant is concerned with the *objective validity of judgements*, a theme that is present throughout his entire critical philosophy. The distinction that Kant makes is that subjective judgements are *necessary* for reaching objective determination or universal assent. However, they are not *sufficient* to lay claim to objectivity. While privately, these subjective judgements are required to gradually form a judgement, they cannot claim “universal assent” or “objective determination.” Especially in the anthropological writings, we see that Kant stresses the importance of the subjective judgement, this time from the perspectivist, lived, first-person and indeed anthropological point of view.

The process of getting acquainted with our intuitions and what they imply for the understanding develops gradually. Kant frames this process as a combination of three syntheses in the first *Critique*, but as a process of reflection throughout the *Anthropology* and as a sequence (*progressus*) his *Reflexionen* (R, 18:379) but also reserves a role for the practical application of the faculty of reflection. This process plays out in real-time and is one of approximation and gradual clarification/exposition. I will have more to say on this later, especially in relation to the synthesis of the manifold through apprehension.

I call the process of (a) progressing from subjective to objective judgements and (b) the use of proto-judgements “Kantian cognitive gradualism” (KCG), and likewise, it will play a role later in my argument.

## 2.2. The Spatialization Argument of Cognition (SAC)

Kant notes that the Thing-in-Itself may well be an unintended or at least misleading side-effect generated by the understanding. The understanding poses a limit-concept of what it can reach. In doing so, it inadvertently also generates another object that is logically possible, but which is an “entirely undetermined concept”.

If, therefore, we wanted to apply the categories to objects that are not considered as appearances, then we would have to ground them on an intuition other than the sensible one, and then the object would be a noumenon in a positive sense. Now since such an intuition, namely intellectual intuition, lies absolutely outside our faculty of cognition, the use of the categories can by no means reach beyond the boundaries of the objects of experience. (CPR, B308)

An object which can be *thought* need not necessarily to *exist*. Kant calls this the “doctrine of the negative noumenon” – that is, an entity which is thought without any relation to our sensible intuition. (CPR, B307–308). To grasp the nature of a Thing-in-Itself, the understanding would have to apply the categories, which cannot be applied to the realm outside our cognition. While we can

have a consistent, logically possible conception of a Thing-in-Itself, this conception is necessarily misleading, as it applies the categories of the understanding in a domain where they hold no power.

What Kant indicates is crucial: the Thing-in-Itself represents the expanse beyond the boundary of cognition. The idea of such an outside expanse is already a stretch – we postulate it by means of a limit-concept, namely the *horizon* that separates the phenomenal and the noumenal. Whatever we may say about the “other side of the horizon” is then always already too much. We simply end up with “entirely underdetermined concepts”. Note also that it places a strict limit on the understanding, not necessarily on sensibility – if we did not have some causal connection to a deeper level of reality, we would be left with the question of how we could be affected at all (see also 3.4).

This insight by Kant that we must deal with two sides of an outer limit, I call the Spatialization Argument of Cognition (SAC). Like KCG, it resurfaces in due course of the argument.

### 2.3. Proto-Conceptuality

We can take Kant’s point about undetermined concepts one step further. That a completely undetermined concept is an impossibility seems obvious – it is a contradiction in terms. But, assuming the validity of KCG, what about *partially* undetermined concepts? As Kant points out:

Consequently, on the one hand, we cannot know the complete concept of any given object because we never dispose of all possible predicates; on the other hand, since each pair of opposed predicates has some content that is either affirmed or denied, we can conceive a priori of one thing whose concept contains all and only positive attributes. This “idea of an All of reality (*omnitudo realitatis*)” represents the concept of a most real being (*ens realissimum*), that is, “the concept of a thing in itself which is thoroughly determined”. [underlining added]. (CPR, A575–576/B603–604)

Here, we witness a Kantian move that possibly inspired Schopenhauer to equate the Thing-in-Itself with the totality of reality (*omnitudo realitatis*) as a universal Will. The “most real being” is that which contains all positive attributes – a kind of all-encompassing totality, the Unconditioned, the Hegelian Absolute, the Deleuzean virtual or Platonic realm of ideas. (Tomaszewska 2007, 79; CPR A702/B730) Kant borrows the Scholastic term *ens realissimum* – a notion usually reserved to designate the ultimate reality of the Divine.<sup>1</sup> To possess all the attributes of a concept would be to draw a map which is the territory. Concepts are instruments to navigate a reality that is remarkably more complex than what they describe. This full reality is a Thing-in-Itself or transcendental object,

---

<sup>1</sup> Indeed, Kant postulated that only a Deity could possess such an intellectual intuition as to survey the whole of reality, as it is not bound to the *Anschauungsformen*. See: Allison 2004 and Winegar 2017 for a discussion of this theme.

fully determined.<sup>1</sup> However, such an entity can only be thought as a regulative ideal, but cannot be perceived.<sup>2</sup> At this point, it becomes so all-encompassing that it can only exist as a logical possibility.

KCG implies a form of proto-conceptuality in Kantian philosophy. For now, it suffices to say that we may usefully extend our conception of concepts from linguistic labels to affect-laden, intentional, practicable, evolving and therefore adaptive cognitive instruments that allow us to navigate the world.

## 2.4. The Problem of Affection

Although Things-in-Themselves fall outside the scope of the innate structure of our minds (i.e. the principle of sufficient reason”), Kant still claims that they *affect* our cognition. This claim caused much philosophical headaches, as it involves a contradiction. There must be some causal connection between the Thing-in-Itself and our cognition if it is able to affect it. One could say here that the inconsistency lies in denying Things-in-Themselves any form of agency: either they affect us, and cause appearances to appear, or they do not. This is *problem of affection*.

Kant seems to defend the idea that the causal agency of Things-in-Themselves extends to our minds, but states that from observing these effects, we cannot learn anything about their constitution. (ML2, 28:567; CPR A42/B59) If Thing-in-Themselves are casually isolated from the world of appearances, how is it that our mind is affected by them, as this implies an efficacious causal connection? Kant seems to make at least two mistakes here:<sup>3</sup>

a) The categories cannot be applied to noumena, and as this applies to the category of “cause”, required for any form of causation, causality becomes impossible.

b) Causation implies a relation between two entities, but Kant says one of those entities is in principle unknowable, failing to provide proof that the “hidden member” is the actual *source* of causation.

We should discuss a second distinction here, namely between noumena in the positive and negative sense. (CPR, B306–309) A positive noumenon is an ideal and/or hypothetical object that can be only conceptually represented and is under that form intelligible. It is logically, although not metaphysically possible. A negative noumenon, on the other hand, is an empirical object considered without the cognitive frame that human cognition imposes upon it.

---

<sup>1</sup> Allison (2004) and De Boer (2014), distinguish in their discussions between and transcendental object, Thing-in-Itself and noumenon. I approach the Thing-in-Itself as *the* transcendental object, as per the passage of the First *Critique* I cited above.

<sup>2</sup> Tomaszewska points out that the Kantian conception of the transcendent realm ultimately terminated with the agency of God as the “prime mover” or ultimate Mind, a position not uncommon to the Enlightenment. At least, it conformed to the religious sentiments of the time, without having to participate in ongoing theological debates. See for a discussion of intellectual intuition and the possibility of knowing Things-in-Themselves Winegar 2017.

<sup>3</sup> See for an extensive discussion of this topic: Tomaszewska 2007



Although the latter possibility seems blocked from the viewpoint of Kantian transcendental idealism, an entire tradition of Eastern philosophy fundamentally diverges on this point.<sup>1</sup> Moreover, recently, emerging frameworks of thinking like Onticology and Actor-Network Theory have also critically interrogated this point.

The very least we can say is that “thinking a transcendental object” provides us with little or no information or knowledge at all. At best, it sets an outer limit for our thought, as SAC indicates. Yet, the problem of affection remains.

Summarizing, I have introduced four notions pertaining to the phenomena/noumena distinction. First, KCG says that judgements are *reflectively developed* over time. Second, the Spatialization Argument of Cognition (SAC) says that there is a discernable demarcation between the sensible and transcendental realms, although this need not be a sharp demarcation (in line with KCG). Third, proto-conceptuality means that concepts develop organically (also in line with KCG). Fourth, the problem of affection highlights to problematic causal relation between *phenomena* and *noumena*.

### 3. The One-World Interpretation and Phenomenal Monism: A Commitment

I pause for a moment to introduce a fundamental philosophical commitment, as well as a related recent contribution to the phenomena-noumena debate that comes close to my position, although it diverges in a few important aspects.

I am committed to *monist metaphysics*, rather than dualism or Deleuze-inspired metaphysics of multiplicity. This commitment derives directly from the doctrine of neo-organicism, some details of which have been worked out elsewhere (Hanna and Paans, 2020, Paans 2022). Following that doctrine, I reject two-world interpretations of the phenomena-noumena distinction, as that would invite an undesirable ontological dualism in.

My sympathies lie with a one-world/two-aspect interpretation of the phenomena-noumena binary, namely *phenomenal monism*:

There is one and only one class of apparent or phenomenal objects (i.e., one apparent or phenomenal world, hence phenomenal monism) such that each member of this class can be considered phenomenally or noumenally by us. (Hanna 2017b, 51)

The object should be taken in a twofold meaning, namely as appearance or as thing in itself. [underlining added] (CPR, Bxxvii)

The upshot of this variety of monism is that we can safely postulate that we can *think* noumena (as purely theoretical, logically possible) entities, while realizing fully well that we can never fully *cognize* or *consistently conceptualize* them. Moreover, there is only a single class of objects in the

<sup>1</sup> In the *Reflexionen* (R, 18:227) there is at least a passage that suggests some cognitive form of access to the Thing-in-Itself.



world. There are many ways to access or experience these objects, but that does not detract from the fact that there is no substance dualism of any kind. If there is any dualism (for instance explanatory dualism or cognitive dualism) involved, it has never to do with postulating more than a single type of object.

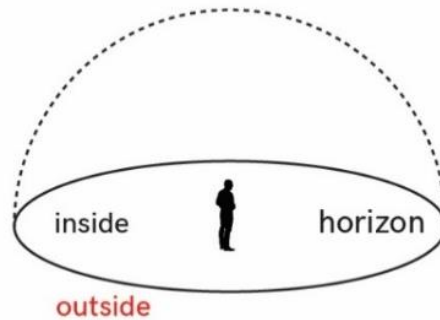
#### **4. The Noumenal Dimension and the Horizon-Structure**

A single important point remains: Kant distinguished between phenomena and noumena. Doing so, he drew a sharp line around the innate cognitive structure of the human mind. This boundary was well-demarcated, involving the *Anschauungsformen*, the categories and the rules of the schematism. But no matter how subsequent thinkers addressed this issue, the fact remained that a horizon had been postulated, dividing the world in an inside and an outside.

The concept of the horizon in relation to lived experience and cognition has been discussed in detail by Bollnow and Shizuteru alike. As Shizuteru builds his analysis on Bollnow's work, I discuss them in parallel, highlighting how their treatments illuminate each other. Arguably, Bollnow was one of the first post-war thinkers to deal with the relationship between human and space in such a detailed fashion. Granted, Heidegger (who was his teacher) had developed the notion of *Dasein* (Being), and conceived being-in-the-world as an inextricable condition for thinking about Being as such, or "the spatial scheme is in a metaphorical sense applicable to the entire human condition." (MR 88) His writings breathe the idea of the "spatiality of being" (Ireland, 2015), including meditations on the polarity between heaven and earth, dwelling and wandering. (MR 130–135; 182–185; 270–273)

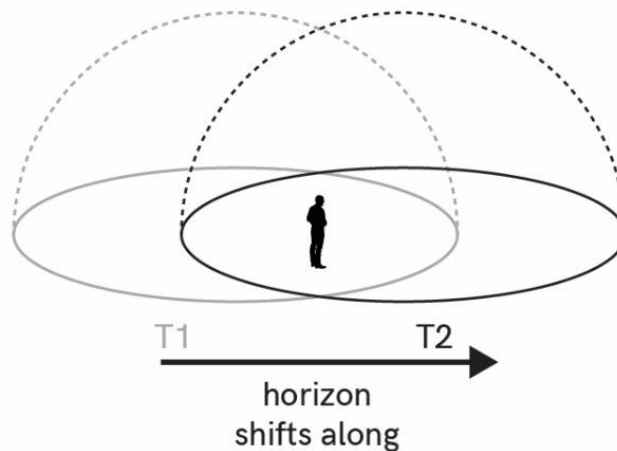
Moreover, Bollnow continues a line of geographical thinking that started with Ludwig Binswanger's 1942 *Grundformen und Erkenntnis menschlichen Daseins* and Gaston Bachelard's 1958 *La poétique de L'espace*, in which the existential import of everyday spaces was analyzed. This line of thinking constructs conceptual bridges between the structure of the environment and the structure of human thinking. In other words, the thought-shaping character of the environment is fully recognized. Bollnow makes it explicit when he states that "the horizon belongs to the essence (*Wesen*) of the human being". (MR 125)

As part of the environment, the horizon is one of the fundamental features of lived, human experience. (MR 118–122; 124–126) Its spatial presence is a necessary consequence of being a bipedal species with stereoscopic vision and the upright constitution of our bodies that it holds such an important place in our thinking. (MR, 91–96) What Kant called "the subjective ground of differentiation" (OT 8:135) is directly related to how individuals experience and cognitively construe a horizon. Within it, three-dimensional spatial cognition is coherently organized. (MR 121)



**Figure 1.** The horizon as limit and demarcation.

However, this boundary is not static, but moves along with the observer. So, it is a dynamic limit on what we can experience and perceive. What at one point lies outside of the horizon may be enveloped by it once we move the vantage point. (MR 121–122)



**Figure 2.** The horizon as an integral, dynamic element of cognition - once the perspective changes, the horizon changes along.

For Bollnow as for Shizuteru, there is a certain methodological advantage in creatively utilizing the literal, geographical and physical meanings of the term “horizon” alongside figurative and metaphorical interpretations. Indeed, it is safe to say that for both thinkers, these two modes of meaning overlap and illuminate each other. So, the visual limitation of the horizon is directly taken as analogous to the limit of our experience.<sup>1</sup> Bollnow borrows from Kant’s philosophy to emphasize how the subject postulates a horizon, and how irreducibly perspectivist one’s subjective position therefore is. (MR 121, 124) Likewise, Shizuteru interprets the horizon as the limit and

<sup>1</sup> Kant himself makes similar conceptual leap in his 1786 essay *What Does it Mean to Orient Oneself in Thinking?* fluidly joining logic, mathematics and anthropology in the course of the discussion.

opening of lived experience. Literally, philosophy becomes in the hands of these two thinkers a *cognitive geography*.

Shizuteru shares Bollnow's conception that the area within the horizon is unified by its presence – without horizon, a world would not be imaginable. (MR 121) For Bollnow, the horizon organizes cognitive experience, while for Shizuteru all “thing-events” take place within it. Importantly, this thought has been more recently taken up in ecological thought and correlationism – as Morton puts it “we have only access to thing-positings” (Morton 2018, 13). But whether one conceptualizes objects as positings or events, they derive their organization from the fact that they appear inside a horizon-structure. Moreover, the horizon is the organizational background for an overall *Gestalt*, or intelligible experiential structure. (MR 125) Without this background, any form of cognitive organization is spinning in the void, and even acting becomes impossible. It is against the background of a limit that an “action space” (*Handlungsraum*) emerges.<sup>1</sup> (MR 125, 274–279)

Both thinkers view the inside of the horizon as the domain of *phenomena*. What appears there necessarily appears within the confines of a prior demarcation by the horizon. Like Kant, Shizuteru emphasizes that *beyond* the horizon exists an expanse which is negatively demarcated for those residing within it.

Shizuteru introduces a way of looking at the inside/outside or phenomena/noumena distinction typical of Kyoto School philosophy. Kyoto School thinkers like Kitarō Nishida and Keiji Nishitani adopted a philosophical perspective predicated on the notion of Nothingness instead of Being, subtly subverting many notions that Western philosophy assumes to be foundational, thereby providing a new perspective.

Shizuteru adopts a similar strategy. For Kant, the noumenon represented a *negative* limit on human sensibility and understanding. It could only be postulated as an “outside” or alternatively “the other side of the (cognitive) horizon”. For Shizuteru, this feature is positive:

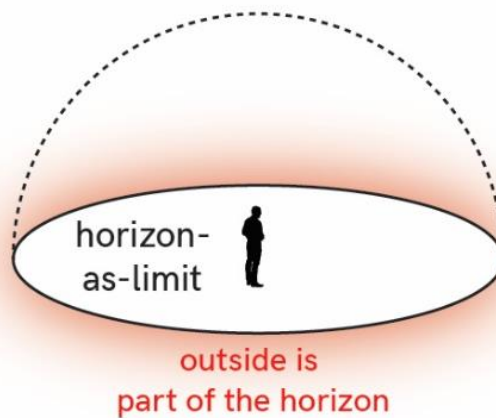
Is it possible to see that there is an area where we cannot see? So that we can see where we can see. (HOH, 97).

Is it not possible for there to be a seeing that there is an area we cannot see in our way of seeing what we can see and that it permeates? (HOH, 97)

This line of thinking inverts Western thought. Usually, we limit ourselves to what we can sensibly perceive – indeed, this is the upshot of Kant's transcendental idealism. It aims to with set the “limits of sense and reason”, thereby setting boundaries on what may be claimed as “knowable and setting metaphysics on the path of a “secure science” The Kantian question “what can I know?” is the most direct expression of this philosophical orientation. Shizuteru invokes a methodological contrast: to demarcate the area where we *can* see, we require precision about where we *cannot* see.

<sup>1</sup> Here, Bollnow is clearly indebted to Heidegger's notion of *Zuhandenheit*, invoking a relational theory of how objects appear to us in the space of phenomena.

*Prima facie*, this sounds vintage Kantian. But the added sting is that we have to see *the area where we cannot see*. Here, East and West part ways. The implicit claim is that there is a way of seeing that which we cannot see. Whereas the classical Kantian procedure demarcates the domain where we can see negatively by means of a limit-concept (the Thing-in-Itself), the Kyoto School procedure emphasizes that there is a continuity between the two domains. That which cannot be seen is nevertheless a part of what can be seen. So, to see where one cannot see is an integral part of seeing and indeed knowing.



**Figure 3.** The horizon as determined from both the inside and the outside.

Consequently, Shizuteru poses the question “whether there is a seeing to indicate an area that our way of seeing prevents us to see?”. The passive form (“a seeing”) derives directly from Buddhist philosophy. As per the classical dictum “the eye that sees cannot see itself”, the pure experience of seeing is regarded as more important than that which is discerned. This line of thinking is directly related to the thought of Shizuteru’s teacher Kitarō Nishida, who postulated the notion of “pure experience”. Put in Kantian terms, pure experience is a deep and thoroughgoing mode of experiencing before sensibility and understanding categorize, order and segment it. Buddhist thought spent considerable resources to detail how one could achieve such a deep and direct understanding. Before our customary frame of thought (up to and including the Kantian *a priori*) applies, pure experience exerts a deep, affective and lasting impact on individuals.

The correlated question that Shizuteru posits extends this idea. The issue is whether there is a way to transcend our customary way of seeing to see the domain where we cannot see. Simultaneously, Shizuteru suggests that we may also perceive a metaphysical continuity: the noumenal domain is not strictly isolated from the realm of phenomena, but “permeates” it. The wording is important: the influence of the noumenal is subtle, indirect and fragile – a dissecting,

analytic gaze quickly overlooks or disregards it.<sup>1</sup> The suggestion lingers in the air: can reason not grasp it because it is too rash, too direct, too forward, too much focused on the “determining judgement”?

I add some philosophical background to appreciate the plausibility of this viewpoint. From within the Kantian framework of thought, this is the issue that caused all the headaches. If the phenomenal and noumenal domains are causally isolated, Things-in-Themselves are merely thinkable, but their causal efficacy (i.e. the problem of affection) remains problematic.

Instead of the binary phenomena-noumena distinction, Japanese philosophy tends to favor a concentric model of cognition, with three concentric circles each signifying degrees of experiential intensity.

If we consider mental representations as theoretical cognitions of objects, as Hume and Kant both did, we place them either in the *field of sensibility* or in the *field of understanding/theoretical reason*. For Hume, this led to the skeptical conclusion that we are stuck with perceiving properties and inferring chains of cause and effects between them, without being able to justify our inferences rationally, but only by recourse to sensory experience and “habit” (Hume, 2008, 55–56, 66). For Hume, we are ineluctably stuck in the field of sensibility. For Kant, the Humean predicament was unpalatable, leading to a crucial refinement in how we conceive of sensibility and understanding/theoretical reason, but also to the ontological binary between *phenomena* and *noumena*.

Keiji Nishitani (1983) followed – in the footsteps of the Buddhist tradition and Nishida – a different path, enriching the taxonomy of experience with an additional step. We can imagine sensibility as the outer circumference of two concentric circles and understanding/theoretical reason as the inner one. We can refine and order sense impressions using understanding and/or theoretical reason, but we are ultimately unable to think objects apart from their Being.<sup>2</sup> But if we penetrate deeper into the heart of the concentric circle, we encounter objects in “the field of Emptiness.”

When we encounter objects in the field of Emptiness, we have left (Kantian) sensibility and its associated understanding/theoretical reason (*Erkenntnis*) well behind, and objects do not look like *objects* any longer, nor do they appear in a discursive space that we can analyze by using concepts or familiar scientific terminology.

Here, we arrive in the domain of *pure experience*, where concepts and words used theoretically or scientifically fall infinitely short of that pure experience itself. The only way we can speak meaningfully about them is obliquely or poetically.

<sup>1</sup> Once more invoking Bollnow: once we move from *Zuhandenheit* (ready-to-hand) to *Vorhandenheit* (present-at-hand), we lose the relational view through the analytical gaze. Importantly, Kitarō Nishida developed a similar line of thinking, arguing against what he called “object logic” or trying to overconceptualize or categorize experience.

<sup>2</sup> This view has a close correlate in Kant’s remark that the Thing-in-Itself (*ens realissimum*) is existence (R, 18.236)

By using the term “experience,” I am not speaking about the type of everyday experience, or the everyday Kantian notion of *Erfahrung*.<sup>1</sup> Instead, I refer to a unique and primordial level of experience that pre-empts, encapsulates, inflects and enables “experience” in the mundane sense. This type of experience is indeed “a seeing”, as the confines of subjective perception have fallen away or weakened at this point.

We can now appreciate Shizuteru’s reasoning. The question is whether we can bypass or circumvent the confines of the phenomenal domain, represented by the horizon-structure of experience. This strategy doesn’t refute or subvert the classical Kantian distinction but extends its area of application. It could even be argued that Kant suggests an affective or speculative access to the realm beyond experience:

By analogy, one can easily guess that it will be a concern of pure reason to guide its use when it wants to leave familiar objects (of experience) behind, extending itself beyond all the bounds of experience and finding no object of intuition at all, but merely space for intuition; for then it is no longer in a position to bring its judgments under a determinate maxim according to objective grounds of cognition, but solely to bring its judgments under a determinate maxim according to a subjective ground of differentiation in the determination of its own faculty of judgment. This subjective means still remaining is nothing other than reason's feeling of its own need (OT 8, 136).

Through “pure reason” that we explore this space for intuition without any object, using our “subjective ground of differentiation”. This process is the outcome of reasons need to venture beyond its boundaries. But, viewed from the anthropological angle, it is the underlying cognitive gradualism that is at work here: through a sequence (*progressus*) of successive judgements, objectivity is approximated.

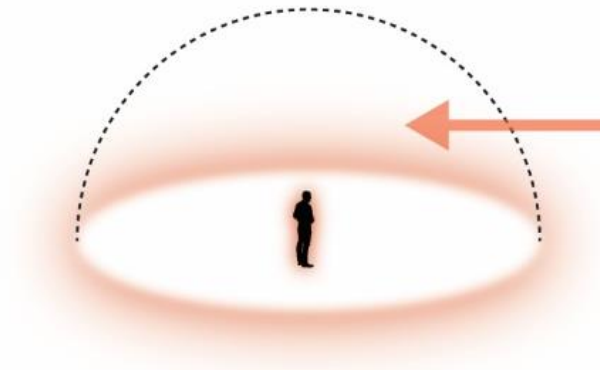
Shizuteru’s argument emphasizes *continuity* between phenomena and noumena. Again, this is a pervasive feature of Japanese philosophy: it emphasizes synthesis and continuity rather than decomposition and confrontation. In this case, this feature is used to “relax” an opposition rather than to emphasize it. (Jullien 2016, 120-122) This relaxation between opposites rejects the causal isolation usually associated with noumena, highlighting the problematic nature of the initial distinction. However, it does not simply *include* noumena within the realm of phenomena. The relation is one of indirect influence:

---

<sup>1</sup> Kant introduced the troubled distinction between judgements of perception and judgements of experience in his 1783 *Prolegomena to any Future Metaphysics*. Where “perception” represents for Kant the subjective observations, experiences etc., the notion of “experience” represents the systematic exposition of necessary features that can safely be ascribed to any object. In the Kantian framework, there is no explicit place for “the field of Emptiness”, save perhaps in aesthetic contemplation, the “free play of the imagination” or possibly in the speculation of pure reason.

Although nothing is added to the areas seen, would not a decisive, even if subdued, change occur in our comportment to the seen? (HOH 97)

Shizuteru raises the question whether the tacit presence of the “other side” of the horizon, once experienced, changes how things appear on “this side”. He suggests that a different “hue” or “tone” suffuses the most basic levels of our experience once we open up towards the noumenal realm.



**Figure 4.** The Shizuteruan proposal: the outside permeates the inside.

Elsewhere, I have written about the notion of “creative piety”, the attitude of consciously striving to experience the world in a different, livelier, informationally richer, humbler and more open sense (Hanna & Paans 2022).

Shizuteru’s suggestion closely resembles this notion, as it draws a link between cognitive faculties and the way in which things sensibly appear. While the cognitive faculties themselves might be usefully organized following the classical Kantian schema, the way in which things appear has to do with what these faculties can accomplish.

The critical point that can be raised here is that the problem of affection has been moved rather than solved. If “the other side of the horizon” is experienced by us, this means that noumena exert causal effects on our cognitive faculties and are *not* causally isolated. And if they are phenomenally experienced, that means that – all protestations to the contrary – noumena have been invited into the phenomenal realm. The rejoinder to this objection would be that nothing changes in the phenomenal domain. Things *appear* suffused to us, but nothing is added or subtracted. The hard-nosed critic might reply that *something* is causing the change and affecting the perception. If this *something* is a noumenon, then it turns into a phenomenon once we experience it. This reasoning conforms to the Kantian scheme: only phenomena can *appear* to us.

From the viewpoint of phenomenal monism, the distinction itself is already problematic, inviting ontological dualism where there doesn’t need to be one.

From the perspective of the horizon-structure of experience, it is equally problematic. The notion of the Thing-in-Itself is a limit-concept, as is the image of the horizon-structure. The

Shizuteruan response is that this limit presents also an opening – thereby extending our cognition. The objection that the noumenon becomes a phenomenon – and that this is problematic – denies the “relaxation of the opposites” its legitimacy. It also forecloses any consideration that the limits of our experience could be different than Kant imagined them to be. Tellingly, Kant understood every determination is a negation: “The person blind from birth cannot form the least representation of darkness, because he has no representation of light.” (CPR A575/B603) He realized very well that the outside is “part of the horizon”. Shizuteru emphasizes this: the horizon is negatively determined by what exists outside it. The “other side belongs negatively to the boundary”. (HOH 96)

If – following the broadly Kantian project – the aim is to investigate the structure of experience, we stop too early when we confine ourselves to what happens within the horizon-structure of experience. The outside of it is as much a part of it as the inside. (HOH 97).

The “possibility of the horizon is limited by the other side”, according to this argument. Shizuteru calls this model the “twofold world” (*nijū chichei*). This is not a form of dualism, but the simple recognition that we “can never know the whole”, just as Western philosophy held that the unconditioned or Absolute cannot be fully known or conceptually represented. To include non-knowing in the way of knowing results in a more informed and humbler picture:

That which is known by us can in no case be the whole. [B]y knowing this non-knowing (*fuchi*) together with knowledge (*chi*) on the basis of the twofold horizon, world knowledge is fundamentally relativized (HOH 99-100)

Sounding almost paradoxically, “the fact that we cannot know the whole becomes a way of knowing the whole” (HOH 100). Just like in the case of “seeing where we cannot see”, there is a way of experientially registering that which remains unknown. Once more, we see how Shizuteru uses this idea to emphasize a continuity between phenomena and noumena:

To know non-knowing means that experience percolates toward the unknown indeterminacy and at the same time the unknown indeterminacy permeates experience (HOH 100)

There is a two-way interaction between known and unknown. Experience “percolates towards” the unknown, while the unknown “permeates” experience. This is what Shizuteru means with things appearing “suffused” with a noumenal presence. All this sounds strikingly familiar from the viewpoint of Kantian non-conceptualism and creative piety. Experience in the full sense of the term encapsulates a forward impetus, a drive to know more and to explore the region that it knows to be outside its reach. As Kant himself well knew:

In fact, if we view the objects of the senses as mere appearances, as is fitting, then we thereby admit at the very same time that a thing in itself underlies them, though we are



not acquainted with this thing as it may be constituted in itself, but only with its appearance. (Pro, 4, 314–15)

Experience points ceaselessly beyond itself; but reason must come to terms with this continuous expansion, this grasping for new frontiers. Given this navigational character, is it possible to conceive the Kantian categorical framework as a toolkit for adaptively structuring our experience? Its structures cannot always make sense of that which is encountered; hence a form of cognitive gradualism is required. Inevitably, human reason seeks to come to terms with what it discovered; but equally, we cannot know the “whole” or “Absolute”. Consequently, wouldn’t it be better to view the categories of the understanding as a system of cognitive probes rather than a Procrustean grid imposed on sensibility? For Shizuteru, this orientation in thinking paves the way to frame the relation between phenomenon and noumenon not as one of ontological exclusion, but of *cognitive continuity*. The noumenal inflects the phenomenal domain.

## 5. Actancy and Grasping the Grounds of Thought

This line of thinking is not entirely unique – it has close correlates in the Romantic conception of the Thing-in-Itself. Both Schopenhauer and Schelling felt compelled to allow some form of access to “the ultimate ground of reality”.

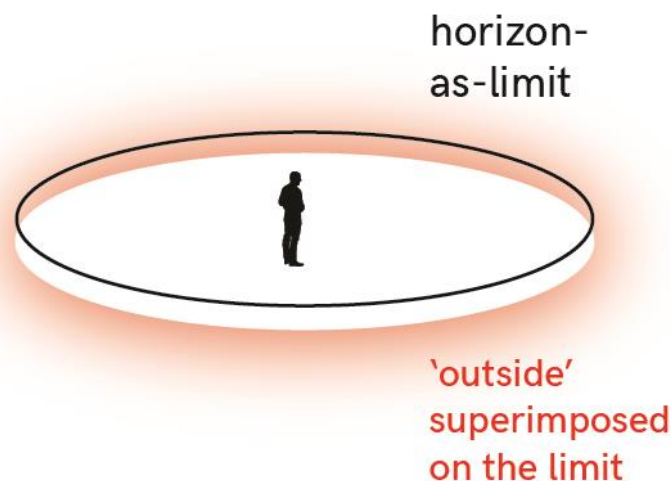
During the Romantic era, the strict phenomena/noumena distinction started to shift and evolve. The post-Kantian philosophies of Hegel, Schelling and Schopenhauer regarded the Thing-in-Itself not so much as a *formal* cause of appearances, but as a *ground* that existed outside the scope of representational thought. For Schopenhauer (2011, 174), this culminated in the “dark drive” of the Will, while for Schelling (2006, 68), it became an unrepresentable abyss (*Abgrund*) or non-ground (*Ungrund*). The *noumenal dimension* – in whatever way it was conceived became emotionally and existentially charged. It represented an experiential surplus that could be dimly apprehended, but not clearly conceptualized. Consequently, it became much more than just the logical upshot of Kantian transcendental idealism, evolving into an evocative description of a creative or generative realm giving rise to the manifestly real world that we perceive, but that itself remains largely invisible.

Being the first thinker who championed what we nowadays call embodied thinking, Schopenhauer establishes a first criterion for noumenal experience: it must allow for *gradual development* and varying *degrees of intensity*. The intimacy of one’s own body sets any embodied experience apart from many others, if only by virtue of the intensity of the event. And since the world is experienced through the body, the noumenal dimension necessarily is joined at the root with the appearances that appear in the phenomenal realm. This viewpoint is fully consistent with the Kantian idea that sensible experiences that have degrees of intensity – or “intensive magnitudes.” (CPR (A168/B201)).

Shizuteru provides a second criterion for access to the ground of reality: it is *indirect* and *non-conceptual*. It is not concerned with that which appears within the horizon-structure of experience, but with that which inflects it:

If we were to take up an image of the twofold world, it would be the horizon colored by the morning glow before the sun rises from the other side of the horizon (...) like the hemming of the horizon, the horizon in such a way enveloped by the other side [HOH 101]

The glow of the rising sun illuminates the world in an incomparable manner. It suffuses the world-as-sensibly-perceived, allowing new ways of looking at it. This way of conceptualizing the bounds of sensibility and the influence of the noumenal domain fully extends Kantian cognitive gradualism: our sensible apparatus may be trained, habituated and accustomed to perceiving hitherto unknown or even formerly unthinkable properties, relations or experiences. That the understanding must support the faculty of sensibility to come to terms with these experiences seems obvious: here, as Kant indicated, do we encounter a case where reason must venture out towards its limits, extending what was formerly thought possible within its limits. As the Kantian theory of cognition allows for sensible experiences that have degrees of intensity – or “intensive magnitudes” (CPR(A168/B201), it seems fully possible that one can learn to recognize or sense different degrees of intensity.

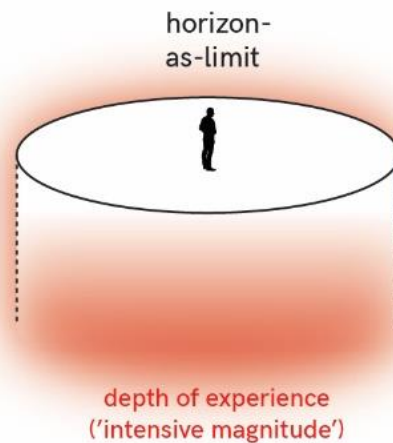


**Figure 5.** The twofold world: the bounds of "sense and reason" as well as the "outside" superimposed on these limits.

The variation in intensity points to the agency of the noumenal dimension, or its *actancy*. I will have more to say on this shortly. However, let's examine first where this emphasis on agency leads according to Shizuteru:

In that situation [of the twofold horizon], the horizon is what makes possible in principle the extent of experience in the horizontal direction, and the layering of the horizon and its other side is that which makes possible in principle the depth of experience in the vertical direction [HOH 99].

Something fundamental is at stake: while transcendental idealism maps out the margins of cognition (the extent), the noumenal domain (or “other side of the horizon”) enables the depth (or intensity) of experience. Like Schopenhauer and Schelling emphasized, there is a certain experiential intensity involved in the encounter with the “other side”. A fruitful way of reconceptualizing this encounter is via several new philosophical movements that have put objects back into the philosophical spotlight, like ecological thought, Object-Oriented Ontology, Onticology and Actor-Network Theory.



**Figure 6.** Within the basic cognitive frame, a degree of experiential intensity is possible.

Imagine that the entire world is made up of agents or assemblages of materials that have a degree of agency. An example would be a thermometer: a contraption of glass, mercury and a visual scale of marks has a degree of freedom to respond to temperature changes. Equally, we could imagine devices like metal detectors, combinations of chemical agents or catalysts as actants:

[A]n actant can be either human or nonhuman: it is that which does something, has sufficient coherence to perform actions, produce effects, and alter situations.  
(Bennett 2004, 355)

In scientific experiments, actants surface all the time. Indeed, when viewed from a Kantian viewpoint, actants are those devices that we use and indeed design to inquire into the structure of

nature itself.<sup>1</sup> Increasingly, ecological thought has started to regard our wider living environment as an actant – one that we know only partially. From the ecological perspective, many factors, material constellations and phenomena function as actants. They appear as “the other side of the horizon”, providing a perceptual hue or coloring that is hard to conceptualize. This actant-power of objects, or “thing-power” as Bennett calls it, deserves to be taken seriously on its own terms. This means that the idealist response that such encounters are “cognitively mediated” must be methodically abandoned for a moment:

My view is that while humans do indeed encounter things only in a mediated way, there nonetheless remains something to be said for the naivete of naive realism. A moment of naivete is, I think, indispensable for any discernment of thing-power, if there is to be any chance of acknowledging the force of matter. (Bennett 2004, 357)

Temporarily suspending the (transcendent) idealistic worldview opens an attitude of receptivity or creative piety. The force of things can be perceived only once we open up to its actancy or material agency.<sup>2</sup> But always, it lingers at the cognitive periphery as an inevitable surplus, an in-itself that makes its presence felt. Once more, Bennett cites several theorists who in one form or the other have dealt with it, among them figures like Adorno (non-identity), Merleau-Ponty (the invisible) Deleuze (the virtual), attempting to formulate the “feeling of internal resistance”:

[T]hing-power makes itself known as an uneasy feeling of internal resistance, as an alien presence that is uncannily familiar. (...) Recent work in cultural theory has highlighted this force that is experienced as in but not quite of oneself. (Bennett 2004, 361)

Bennett’s examples obliquely point to this surplus, to the “other side of the horizon” that inflects what appears on this side. However, it is instructive to return to Schelling, who coined the term “indivisible remainder”, pre-empting more recent thinkers:

[E]verything in the world is, as we see it now, rule, order and form; but anarchy still lies in the ground, as if it could break through once again, and nowhere does it appear as if order and form were what is original but rather as if initial anarchy had been brought to order. This is the incomprehensible base of reality in things, the indivisible remainder, that which with the greatest exertion cannot be

---

<sup>1</sup> See also Kant’s example of Evangelista Torricelli’s invention of the barometer in CPR Bxii–xiii.

<sup>2</sup> Kant pre-empts this point in his treatment of the Sublime: encountering the Sublime is actually a point where the transcendent-idealist structure seems to break down, rendering the work of the imagination necessary to construct a cognitively coherent picture.

resolved in understanding but rather remains eternally in the ground. (Schelling 2006, 29)

For Schelling, reality itself emerges from a surplus or ground-of-thought that reason cannot grasp. The categories of the understanding have indeed no power there. Yet, as an affective force, it threatens to “break through”. The Schellingian point here has to do with antecedence: the ground appears first and is itself chaotic and without rules (*Regellos*). Yet, what appears is ordered and structured only afterwards. (Dunham et al. 2014, 139–140) For Schelling as for Kant and Schopenhauer, an antecedent ground must be present for reality to arise in a minimally ordered, intelligible form. Yet, this ground can still be dimly grasped or approached rather than conceptualized. What Kant would have called the “limits of reason” is conceptualized as “personality” by Schelling:

[A]ll personality rests on a dark ground that indeed must therefore be the ground of cognition as well. But it is only the understanding that develops what is hidden and contained in this ground merely *potentialiter* [potentially] and raises it to actuality [*zum Aktus*]. This can only occur through division, thus through science and dialectic, of which we are convinced that they alone will hold fixed and bring permanently to cognition the system which has been there more often than we think but has always again slipped away, hovering before us and not yet fully grasped by anyone. (Schelling 2006, 75)

From the viewpoint of SAC, this is only to be expected, as the horizon-structure of experience is determined from the outside *and* the inside – it always hovers before us but resists conceptual closure. As every determination is a negation, a prior ground must be postulated for anything that is grounded by it to appear. This argument extends the reach of KCG, which says that many forms of experience develop gradually. This applies to learning new skills, getting acquainted with a body of knowledge or even bodily ingraining knowledge through gesture and mnemonics. But now, we can see that such gradual development is not just contingent, but necessary. The world is always larger than the perception of the perceiving agent – it appears temporally before them. Yet, to come to grips with in, an act (*Aktus*) is required. What is potentially present in the manifestly real world is through an act of the understanding developed into a cognitive handhold. Schelling – like Hegel – credits the differentiating nature of dialectical thought with this developmental process:

If the dialectical principle, that is, the understanding which is differentiating but thereby organically ordering and shaping things in conjunction with the archetype by which it steers itself. (Schelling 2006, 76)

The understanding organizes organically what comes before it through differentiation. The categories of the understanding form the structure along which this differentiation is accomplished.

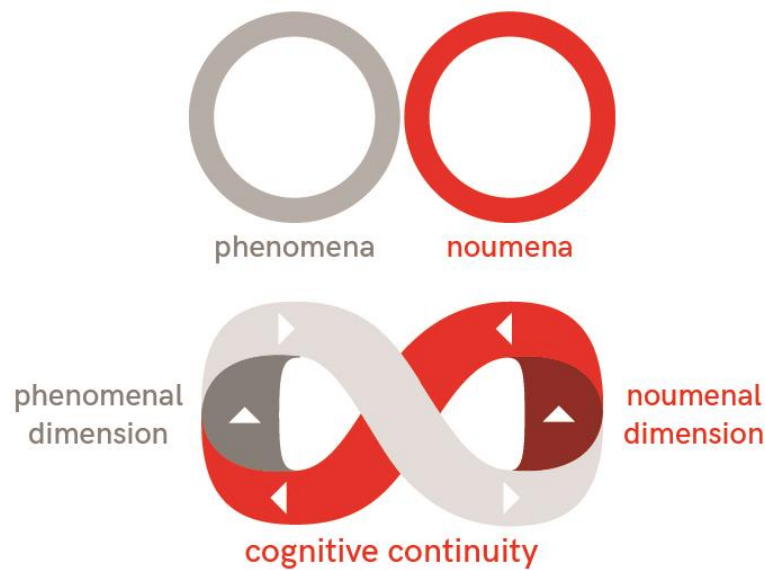
It is required for orienting oneself in thinking: instead of forming a rigid structure, the understanding organically orders intuitions, some of which are more developed than others. A version of this Schellingian thought can already be found in Kant, especially in the notion of transcendental reflection (CPR A261/B317). That faculty, according to Kant, is a state of mind that utilizes oppositional conceptual pairs to structure experience and determine whether an intuition belongs either sensibility or understanding. But apart from the formal structure that Kant proposes, we can clearly see the implication that thought-processes develop organically and through careful differentiation of its subject matter. Notably, such processes include non-conceptual and conceptual contents, therefore including concepts and affects as well as emotive impacts. This, then, allows us to grasp “the ground of thought”: it is composed of gradually enveloping experiences, some of which are experienced more intensely than others.

## **6. From the Things Themselves Towards Ecology**

I argued that the Kantian account of experience can be fruitfully extended with the Shizuteruan/anthropological approach. Instead of maintaining a strict division between noumena and phenomena, I proposed to consider Ueda Shizuteru’s model of the “horizon-structure of experience”. This model postulates a cognitive continuity between phenomena and noumena, allowing a noumenal dimension (or “Emptiness”) to permeate the phenomenal realm. So, the “great unknown” or noumenal domain becomes an integral part of our everyday experience. Consequently, our anthropological place in the world qualitatively changes and deepens due to this influence. By inviting the Things-in-Themselves in, we enrich our lived experience. In an insightful essay dealing with architecture and phenomenology, Jacques and Giroud write:

[A]rchitecture must pursue the task of securing its own possibility by longing for the thing. It is in this sense that phenomenology makes a valuable contribution to the work of architects, providing them with a means to recover the things themselves, on the ground of which rests the possibility of the discipline of “architecture”. (Jacques and Giraud 2012, 5).

But what applies to architecture applies as well to philosophy. To fully recover Things-in-Themselves, we should grant them their actancy in the world. The way in which we are affected by the depth of each object – no matter how mundane – lies within the faculty of sensibility, and is gradually worked out with proto-concepts. So, instead of regarding Things-in-Themselves as metaphysical redundancies, we should receptively and appreciatively open towards them.



**Figure 7.** From the standard model of two separated domains (top), to a situation of cognitive continuity (bottom).

The Kyoto School innovation is to fully recognize the actancy of the noumenal and integrating it within the structure of experience. So construed, the idea of a limit or horizon determines what we can experience, but it extends the bounds of sense. If we put it in Robert Pirsig's term: is there a "seeing" that sees Quality the elusive hue that suffuses everyday experience? Is there a register of experience that allows for unveiling and creating Quality? The process of furnishing experiences is gradual, as per KCG. It unfolds in real-time, allows of degrees of intensity, involves the practice of creative piety, and fully embraces non-conceptuality, as well as embodied forms of cognition. Thus, the relation between phenomena and noumena is better visualized as a Möbius strip or Klein bottle: it has separate sides or aspects but can be traversed as a continuum.

As Kant argued, notions and ideas must be worked out (*Ausprägen*). The sensible impression is conceptually grasped through the understanding, but the materials on which it exercises its structuring activity is variegated and phenomenologically deep. As Schelling put it, it is a "generative dialectic" or (proto-conceptual) differentiation that works out possibilities, venturing beyond its own boundaries. The sensibility that Kant speaks about extends towards a dimension that grounds all the rest. At this point, what Pirsig designated as "Quality" or experiential depth enters, if one possesses a sufficiently developed sensibility. As discussed, Kant noticed in his account of intensive magnitude that certain experiences allow of varying degrees of intensity, and so it is only prudent or at least epistemically humble to assume that certain types of experiences can be further developed than we currently know. We use routinely the category "Quality" to describe with the fact that our sensibility picks up on the varying degrees of intensive magnitudes.

Actually, this is simplification – a category that describes the fact that sensations register variety of intensity.

What emerges is a picture that is inherently relational yet still monistic. We can take the Thing-in-Itself under its phenomenal and noumenal aspect and consistently argue that these are two deeply intertwined domains fused in appearances, yet always containing an indivisible remainder. This requires that we postulate a deep and thoroughgoing relationship between the noumenal and the phenomenal – in other words, we require the “other side of the horizon” or “ground of thought” to gradually advance our thinking. This means developing the skills and attitudinal dispositions to detect and access it. Creative piety is the attitudinal disposition required to achieve this access.

This last point is particularly relevant while thinking about ecology and organicism. The physical world appears perceptually before humanity and was there before human consciousness developed. But the presence of thing-power, or the agency that the environment exerts is largely unknown – yet not unperceived – by us. The resonance of for instance ecological thought, as well as Glenn Albrecht’s work on finding new words for a new world point towards the necessity of having to orient ourselves in thought once more (Albrecht 2009; Morton 2013, 2018). We live in what has been called an “Ignorance Society”, whereby the limits of what we do not know force ourselves on our collective consciousness. (Brey, Innerarity and Mayos 2009) The *omnitudo realitatis* becomes paradoxically “larger than we thought”. Likewise, our ecological predicament appears to us as the “great unknown”, yet it structures our experiences, and exerts its influence from the cognitive periphery. Our action space (*Handlungsraum*) is larger than we may realize, but it requires a refined and developed sensibility to explore its limits and possibly even attempting to think what lies beyond them.

Against this background, Kant’s plea for cultivating the sensibility for navigating the space of thought acquires renewed relevance. The philosophical project of transcendental idealism is more necessary than ever, if only to prevent humanity from disregarding its limitations and attempting to play God. Conversely, Kantian philosophy stresses the importance of limitations, leading to an attitude of epistemic humility and taking moral responsibility. At the point where we live this attitude collectively, we might well be able to answer that quintessential Kantian question “what can we hope for?”

## References

- Albrecht, G. (2009). *Earth Emotions. New Words for a New World*, Cornell University Press.
- Allison, H. (2004). *Kant’s Transcendental Idealism: An Interpretation and Defense*, Yale University Press.
- Augé M. (2006). *Non-Places. An introduction to super modernity*, Verso.
- Bachelard, G. (1994). *The Poetics of Space*, Beacon Press.
- Beiser, F. (2014). *The Genesis of Neo-Kantianism 1796–1880*, Oxford University Press.
- Bennett, J. (2004). The Force of Things. Steps Towards an Ecology of Matter, in: *Political Theory* 32(3), 347–372. DOI: [10.1177/0090591703260853](https://doi.org/10.1177/0090591703260853)



- Bennett, J. (2010). *Vibrant Matter: A Political Ecology of Things*, Duke University Press.
- Boer de, K. (2014). Kant's Multi-Layered Conception of Things in Themselves, Transcendental Objects, and Monads, in: *Kant-Studien* 105(2), 221–260. <https://doi.org/10.1515/kant-2014-0011>
- Bollnow, O. F. (2022). *Men's en Ruimte*. Uitgeverij Noordboek.
- Bryant, L. (2011). *The Democracy of Objects*, University of Michigan Press/Open Humanities.
- Brey, Antoni, Innerarity, D. & Mayos, G. (2009). *The Ignorance Society and Other Essays*, Infonomia.
- Clewis, R. (2018). Kant's Physical Geography and the Critical Philosophy, *A Journal for the History of Philosophy* 22(2), 411–427. <http://dx.doi.org/10.5840/epoche201819109>
- Dunham, J. Hamilton Grant, I. & Watson, S. (2014). *Idealism. The History of a Philosophy*, Routledge.
- Griffin, D. R. (1998). *Unsnarling the World-Knot. Consciousness, Freedom and the Mind-Body Problem*, Wipf Stock.
- Hanna, R. (2017a). Life-changing Metaphysics: Rational Anthropology and its Kantian Methodology, in: *The Cambridge Companion to Philosophical Methodology*, pp 187–210, Edited by D'Oro, G. & Overgaard, S.
- Hanna, R. (2017b). Kant, Radical Agnosticism, and Methodological Eliminativism About Things in Themselves, *Contemporary Studies in Kantian Philosophy* 2, 38–54. <https://doi.org/10.5922/020769182017-4-4>
- Hanna, R. & Paans, O. (2021). Thought -Shapers, *Cosmos and History*, 17(1), 1-72
- Hanna, R. & Paans, O. Creative Piety and Neo-Utopianism: Cultivating Our Global Garden, *Cosmos and History*, 18(1), 1-82
- Harman, G. (2013). *Bells and Whistles. More Speculative Realism*, Zero Books.
- Harman, G. (2018). *Object-Oriented Ontology. A New Theory of Everything*, Penguin Random House.
- Hume, D. (2008). *An Enquiry Concerning Human Understanding*, Oxford Univ. Press.
- Ireland, T. (2015). The Spatiality of Being, in: *Biosemiotics* 8, 381–401. DOI 10.1007/s12304-014-9227-7
- Jacquet, B. & Giraud, V. (eds.). (2012). *From the Things Themselves: Architecture and Phenomenology*, Kyoto University Press.
- Jullien, F. (2016). *The philosophy of living*. Translated by M. Richardson & K. Fijalkowski, Seagull Books.
- Marshall, C. (2013). Kant's appearances and things in themselves as qua-objects, in: *Philosophical Quarterly* 63 (252) 520-545
- Meillassoux, Q. (2018). *After Finitude. An Essay on the Necessity of Contingency*, Bloomsbury.
- Morton, T. (2013). *Hyperobjects. Philosophy and Ecology after the End of the World*, University of Minnesota Press.
- Morton, T. (2018). *Dark Ecology. For a Logic of Future Coexistence*, Columbia University Press.
- Nishitani, K. (1983). *Religion and Nothingness*, University of California Press.
- Paans, O. (2022). Cold Reason, Creative Subjectivity: From Scientism and the Mechanistic Worldview to Expressive Organicism, *Birderless Philosophy*, 5, 161- 212.
- Paans, O. (2023). Kant's Cognitive Gradualism: Reflection and Experience, *Contemporary Studies in Kantian philosophy*, 8, 102-121
- Pirsig, R. M. (2006). *Zen and the Art of Motorcycle Maintenance*, Harper Torch.
- Schelling, F.W.J. (2006). *Philosophical Investigations in the Essence of Human Freedom*. Translated by J. Love & J. Schmidt, State University of New York.

- Schopenhauer, A. (2018). *The World as Will and Representation*. Translated by J. Norman, A. Welchman, & C. Janaway, Vol. 2, Cambridge University Press.
- Shizuteru, U. (2019). Horizon and the Other Side of the Horizon, in *Contemporary Japanese Philosophy. A Reader*, Edited by J. W. M. Krummel. Pp 93- 106, Rowman & Littlefield.
- Staton, C. (2023). Imagination and Transcendental Objects: Kant on the Imaginary Focus of Reason, in *The Being of Negation in Post-Kantian Philosophy*, Edited by G.S. Moss, pp. 57-75, Springer.
- Tomaszewska, A. (2007). Transcendental object and the “problem of affection”. Remarks on some difficulties of Kant’s theory of empirical cognition, in: *Diametros* 11 (26), 61–82  
<https://doi.org/10.13153/diam.11.2007.268>
- Whitehead, A. N. (1960). *Science and the Modern World*, Cambridge University Press, 1960.
- Winegar, R. (2017). Kant on intuitive understanding and things in themselves, *European Journal of Philosophy*, 26(1), 1–15. <http://dx.doi.org/10.1111/ejop.12320>

# Consciousness and Cognition in Kant's First Critique

Roberto Pereira 

Federal University of Rio de Janeiro, Brazil. Email: [robertohsp@gmail.com](mailto:robertohsp@gmail.com)

---

## Article Info

### Article type:

Research Article

### Article history:

Received 20 May 2024

Received in revised form  
03 May 2024

Accepted 10 July 2024

Published online 07  
August 2024

### Keywords:

transcendental  
consciousness; access  
consciousness;  
nonconceptual sensible  
intuition; cognition.

## ABSTRACT

This paper has the ambitious aim to clarify the putative different meanings of "consciousness" in Kant's *Critique*, particularly focusing on the concept of apperception. Often misinterpreted merely as the potential for self-attributions of experiences and mental states—technically, as the individual's ability to knowingly refer to himself—such readings overlook the pivot role of transcendental apperception in bridging the inherent gap between nonconceptual content of sensible intuitions and the higher-level conceptual content of propositional attitudes, essential for reasoning and the rational control of actions. In this context, "consciousness" or "self-consciousness" means cognitive accessibility (in Block's sense). But Kant's texts reveal additional meanings of consciousness. Notably, "sensation" means the raw material of intuition when it is apprehended through a synthesis of imagination without conceptual determination, capturing the subjective "what-it-is-like" phenomenal aspect of perception. Conversely, its objective correlate—the ability to discriminate and single out objects from their surroundings—embodies what can be described as "de re awareness" of a yet conceptually undetermined object of intuition.

---

**Cite this article:** Pereira, R. (2024). Consciousness and Cognition in Kant's First Critique. *Journal of Philosophical Investigations*, 18(47), 139-160. <https://doi.org/10.22034/JPIUT.2024.61729.3770>



© The Author(s).  
<https://doi.org/10.22034/JPIUT.2024.61729.3770>

Publisher: University of Tabriz.

---

## Introduction

Kant's concept of transcendental apperception holds a central position within his theoretical framework, being referred to as the "highest point" and a "Radikalvermögen," which serves as the pinnacle of transcendental philosophy. It encompasses "all employment of the understanding, even the whole of logic and transcendental philosophy" (Kant, 1781/1787, B135). In the transcendental analytic, transcendental apperception plays a crucial role in the transcendental deduction of the categories; in the transcendental dialectic, it is central to the paralogisms of pure reason. This raises the following questions: first, what is apperception? Second, are there other irreducible senses of "consciousness" in Kant's work?

This paper has the ambitious aim to clarify the putative different meanings of "consciousness" in Kant's philosophy, particularly focusing on the concept of transcendental apperception. Often misinterpreted merely as the potential for empirical self-consciousness or self-attributions of experiences—technically, as the individual's ability to knowingly refer to oneself—such readings overlook the pivot role of transcendental apperception in bridging the inherent gap between nonconceptual content of sensible intuitions and the higher-level conceptual content of propositional attitudes, essential for reasoning and the rational control of actions. In this context, "consciousness" or "self-consciousness" mean cognitive access (in Block's sense). But Kant's texts reveal additional meanings of consciousness. Notably, "sensation" means the raw material of sensory intuition when it is apprehended through a synthesis of imagination without conceptual determination, capturing the subjective "what-it-is-like" or phenomenal aspect of perception. Conversely, its objective correlate—the ability to discriminate and single out objects from their surroundings—embodies what can be described as "de re awareness" of a yet conceptually undetermined object of sensory intuition.

This article is structured as follows: In Section 1, following this historical introduction, I present and critique the "language-analytic" reading of Kant's transcendental apperception. As expected, Strawson's interpretation in his seminal 1959 and 1966 books serves as the starting point. However, a prominent figure of the German Heidelberg School, Konrad Cramer, offers the finest and most detailed linguistic analyses in his papers from 1987 and 2013, inspired by Ernst Tugendhat's "language-analytic" approach to the phenomenon of epistemic immediate self-consciousness (see Tugendhat, 1979). Additionally, in a systematic paper unrelated to Kant's scholarship, Lynne R. Baker provides the clearest account of the analytic reception of Kant's apperception (see Baker, 1998).

In Section 2, I critically examine the unique perspective of Kant's apperception in relation to both the global workspace theory of consciousness (GWT) in neuroscience and the higher-order theories of consciousness (HOT). While Kant's usage of the term "consciousness" typically aligns with "propositional consciousness," Section 3 is dedicated to showcasing the diverse and

independent interpretations of what we commonly refer to as "consciousness" in Kant. This is a novel aspect of Kant's hierarchical arrangement of representations in the *Critique* and his logical writings.

In section 4, I present and defend my reading of Kant's transcendental apperception. What Kant had in mind, distinct from the contemporary understanding of self-consciousness, is the "cognitive accessibility" of nonconceptual representations. This understanding allows us to draw a connection between Kant's transcendental apperception and global workspace theory, not as a theory of consciousness but as a theory of cognition. In the final section, 5, I reiterate: My reading provides robust additional support for the nonconceptual reading of the *Critique*, reinforcing its validity.

### 1. The potential Self-attribute of Experiences.

Strawson's interpretation of Kant's apperception fits into his "austere" analytical reconstruction of Kant's Deduction. This reconstruction is meant to show that the sense-data skeptic (as a straw figure) who challenges us to demonstrate that our experience is not limited to fleeting sense-data. While it can be disputed whether Kant's Deduction has the goal that Strawson ascribes to it, there is no question that Kant holds that our experience is about a mind-independent reality. Kant assumes that we experience objects and tries to show that the concept of object in this strong/weigh sense is presupposed by the self-ascribing of experiences. Strawson's reading is as follows:

(i) The experience contains a diversity of elements (intuitions) that must somehow be united in a single consciousness capable of judgment, capable, that is, of conceptualizing the elements so united (Strawson, 1966, 87).

(ii) This [subjective] unity requires another kind of unity or connectedness on the part of the multifarious elements of experience (Strawson, 1966, 87).

(iii) This [another kind of unity or connectedness] is required for the experience to have the character of the experience of a unified objective world and, hence, to be capable of being articulated in objective empirical judgments. [From (i) and (ii)]

Strawson conceives the transcendental apperception as follows:

Unity of the consciousness to which a series of experiences belong implies [...] the possibility of self-ascription of experiences on the part of a subject of those experiences; it implies the possibility of consciousness, on the part of the subject, of the numerical identity of that to which those different experiences are by him ascribed. (Strawson, 1966, 98)

Strawson's particular reading of transcendental apperception is closely tied to his misinterpretation of the aim and structure of the deduction. To be sure, any interpretation of transcendental apperception is connected to an interpretation of the aim and structure of the transcendental deduction. However, I want to reassure you that I will not be engaging in these debates, as they are beyond the scope of this article. While I acknowledge the connection between the interpretation of transcendental apperception and the understanding of the deduction, I will focus on the transcendental apperception and steer clear of the controversies surrounding the aim and structure of the deduction that have been ongoing since the first edition of the *Critique*.

The most thorough linguistic-analytical exposition of Strawson's interpretation of transcendental apperception originates not from the Anglo-Saxon sphere but from Germany, specifically in essays by Konrad Cramer (1987, 2014). Cramer's linguistic elaboration of Strawson's interpretation is based on Ernst Tugendhat's linguistic approach to self-consciousness (see Tugendhat, 1979)—which lays the groundwork for understanding self-consciousness but does not directly engage with Kant's particular concept of transcendental apperception. Cramer's essay focuses entirely on Kant's renowned passage in paragraph 16 of the transcendental deduction, a passage that we will reference multiple times throughout this paper. In an oracular tone, Kant says:

The *I think* must be able to accompany all my representations [Vstellungen]; for otherwise something would be represented in me that could not be thought at all, which is as much as to say that the representation would either be impossible or else at least would be nothing for me. That representation that can be given prior to all thinking is called intuition. (B131/B132, original emphasis)

Regardless of the pivot role that apperception plays in the main argument of B-deduction, the passage contains a dense and obscure argument of its own:

- 1) That I think, must be able to accompany all my representations.
- 2) For otherwise something would be represented in me that could not be thought, which means just as much as:
  - 3) The representation would either be impossible,
  - 4) Or at least nothing for me.

---

Therefore, the I think I must be able to accompany all my representations.

The key to the dense argument is how the "for" clause justifies the general claim 1). According to Cramer (2014), the argument may seem flawed at first glance. First, the conclusion Kant draws, namely: "Therefore, all the variety of intuitions must have a necessary relation to 'I think,' in the same subject to which this manifold appears" (B 132), is not a restatement of his thesis that intuitions without concepts are "blind," and concepts without intuitions as thoughts without

content are "empty" (A 51/B 75) (see Cramer, 2014, 26). Indeed, even if the cooperation between intuitions and concepts is necessary for the possibility of cognition of objects, that is something that Kant does not state when he claims that the "I think" must accompany all my representations.

Moreover, the simple, dense and obscure argument raises several questions: (1) How do representations qualify as "my representations?" Is the accompaniment of "I think" a necessary condition for the ownership of "my representations?" (2) What is the difference between *within me* and the ownership of my representations? According to Cramer, the claim: "this representation is mine" ("belonging-to-me," the *Meinigkeit*) is not an answer to the trivial question of what representation I have? (see Cramer 1987, 189). Finally, (3) What does "impossible" and "nothing for me" mean?

According to Cramer, representations are rather mental states that have always been considered almost definitions of the "soul" (A 19/B 33; A 22/B 37; B 67 et seq.) or "subject" (B 67; B 132 et seq.) and which "enjoy" in it (B 132). The expression "within me" is evidently related to the expressions "within my soul" and "within my mind." The phrase "to represent something within me" accordingly means "to represent something within my soul." This, in turn, according to the above definition of representation, means representing something through the determination of my soul. That is to say, the determination of my soul represents an object through its structural similarity.

There are two possible interpretations of the expression "to represent something within me" (see Cramer 1987, 181n.). Firstly, Kant could have emphasized the owner of the representation. Accordingly, Kant wanted to highlight that I represent something within myself and not within another. However, this interpretation is inappropriate or at least imprecise because the possibility of external ownership is not addressed in the context of the principle of apperception (at least from the perspective of the third person) and is, therefore, irrelevant. Cramer (1987, 181n.) himself prefers the view that "to represent something within me" designates a mental event or state.

But from this characteristic of representations, it does not follow that the soul or subject, whatever its definition, must also already be aware that representations are in it. It is here that following the Strawsonian tradition, Cramer introduces self-attribution as a way to differentiate "representation *within me*" from "representation *for me*:"

But if the subject has the consciousness of the representation in him, then he attributes these representations to himself. The self-ascription of representations has a linguistic form "I know that I  $\phi$ " or "I  $\phi$ ," where  $\phi$  is a predicative variable, for which every time descriptive predicates, which are concepts by which mental states or mental events such as what Kant calls "representation" are specified. If these representations are givennesses of sensibility in the form of intuition, then I must have concepts to describe them so that I can attribute them to myself. Concepts for intuition are not intuitions; they relate to intuitions. That description of intuitions is possible only through the

use of concepts through which they are thought to be a trivial truth. Therefore, if self-ascription of intuitions contains descriptions of this intuition, then it is also true that "I think" must be that which must be able to accompany intuition. So, it all comes down to the self-ascription of representations. In judgments like "I know that I  $\phi$ ," the one who makes the judgment not only describes the representation, which is something in it, but also, through the use of the singular term "I," relates it to oneself. Thus, the singular term "I" is not only synonymous with what it refers to but also with the one who uses it through speech. Therefore, the singular term "I" itself has a semantic reference to a conceptual form of such kind that Kant - even if not so explicitly - could only speak about "the concept of thinking beings in general" (KrV, B 418), and also, about "the concept of non-self in thought" and about "the concept of I, as it arises in any thought" (KrV, A 342/B 400). (Cramer, 2014, 26)

Cramer adopts Tugendhat's linguistic analysis of self-consciousness (Tugendhat, 1979). The conscious awareness that a representation is "in me" takes the following canonical form: "I know [ich weiss] that I  $\phi$ " (where " $\phi$ " denotes a psychological predicate), such as "I know that I am bored" (Tugendhat, 1979, 13). Let me call this the "semantic interpretation." When we interpret Kant's statement, "that the I think must be able to accompany all my representations," as the potential self-attribution of psychological mental states in the form of propositional content, we have the following statement:<sup>1</sup>

➤ If some representation  $\phi$  is means something for me, then it must be possible that "I know that I  $\phi$ ."

Now we know (1) what qualifies a representation as my representation, (2) why the accompaniment of "I think" is a necessary condition for the ownership of "my representations," what the difference is between "in me" and the ownership of my representations, and (3) what "impossible" and "nothing for me" mean is the absence of self-attribution. Finally, however dense and obscure, the argument is as cogent as possible: the "I think must..., i.e., "I know [ich weiss] that I  $\phi$ ," is a condition for representing a representation in me, i.e., a mental state of myself.

---

<sup>1</sup> The full statement "I know that I  $\phi$ " may sound odd in everyday language, as we usually just say "I  $\phi$ ." However, the sentence "I  $\phi$ " is implicitly understood to mean the self-ascription of a complete psychological propositional content ("I know that I  $\phi$ ") rather than just a single psychological predicate. For those who find the full statement "I know that I  $\phi$ " uncomfortable, Lynne Rudder Baker (1998) offers several examples with a similar propositional structure by substituting the verb "know" with other intentional verbs, such as "I wish I were tall," "I hope I will win," or "I am afraid that I am going to die." While she has not analyzed Kant's apperception, her examination of self-consciousness bears resemblance to Cramer's approach: self-consciousness is what she describes as a "strong first-person perspective," which refers to the self-ascription of the "weak first-person perspective" that is common to animals and infants. She defines this as follows: "The former [I am tall] *makes* a first-person reference; the latter [I wish I were tall] *attributes* first-person reference to *herself*" (Baker, 1998, 330, original emphasis).



The semantic interpretation of transcendental apperception is flawed for several reasons. First, it is flawed for pure exegetical reasons. If the given representation is a sensible nonconceptual representation, then—Cramer argues—the subject must conceptualize it, otherwise, it could not attribute it to himself as a mental state "I  $\phi$ " ("If these representations are givennesses of sensibility in the form of intuition, then I must have concepts to describe them so that I can attribute them to myself" (Cramer, 2014, 26). In these terms, Cramer puts the cart before the horse. How could the subject "conceptualize" the nonconceptual content of his sensible intuition if not by the "I think must accompany my sensible intuition?" If the usual self-attribution ["I think that I  $\phi$ ,"] relies on the conceptualization of some intuition as " $\phi$ ," we should not forget that transcendental apperception is the "vehicle" of all concepts, as Cramer recognizes, the statement "I think, must be able to accompany all my representations" cannot mean the self-attribution of mental states in the canonical form: "I  $\phi$ ." Either what Cramer calls "self-attribution" constitutes the conceptualization of the nonconceptual representation and in that case self-attribution cannot mean "I think that I  $\phi$ ," or Cramer's account runs in a vicious circle.<sup>1</sup>

Furthermore, as Cramer himself acknowledges, "Whoever uses the term 'I' uses the first-person pronoun and thereby relates to themselves as individuals, to the very person they are, therefore, to individuals are entities in space and time" (Cramer, 2014, 26). The question arises as to how Cramer reconciles this triviality with his previous remark that the singular term 'I' itself refers to "the concept of thinking beings" (Cramer, 2014, 26). On the contrary, when Kant claims that "I think" is the "vehicle" of concepts, even of transcendental ones, he denies that "I think" is a concept and, hence, cannot occur in self-ascriptive propositions (see A 341/B 399; A 348; B 406).

In the same vein, Kant rises self-consciousness to the status of "highest point" or *das Radikalvermögen*, claiming that "all employment of the understanding, even the whole of logic and transcendental philosophy" (B135). The self-attribution of experiences ("I think that I  $\phi$ ") does not play this central role. On the contrary, self-attribution of experiences ("I  $\phi$ ") relies on a range of cognitive abilities. Self-attribution of experiences typically develops between the third and fourth year of life, after the acquisition of language, the mastery of numerous concepts and the ability to attribute false beliefs to others.<sup>2</sup>

<sup>1</sup> By far, Cassam's book is the most ingenious reading of Kant's self-consciousness as an embodied individual who is conscious of his own experiences (see Cassam, 1997). It weaves a complex yet nuanced array of arguments supporting the idea that I am conscious of myself as a subject of experiences by being conscious of myself as a physical object among physical objects. While it remains open to debate whether Cassam has accurately captured Kant's perspective on self-consciousness, the merit of his arguments stands on their own.

<sup>2</sup> We must, therefore, concur with Longuenesse (2006). If the first personal pronoun "I" does not refer to a person but rather linguistically manifests a sentence operator, the most reasonable conclusion is that in Kant's work, the first-person pronoun "I" does not serve as a self-referring expression. Instead, in Kant's work, the "I" is ontologically committed to what Longuenesse calls "the spontaneous thinking agency" responsible for the synthesis of

## 2. The 'I Think,' the GWT and the HOT

Contemporary cognitive science, with its cutting-edge insights, offers a fresh perspective on Kant's apperception. Our novel approach is to link Kant's distinction between unconscious and conscious representations with Dehaene et al.'s distinction between subliminal and conscious representations (see Dehaene, 2006). This connection—we believe—sheds new light on Kant's ideas. The global workspace theory, a key component of our analysis, suggests that consciousness arises from integrated information that is globally available to guide actions and verbal reports (see Baars, 1988).

Dehaene has further developed Baars' original functional model. He assumes that the global working space certainly may comprise brain structures in the prefrontal cortex and the frontal areas of the brain that are frequently associated with known cognitive functions. For an object stimulus to become a "conscious" representation, it must first activate excitatory neurons that have extensive connections in the cortex. In this case, the stimulus must be able to support activity throughout the brain. Be that as it may, the neurological regions involved are not relevant to our topic here.

One of the intriguing aspects of attention and consciousness is the competition for cognitive accessibility to the global workspace. Many sensory representation objects "compete" for this access at any given time, but only some can be consciously experienced at any given time. At the neuronal level, "activated neuronal coalitions compete to initiate recurrent activity" and fight for access to the global workspace (Block, 2023, 7). According to this view, attention is a prerequisite for consciousness (see Schlicht & Newen, 2015). This perspective sparks a fascinating discussion

---

representations. However, how are we to interpret positively the activities of "combining representations" and "thinking agency?"

Longuenesse proposes an intriguing modern counterpart to Kant's formal "I," suggesting that Kant's notion of transcendental apperception parallels the Freudian Ego in opposition to the Id, rather than aligning with contemporary views of self-reference. She posits that like Freud's ego ('das Ich'), which organizes mental activities to create a reliable world image conducive to life preservation, Kant's apperception synthesizes representations to enable object representation (Longuenesse, 2017, 20).

However, Longuenesse acknowledges crucial distinctions between Kant's apperception and Freud's concept of the ego. Kant emphasizes the spontaneous agency vital for object cognition, centering his critique on evaluating the cognitive claims of pure reason and addressing the question, "What can I know?" (A805/B833). Conversely, Freud's focus does not include a theory of cognition or epistemological inquiries.

But the most glaring difference is one that Longuenesse seems to overlook: Freud's second "topic"—the division of the psyche into the id, ego, and superego—focuses on impulses and sexual drives (conative states) rather than cognition. Therefore, the organizing principle in Kantian is cognitive, while in Freud, it is conative. Furthermore, although the Freudian ego is exactly a "person," it is an embodied subject (see Muller & Tillman, 2007).

on the interplay between attention and consciousness. Critics contend that the extent of attention does not constrain cognitive access (see Block, 2011), a viewpoint I concur with.<sup>1</sup>

In line with paragraph 16, one might suspect a close analogy between Kant's emergent theory of consciousness and global workspace theory (GWT). Kant claims that "something would be represented in me that could not be thought at all, which is as much as to say that the representation would either be impossible or else at least would be nothing for me" (B131/B132). In accordance with this proposal, a sensory representation (which takes place in my sensory cortex at the back of my head) without the accompaniment of "I think" would not count as a "representation with consciousness."

In the framework of Global Workspace Theory (GWT), the phrase "impossible or nothing for me" denotes a representation devoid of consciousness. The resulting understanding is as follows: The cognitive assertion "I think" must accompany sensible representations; without it, these representations would be meaningless or nothing for me—labelled "unconscious." Sensory representations require the accompaniment of an "I think" to be integrated into the global workspace, thereby becoming accessible for propositional attitudes, reasoning, and the rational control of behavior as intentional actions. This integration confirms the conscious status of these representations and their cognitive availability. Stanislas Dehaene's use of "subliminal" in Kantian terms refers to those representations that could potentially be accompanied by the "I think" but are not in reality. These are the representations that disappear without ever being integrated into the global workspace, remaining unconscious or subliminal.

It has also been argued that Kant was a precursor to what is now known as the Higher-Order Thought (HOT) theory of consciousness (Gennaro, 1996, §3.4, 48–54). Within Kant's framework, the theory suggests that "meta-psychological thoughts" are what render mental states "conscious" in the first place. The assertion that "I think" must accompany all my representations indicates a meta-representation, making a first-order representation meaningful and conscious. In addition, other passages seem to support the HOT assertion that thoughts render sensory states conscious. For instance, in a reflection from 1769, Kant asserts, "In fact, the representation of all things is the representation of our own state" (HN, Ref. 3929, Ak. 17, 351). Furthermore, he explains,

---

<sup>1</sup> I believe that it is what Kant has in mind in the famous passage where there is a competition between the musician playing an improvisation and a friend talking to him:

"The same thing is true of the sensation of hearing when the musician plays a phantasy on the organ with his ten fingers and two feet while, at the same time, he is speaking to someone standing beside him. Within a few seconds, a host of ideas is awakened in his soul, and every idea requires special judgment as to its appropriateness since a single stroke of the finger, not fitted to the harmony, would immediately be heard as discord. Yet the whole comes out so well that the improvisator must often wish to have preserved in a score many a passage which he has performed in this happy fashion but which he could not have performed so well with real diligence and attention." (*Anth*, Ak, 7, 135). I will be back to this famous passage.

"consciousness is...a representation of my representations; it is a self-perception" (VM, Ak. 28, 227). Thus, Kant establishes a significant link between consciousness and self-consciousness, as discussed by Gennaro (1996, 49). The "I think" pertains to a specific thought that targets one's representations, illustrating that consciousness entails self-consciousness or HOTs.

Assuming Kant as a precursor to the Higher-Order Thought (HOT) theory, it follows that he viewed all representations, prior to being accompanied by "The I think," as nonconscious (Gennaro, 1996, 49). However, Gennaro notes the ambiguity of the term "representation" in Kant's writings, which can be interpreted either as a "mental state" or as a "conscious state" (Gennaro, 1996, 49). This ambiguity leads to significant questions: What does it mean for "the I think" to accompany a representation, and is this accompaniment a condition for consciousness? Additionally, the ambiguity prompts an inquiry into whether representations lacking "the I think" are deemed "unconscious" and whether various forms of consciousness exist.

In his analysis, however, Gennaro encounters another complication. He suggests that the phrase "I think" can be interpreted in two ways: "I think unconsciously" and "I think consciously" (Gennaro, 1996, 49). If Kant intended the latter, "I think consciously," then self-consciousness equates to introspection. Gennaro explains, "To have a conscious mental state is to be able to think consciously about that state, i.e., engage in introspection" (Gennaro, 1996, 49). He argues that Kant's conflation of the two perspectives leads to the assumption that any unconscious representation must exist within a self-conscious being. This implies that a representation must be subjected to introspection to become conscious.

The only alternative interpretation is to understand Kant's claim that "the 'I think' must be able to accompany all my representations," meaning that "having conscious mental states presupposes being able to have (unconscious) thoughts about them" (Gennaro, 1996, 51). Gennaro's critique of Kant's supposed conflation of "I think unconsciously" with "I think consciously" is significant. Gennaro argues that Kant could have supported his alternative interpretation: "The application of concepts to one's representations need not be conscious in order to become conscious" (Gennaro, 1996, 52).

It is with a sense of respectful disagreement that I approach this interpretation. Gennaro (1996, 49) offers two interpretations of the phrase "I think": "I think unconsciously" and "I think consciously." The ambiguity, however, lies in the distinction between "I think about what the sensible intuition represents" and "I think that the sensible intuition is a state of my mind." In the first case, we are dealing with a simple propositional consciousness, such as "I think that the cinnabar is red" (Kant's example: A101). In the second case, we enter the realm of introspection, as in "I think that I perceive the cinnabar as red." Here we are back to Cramer's linguistic reading: "I think that I  $\phi$ ." Kant's distinction between a transcendental and an empirical understanding of the expression "my representation" clarifies this dual interpretation. In the transcendental sense,

everything that appears is "my representation." In contrast, in the empirical sense, every manifestation is an appearance of a thing in itself in space and time outside of me.

Nevertheless, "my representations" in the transcendental sense do not constitute what we currently describe as "introspection." They are not objects of "I think" as a form of second-order representation that also serves as a first-order representation. The transcendental sense merely offers a worldview. If I intuitively perceive something as red, and this sensible intuition is accompanied by an "I think," it merely implies that I think that the cinnabar is red, that ripe apples are red, and that ripe tomatoes are red, among others.

In summary, it is highly implausible to interpret Kant's "I think" as meaning "I nonconsciously think." Equally implausible is the suggestion that Kant's "I think" is a second-order thought that takes some first-order representation as its object (a meta-representation). In this sense, transcendental apperception is more aligned with Global Workspace Theory (GWT) than with Higher-Order Thought (HOT) theory. However, as I will demonstrate in the following section, Kant should not be viewed as a precursor to both GWT and HOT. This is because neither the integration of representations into the global workspace nor a higher-order thought that takes a representation as its object qualifies as a condition for *consciousness*. In the conclusion of this paper, I will argue that integration into the global workspace makes representations *cognitively accessible*.

But let us return to Kant's argument of the paragraph 16:

- 1) That I think, must be able to accompany all my representations.
- 2) For otherwise, something would be represented in me that could not be thought, which means just as much as:
  - 3) The representation would either be impossible,
  - 4) Or at least nothing for me.

---

Therefore, that I think must be able to accompany all my representations.

As we have observed, the crux of this argument hinges on how premise (2) substantiates the general assertion in premise (1). Following the Global Workspace Theory (GWST) or Higher-Order Thought (HOT) theory, premise (3), "impossible or nothing for me," refers to a *representation lacking consciousness*; thus, the argument is as cogent as possible. The argument posits that the intellectual representation "I think" (whether HOT or not) must accompany any representation; otherwise, it would be devoid of meaning in the sense of lacking consciousness. If this interpretation is correct, Cramer (1987; 2014) is accurate in asserting that § 16 is not merely a reiteration of Kant's distinction between intuitions and concepts, specifically the claim that sensible intuitions without discursive concepts are "blind" and concepts without intuitions are "empty" (A 51/B 75), because what is absent is consciousness, not merely understanding or cognition.

### 3. The 'I Think' is not a condition for Phenomenal Consciousness.

The complex question we address is whether Kant's transcendental apperception constitutes a necessary condition for (phenomenal) consciousness. Collins (1999) affirms this, stating that "conscious experience involves reflexive self-consciousness" (p. 12). More explicitly, he explains that "the overall plan of Kant's account of experience suggests that consciousness of objects is only achieved when elementary representations, of which the subject is not conscious, are suitably concatenated" (p. 79).

However, Kant's "progression" ("Stufenleiter") in the *Critique* suggests otherwise:

The genus is representation in general ("repraesentatio"). Under it stands the representation with consciousness ("perceptio"). A perception [Perception] that refers to the subject as a modification of its state is a sensation ("sensatio"); an objective perception is a cognition ["Erkenntnis"] (cognitio). The latter is either an intuition or a concept (intuitus vel conceptus). (A319-20/B376-77).

If we consider sensible intuition solely in relation to the individual—as a modification of his mind, not in terms of its objective correlation with what it represents—we arrive at what Kant terms "sensation" ("sensatio"/"Empfindung"). Kant's notion of "sensation" and "inner sense" aligns closely with what is currently referred to as "phenomenal consciousness," the subjective aspect of sensory experience. When I focus exclusively on my personal experience of seeing something red, irrespective of its actual representation, this act of seeing red takes on a distinct "what-it-is-like" phenomenal character of perception. Kant asserts that transcendental apperception is not required for sensation. However, if there are conscious representations, which Kant identifies as perceptions, there must also exist "unconscious" representations. Kant addresses this issue in his *Anthropology from a Pragmatic Point of View* by proposing that:

The field of sensuous intuitions and sensations of which we are not conscious, even though we can undoubtedly conclude that we have them; that is, *obscure* representations in the human being (and thus also in animals), is immense. *Clear* representations, on the other hand, contain only infinitely few points on this field that lie open to consciousness (*Anth*, Ak, 7, 135, emphasis added.)

A brief digression on the concept of "consciousness," as adopted by Kant from the Leibniz-Wolff tradition, is necessary here. Leibniz and Wolff fundamentally distinguish between two types of non-conscious representations. First, there are what they term "obscure ideas." Leibniz famously describes these as "perceptions without apperception" or "petites perceptions" (Leibniz, 1704, I. xxix, § 13, 242f). The essence of "obscure ideas" is that they place us in a mental state where we are completely unaware of what our mental states represent. Second, there are "confused

ideas," which are clear as opposed to obscure, allowing us to represent objects but not distinctly. Wolff claims that we are clearly conscious of something only insofar as we can differentiate it from something else (Wolff, 1751, § 729). Consequently, it is not possible to be aware of one thing in isolation without also being aware of another thing from which it is distinguished. Conscious awareness thus requires differentiation between multiple entities.

Christian August Crusius challenges Wolff's conception of consciousness as the capacity to differentiate. He contends that differentiation presupposes consciousness, not the reverse (Crusius, 1745, § 444). Crusius critiques what he perceives as Wolff's assumption that "differentiation" involves a negative categorical judgment. He notes that to make a judgment such as "A is not B," one must first be aware of both A and B (Crusius, 1747, §§ 435, 436). Crusius defines consciousness as "distinct sensation" and considers sensation as the origin of the mind's passive faculties (Crusius, 1747, § 86), (see Indregard, 2018, 175).

Kant agrees with Crusius that there is a type of subjective consciousness that is not based on "differentiations" through negative categorical propositions. This is the peculiar "what-it-is-like" aspect of sensible intuition (phenomenal consciousness), which Kant, in agreement with Crusius, calls "sensation." Unlike Crusius, however, Kant still agrees with Wolff when he claims that we are only conscious *of something* to the extent that we "differentiate" it. Suppose you are staring at a perpetually well-camouflaged moth on a tree. You are not conscious of it because you are unable to differentiate it from other things and its background. But Kant distances himself from Wolff when he makes an important distinction between two kinds of "differentiations" that Wolff overlooked:

I would go still further and say: it is one thing to differentiate [unterscheiden] things from each other, and quite another thing to recognize the difference between them [den Unterschied der Dinge zu erkennen]. The latter is only possible by means of judgments and cannot occur in the case of animals, who are not endowed with reason. The following division may be of great use. Logically differentiating means recognizing that [erkennen *daß*] a thing A is not B; it is always a negative judgment. Physically differentiating [physisch unterscheiden] means being driven to different actions by different representations. The dog differentiates the roast from the loaf. It does so because the way in which it is affected by the roast is different from the way in which it is affected by the loaf (for different things cause different sensations). The sensations caused by the roast are a ground of desire in the dog which differs from the desire caused by the loaf, according to the natural connection which exists between its drives and its representations. (*FSS.*, § 6, Ak, 2: 60; 104)

The concept is straightforward: Wolff did not acknowledge that the term "differentiate" carries two distinct meanings. The first is conceptual or logical differentiation, which invariably involves a negative categorical judgment. This form of differentiation is always "de dicto," meaning it is performed by means of a "dictum" or proposition, exemplified by the negative categorical proposition that A is not B. Conversely, the dog is "de re" aware of both the roast and the loaf as it can sensibly discriminate between them. In contrast, my "de dicto" consciousness that the roast is not the loaf manifests as I judge that what is categorized under "roast" does not fall under the category of "loaf."

Now, one may ask how Kant could align with both Crusius and Wolff in the controversy at the same time. The answer is that there are three concepts of consciousness in the debate. Both Crusius and Kant use the term "sensation" to refer to what we call today "phenomenal consciousness," denoting the peculiar subjective aspect of perception/awareness. In contrast, both Wolff and Kant utilize the term "differentiation" to signify the objective character of the same perception. It is here that the idea of "unconscious" representation makes sense: I may represent the mouth on the tree insofar as my vision is somehow affected by the presence of the mouth. Nonetheless, my representation is unconscious since I am unable to discriminate the mouth from its surroundings. Therefore, there is nothing for me that it is like to see the mouth on the tree without discrimination.

Although Kant agrees with Leibniz and Wolff about the existence of "obscure representations," i.e., representations that we possess without being aware of them, he distances himself from the Leibniz-Wolff tradition on three important points. Firstly, Kant rejects Leibniz-Wolff's contrast between obscure and clear representations. Clear representations are either distinct or indistinct. For example, suppose I see a "human [face] far away from me in a meadow" (*Anth*, Ak, 7, 135), although I am not visually aware of its eyes, nose, mouth, etc. We lack "de re" awareness, making this representation "obscure." In contrast, I have a clear "indistinct representation" when I see a human face and can distinguish single it out from the background, without being able to discriminate its features, indicating a partial "de re" awareness of the representation. Finally, when I see a human face can accurately single it out from the background and discern all its relevant features; even without any concept of a human face, I possess a clear and distinct representation: I am consciously representing (see Heidemann, 2017, 50).

Secondly, Kant decisively rejects the rationalist Leibniz-Wolff assumption that all sensible representations are "obscure." Consider again Kant's example of a savage who sees a house from a distance (*JL*. Ak., 9, 33). One could argue that the savage who perceives a house from a distance has an "obscure" sensible representation, assuming that he is able to single out something from its background and surroundings but unable to distinguish all the relevant features of the house: roof, doors, windows, etc. Kant's argument, however, is that it is a mistake to regard all sensible representations as "obscure." Suppose the savage might sensibly discriminate certain features,



such as the roof and the windows, while overlooking others, such as the doors and the balconies. In this context, the savage has an indistinct yet sensible representation of the house. There is also a possibility that the savage discriminates all of the house's important features. In such a scenario, the savage has a clear or distinct sensible representation of the house—a conscious representation—even though he lacks any concepts, including the concept of a house.<sup>1</sup> In both cases, there is something for the savage that is what it is like to see the house at a distance.

If we appropriately isolate sensible from conceptual representations, the overall picture becomes clear. When our senses are "affected," sensation arises. However, in Kantian terms, this is not already what we today call phenomenal consciousness: the what-it-is-like aspect of sensible representation. The conscious character of sensation/*Empfindungen* relies on the fact that the subject consciously represents the object by being able either to discriminate some of its features ("indistinct" representation) or most of them ("distinct" representations). If, on the other hand, we are affected but unable to even single out the object from its background and surroundings, we have an "obscure" or unconscious representation of it (the case of a camouflaged mouth). The objective counterpart of the phenomenal character of sensory perception is, therefore, the ability to single out a particular object against a background. Therefore, "perception," at the very least, as what Kant calls a conscious representation, is "de re" *awareness of* something conceptually indeterminate but presupposing the ability to single out particulars.

Irrespective of "the I think" of apperception, we already have sensible intuition with and without consciousness. The intriguing question is how there can be sensations with and without consciousness: "The field of sensuous intuitions and *sensations of which we are not conscious is immense*" (*Anth*, Ak, 7, 135, emphasis added). If we put Kant aside, a sensation without consciousness seems to be a contradiction in terms. After all, it seems to be an analytical proposition that to feel something is to be conscious of something. If I feel something smooth or soft, how could I not be conscious of such qualities? What I may lack are the respective concepts of smoothness or softness and, therefore, the propositional consciousness that smooth or soft is what I am sensing.

However, Kant's use of the word "sensation" does not match our common use. According to Longuenesse: "a sensation 'with consciousness,' or 'of which the representational subject is conscious,' only if it is seized upon, 'apprehended' by the active capacity of the mind, which affects inner sense with the sensory content thus seized upon" (Longuenesse, 2023, 9). Longuenesse connects Kant's view with "David Rosenthal's theory of consciousness as 'higher-order thought'" (Longuenesse, 2023, footnote 17) with two important caveats. First, the active apprehension of

---

<sup>1</sup> However, as Dietmar H. Heidemann correctly notes, Kant is unfortunately not terminologically rigorous: "In his work, he frequently uses "obscure" in the broadest sense of "unconscious." Thus, he even refers to clear but indistinct representations as obscure representations" (Heidemann, 2017, 50).

sensations is not an act of thinking. Second, consciousness, in general, does not rely on a higher-order representation. I fully concur. However, according to Longuenesse's intellectualist reading, *all* synthesis of apprehension of imagination is determined by conceptual rules, at least insofar as the result is the *representation* of an object (see Longuenesse, 1998, 50).<sup>1</sup>

I am afraid I have to disagree. Considering only the representation of an object, no conceptual activity is required but just the ability to single out something as a unity from its background and its surroundings. Again, I can single out a camouflaged mouth on the tree, if it moves when I am staring it. However, this apprehension is quite independent of my possessing the relevant concept of "mouth" and from any other concepts. Concepts are required only for the cognition (*Erkenntnis*) of a represented object *as an object*, i.e., as a mind-independently or objectively existing particular.

Kant's most cited example of "nonconscious sensation" is the passage quoted above:

The same thing is true of the sensation of hearing when the musician plays a phantasy on the organ with his ten fingers and two feet while, at the same time, he is speaking to someone standing beside him. Within a few seconds, a host of ideas is awakened in his soul, and every idea requires special judgment as to its appropriateness since a single stroke of the finger, not fitted to the harmony, would immediately be heard as discord. Yet the whole comes out so well that the improvisator must often wish to have preserved in a score many a passage which he has performed in this happy fashion but which he could not have performed so well with real diligence and attention. (*Anth*, Ak, 7, 135)

The conversation with someone standing next to the musician awakened various ideas in him as he played the fantasy in the improvisation. He might have lost track of the notes he was playing on the keyboard, which resulted in his listening being devoid of "the subjective character." Under this assumption, the musician has an "unconscious sensation" because he was unable to apprehend, i.e., to discriminate one note from another while he was playing and representing them as a whole. However, this does not justify the assumption that the conceptual rules of understanding, a higher-order cognitive faculty, constrain the apprehension of the notes as a whole piece of music.

---

<sup>1</sup> The passage is the following:

"But to have made a concept of body out of this conjunction is more than the expression of this merely subjective habit. My apprehension of empirically given spatial multiplicities is henceforth guided by the concept. I consider the marks contained in the concept 'body' as necessarily belonging to the given multiplicities" (Longuenesse, 1998, 50). Moreover, when Longuenesse claims that "space and time are given only if understanding determines sensibility" (1998, p. 216), she seems to suggest that Kant should rewrite his *Transcendental Aesthetic* in the second edition since Kant emphatically denies any dependence on space and time on understanding. In private conversation, she denied any suggestion to me. However, how should we make sense of her claims of 1998, 2016

Furthermore, an essential ambiguity surfaces in the ongoing debate, as delineated in Longuenesse's analysis: the distinction between "a sensation with consciousness" and "a sensation of which the representing subject is conscious" (Longuenesse, 2023, 9). Longuenesse, in relating Kant's ideas to those of Rosenthal, appears to encounter the same ambiguity highlighted by Gennaro (Longuenesse, 2023, footnote). This raises a critical question: Is introspection a necessary condition for conscious sensations? Kant describes introspective consciousness as a sophisticated cognitive process that involves apprehension through a synthesis of understanding, aligned with conceptual rules—referred to in contemporary discussions as "phenomenal concepts." Considering a practical example, if a newborn experiences pain from intestinal colic, would its pain sensations be unconscious merely because it lacks the capacity for introspection?

#### **4. The 'I think' is Access Consciousness.**

Kant differentiates himself from the Leibniz-Wolff tradition on a third significant aspect. In Kant's framework, "conscious representations" are those governed by the laws of transcendental apperception or the spontaneous activity of "the I think," which must be capable of accompanying all my representations. Specifically, the intellectual representation "I think" must accompany my representations; otherwise, they would be either (i) "impossible," which implies illogical, or (ii) "be nothing to me" (B131). Here, the term "impossible" denotes illogical. However, this essay focuses on the latter scenario: if "the I think" cannot accompany representations, they are "nothing for me." This situation implies what might be termed "unconscious representations." We return to our initial question: In what sense of "consciousness" are representations that lack the accompaniment of "the I think" considered nonconscious?

These representations are clearly not unconscious in the phenomenal sense—, i.e. they are not free from the subjective character of perception, regardless of whether they are part of a synthetic act of thought. Moreover, they are not unconscious in the sense of lacking "de re awareness," which refers to the ability to distinguish particulars. As Kant says: "The representation that can be given prior to all thinking is called intuition" (B131/B132). So, when representations are described as "nothing for me"—suggesting "unconscious representations"—they have nothing to do with either phenomenal consciousness or "de re" consciousness of particulars.<sup>1</sup>

A dog can distinguish a loaf from a roast and experiences a unique sensory quality when seeing, smelling, and tasting the roast as opposed to the loaf. However, lacking the faculty of "apperception," the dog is unable to conceptualize that loaves are not roasts and vice versa.

---

<sup>1</sup> This independence is enough to reject the proposed close analogy between Kant's emergent theory of cognition and the global workspace theory (GWT) or any higher-order theory of consciousness (HOT). A representation without the accompaniment of the spontaneous "I think" would not qualify as a representation without any kind of consciousness whatsoever.

Similarly, consider a scenario where a person described as a "savage" sees a house from a distance. The object impacts his vision, providing a distinct experiential quality when viewing the house. Assume further that he can differentiate certain features of the house and distinguish it from its surroundings. In this context, he possesses "de re" awareness of the house as an object, yet it remains conceptually undetermined. Given these conditions, to what extent can his representation of the house be considered "unconscious?"

A closer look reveals that the kind of "consciousness" that is missing in the intuitions that are not accompanied by "I think" refers specifically to the "de dicto" or propositional form of consciousness. This form of consciousness involves self-conscious thoughts such as "I think about the object that I represent" or "I think that doors are features of a house." Without propositional consciousness, as in the case of someone who lacks the concept of a "house," the individual is unable to conceptualize that a roof is part of "a house" or to recognize doors and windows as features of the house.

Let us untangle the web of consciousness. The first meaning of consciousness in Kant's work is closely related to what is called "phenomenal consciousness" in the philosophy of mind. "Sensation" becomes conscious when the raw material of sensible intuition is apprehended without any conceptual determination. This apprehension is completely independent of any kind of introspection or self-knowledge. It is a blind synthesis of apprehension, a product of imagination. For example, when I perceive the notes played in an improvisation, there is "something for me that is like" to hear the improvisation, a "what-it-is-like" subjective phenomenal aspect, because I can discriminate the notes and single out the piece of music as a whole, even in the absence of any conceptual determination. The second meaning of consciousness is the objective correlate of conscious sensation, the "de re" consciousness of the conceptually undetermined object of sensible intuition. In this sense, I am "de re aware" of the improvisation since I can discern the notes and single out the piece of music as a whole, independently of my conceptual understanding. And finally, consciousness, or self-consciousness, means propositional consciousness or, more precisely, cognitive access. It is important to note that neither the subjective phenomenal aspect of consciousness nor its objective "de re consciousness" is based on propositional consciousness. Both are perceptual aspects—subjective and objective—while propositional consciousness par excellence is the form of cognition.

Though Kant states that the "manifold of intuition" has "a necessary relation to the 'I think,'" his claim is not that the "I think" *necessarily* accompanies all my intuitions. Rather, Kant uses the modal "must be able:" "The 'I think' *must be able* to accompany all my representations" (B131/B132). This implies that "I think" may not accompany certain representations, i.e., the unconscious representations. However, "I think" must be able to accompany them unless they are

"impossible" (illogical or contradictory) or meaningless to me (lacking consciousness in the propositional sense).

Since "means something for me" means propositional consciousness, the modal "must be able" can only signify what one today calls "cognitive accessibility." The "I think" must be able to accompany my representation is what Kant calls "analytical unity of consciousness." According to the major claim of transcendental deduction, this "analytical unity of consciousness" would not be possible without a synthetic unity:

The analytical unity of consciousness pertains to all common concepts; e.g., if I think of red in general, I, at this moment, represent to myself a feature that (as a mark) can be encountered in anything or that can be combined with other representations; [...] A representation that is to be thought of as common to several must be regarded as belonging to those who, in addition to it, also have something different in themselves (B133, footnote.).

Being conscious of general concepts takes place in judgment and presupposes the synthetic activity of transcendental apperception in judgments. For example, I become propositionally "conscious" of something by means of the concept of red (analytic unity of consciousness) if I am able to judge that red is the color of blood, of apples, of tomatoes etc. (synthetic unity). The "I think" that accompanies my representations, which would otherwise be "unconscious," is thus what makes these representations *available for judgment reason and guiding action* (Block's access to consciousness).<sup>1</sup>

Consequently, intuitions as sensory representations that cannot be accompanied by the propositional attitude "I think that" can still be conscious in a phenomenological sense or in the "de re awareness" of appearances, provided they are not deemed "impossible." Nevertheless, these intuitions remain "unconscious" in the sense of their cognitive "inaccessibility" to propositional thinking, reasoning, and the control of rational action. It is only through the act of "I think" that

---

<sup>1</sup> In Block's words: "A perceptual state is access-conscious [...] if its content [...] can be used to control reasoning and behavior." (Block, 1995, 229.)

Again, Longuenesse appears to share a similar interpretation when she claims: "With intuition, we enter a realm in which phenomenal and what I will call, borrowing the term again from Ned Block, access consciousness (where a representation is "with consciousness" insofar as it is available for judging, reasoning and guiding action) are inseparable." (Longuenesse, 2023, 11).

However, her following remark seems incompatible with Block's famous distinction: "If intuition is a representation of which we are phenomenally conscious (there is something it is like for us to have that intuition), *it is also a representation of which we have access consciousness* (Longuenesse, 2023, 11, emphasis added). According to Block, a representation may be phenomenal consciousness without being cognitively accessible: "These conclusions have not convinced all readers, but the infant color work of Chapter 6 allows *for a quite different argument for phenomenal consciousness without access-consciousness*." (Block, 2023, 421, emphasis added). I believe that what Longuenesse has in mind is the global workspace theory (GWT).

representation gains access to what is referred to as the "global workspace." In Kant's framework, transcendental consciousness serves as the guiding principle for cognitive access to representations. In this context, it resembles the modern theory of the global workspace as a model of cognition rather than a theory of consciousness per se. In the Kantian sense, the global workspace does not account for consciousness but for cognitive accessibility.

We are now better equipped to elucidate the enigmatic passage in paragraph 16 of the transcendental deduction concerning "Meinigkeit" or "mineness." Kant asserts, "For the manifold representations that are given in a certain intuition would not altogether be *my* representations if they did not altogether belong to a self-consciousness" (B132, original emphasis). As argued throughout this discussion, "My representations" do not refer to my mental states, which could be described using predicates (potential self-attribution of experiences). Rather, the possessive first-person pronoun refers to all phenomena that appear to me and are cognitively accessible within the same cognitive apparatus.

Moreover, we are now in a better position to explain the metaphor of "thinking agency." Spontaneous thinking agency is the widespread activity of the mind in the global workspace. It is the thinking activity of the mind that combines sensible and conceptual representations into thoughts, judgments, and inferences, thereby transforming perceptual nonconceptual representations into cognitions ("Erkenntnis").

## 5. Concluding Remarks

Let us return to Kant's argument of the paragraph 16:

1) That I think, must be able to accompany all my representations.

For otherwise, something would be represented in me that could not be thought, which means just as much as:

The representation would either be impossible, Or at least nothing for me. Therefore, that I think must be able to accompany all my representations.

If my reading is correct, then Kant is talking about "unconscious representations" in paragraph 16. These are not representations devoid of phenomenal consciousness, nor are they devoid of "de re" awareness—awareness of something that we single out from its background and discriminate its features. Rather, they are "unconscious" in the sense that they are cognitively inaccessible because they are not accompanied by the representations "I think." They contribute nothing to what Kant calls cognition ("Erkenntnis"). The second premise grounds the first in the simple argument in the sense that without the "I think" representations in me are not representations for me, i.e., cognitively accessible.

Cramer is correct in asserting that the argument extends beyond a mere rephrasing of Kant's well-known principle: "without concepts, intuitions are blind, and without intuitions, concepts are

empty." However, he seems to have overlooked the foundational role of paragraph 16 in transitioning blind intuitions into cognitively accessible representations. It is specifically through the accompaniment of "the I think" that these blind intuitions become accessible to cognition.

This reading has a crucial explanatory value for the question of whether there are nonconceptual mental contents and how this would be possible. The central claim of nonconceptualism is that there are sensory representations of objects that are conceptually indeterminate. Because the crucial distinction between nonconceptual and non-propositional perception and cognition informs all of contemporary cognitive science (see Block, 2023; Burge, 2022), this issue is of utmost importance. Kant's explanation of unconscious representation is strong support for non-conceptualism since sensory representations are conscious in both the "phenomenal" and "de re" awareness senses of "consciousness." They are "unconscious" in the propositional sense because they lack conceptual determination. Kant's example of the house illustrates my view. Whether or not the savage has the concept of a house, the idea is that he can be "de re" aware of the house without the corresponding concept determining the house.

## References

### Works of Kant

References to Kant's works are given in the German Academy edition: *Gesammelte Schriften*, herausgegeben von der Königlich Preussischen Akademie der Wissenschaften, 29 vols. (Berlin: 1902-83; 2d ed., Berlin: De Gruyter, 1968, for vols. I–IX). The abbreviation of the work's title is Ak., volume, and page. For the Critique of Pure Reason, the references are shortened, in keeping with current practice, to the pagination of the original edition, indicated by A for the 1781 edition and B for the 1787 edition.

### Abbreviations of Kant's works

*Anthr*—*Anthropology from a Pragmatic Point of View* (Ak 7). Trans. V.L. Dowdell. Carbondale & Edwardsville: Southern Illinois University Press, 1978.

*BL*—*The Blomberg Logic*, Ak. 24, in *Lectures on Logic*, edited and translated by J. Michael Young. Cambridge University Press, 1992, pp. 15-246.

*FSS*—*Die falsche Spitzfindigkeit der vier syllogistischen Figuren* Ak. 2 (1762). *The False Subtlety of the Four-Syllogistic Figures*, translated by David Walford in collaboration with Ralf Meerbote. Edited by Paul Guyer and Allen W. Wood. Cambridge: Cambridge University Press, 1992.

*HN*—*Handschriftliche Nachlaß* Ak. 14-19. *Notes and Fragments*, translated by Curtis Bowman, Paul Guyer, Frederick Rauscher, edited by Paul Guyer. Cambridge: Cambridge University Press, 2005.

*JL*—*Jäsche Logik*, Ak. 9 (1800). *Logic*, edited by J. B., in *Lectures on Logic*, edited and translated by J. Michael Young. Cambridge: Cambridge University Press, 1992, pp. 521-640.

*VM*—*Vorlesungen über die Metaphysik* Ak. 28 (1821). *Lectures on Metaphysics* edited and translated by Paul Guyer and Allen W. Wood. Cambridge University Press, 1997.

### Secondary Literature

Baars, B. J. (1988). *A Cognitive Theory of Consciousness*. Cambridge University Press.

- Baker, L. R. (1998). The First-Person Perspective: A Test for Naturalism. *American Philosophical Quarterly*, 35 (4), 327–348.
- Block, N. (1995). On a confusion about a function of consciousness. *Brain and Behavioral Sciences*, 18 (2), 227–247.
- Block, N. (2023). *The Border Between Seeing and Thinking*. OUP USA.
- Burge, T. (2022). *Perception: First Form of Mind*. Oxford University Press.
- Cassam, Q. (1997). *Self and World*. Oxford University Press.
- Cramer, K. (1987). Über Kants Satz: Das: Ich denke, muß alle meine Vorstellungen begleiten können. In *Theorie der Subjektivität*, Edited by K. Cramer et al., Frankfurt am Main: 167–202.
- Cramer, K. (2014). Kant's "I think" and Fichte's "I am." *Sententiae*, 31 (2), 25-51. <http://dx.doi.org/10.22240/sent31.02.025>
- Dehaene, & et al. (2006). Conscious, preconscious, and subliminal processing: A testable taxonomy. *Trends in Cognitive Sciences*, 10(5), 204-211. <http://dx.doi.org/10.1016/j.tics.2006.03.007>
- Gennaro, Rocco J. (1996). *Consciousness and Self-Consciousness: A Defense of the Higher-Order Thought Theory of Consciousness*. John Benjamins.
- Heidemann, D. H. (2012). The 'I Think' Has to Be Possible to Accompany All My Representations: Kant's Theory of Apperception and the Unconscious. In book: Kant's Philosophy of the Unconscious (pp.37-60) <http://dx.doi.org/10.1515/9783110265408.37>
- Indregard, J. J. (2018). Consciousness as Inner Sensation: Crusius and Kant. *Ergo: An Open Access Journal of Philosophy*, 5 (7), 173–201. <http://dx.doi.org/10.3998/ergo.12405314.0005.007>
- Leibniz, G. W. (1704). Nouveaux Essais sur l'Entendement Humain. In *Die philosophischen Schriften von Gottfried Wilhelm Leibniz*, Vol. 5, Edited by C. I. Gerhardt. Flammarion
- Leibniz, G. W. (1714). Les Principes de la Philosophie ou la Monadologie. In *Die philosophischen Schriften von Gottfried Wilhelm Leibniz*, Vol. 6, Edited by C. I. Gerhardt. Presses universitaires de France
- Longuenesse, B. (2006). Self-Consciousness and Consciousness of One's Own Body. Variations on a Kantian Theme. [Em: http://philosophy.fas.nyu.edu/docs/IO/2575/longuenesse.pdf](http://philosophy.fas.nyu.edu/docs/IO/2575/longuenesse.pdf)
- Longuenesse, B. (2007). Kant's 'I think' versus Descartes' 'I am a Thing that Thinks.' In *Kant and the Early Moderns*, edited by Beatrice Longuenesse and Daniel Garber, 9–31. Princeton University Press.
- Longuenesse, B. (2012). Two Uses of 'I' as Subject? In Simon Prosser & François Recanati (Eds.), *Immunity to Error through Misidentification*. Cambridge University Press <http://dx.doi.org/10.1080/00048402.2014.989243>
- Longuenesse, B. (2017). *I, Me, Mine: Back to Kant, and Back Again*. Oxford University Press.
- Longuenesse, B. (2023). Kant on Consciousness and Its Limits. *Revue de Métaphysique et de Morale*, 117(1), 7-26
- Rozzo, & M. Sgarbi (Eds.), *Kant's Philosophy of the Unconscious* (pp. 37–60). de Gruyter.
- Strawson, P. F. (1959). *Individuals: An Essay in Descriptive Metaphysics*. Methuen.
- Strawson, P. F. (1966). *The Bounds of Sense*. Methuen.
- Tugendhat, E. (1979). *Selbstbewusstsein und Selbstbestimmung. Sprachanalytische Interpretationen*. English Translated by P. Stern. *Self-Consciousness and Self-Determination*. MIT Press, 1986.





Journal of Philosophical Investigations

## Journal of Philosophical Investigations

Print ISSN: 2251-7960 Online ISSN: 2423-4419

Homepage: <https://philosophy.tabrizu.ac.ir>



University of Tabriz

# The Mathematical Basis of the Phenomenal World

Bernard Riley 

Independent Researcher, UK. Email: [bernardriley125@gmail.com](mailto:bernardriley125@gmail.com)

### Article Info

#### Article type:

Research Article

#### Article history:

Received 24 May 2024

Received in revised form  
05 July 2024

Accepted 10 July 2024

Published online 07  
August 2024

#### Keywords:

Number, Cognition,  
Concept, Mathematics,  
Science, Phenomena

### ABSTRACT

In the *Critique of Pure Reason* Immanuel Kant said that cognition (objective perception) is acquired in the unity of sensibility (the receptivity of the mind to receive empirical representations of things, which yields intuitions) and the understanding (in which concepts – general representations of things – arise) and is mediated by the imagination. Here, it is shown that numbers, either pure or denominate, are cognized in the synthesis of intuition and mathematical concept, and that the phenomenal world of the cognizer is shaped accordingly. Any number can be related to any other number through a *general* mathematical formula conceived by the cognizer for the purpose. The judgment of the cognizer is manifest in the *specifics* of the mathematical relationship established between the two numbers in cognition. If the cognized number is the numerical value of a physical constant, then in the (consistent) phenomenal world it will always have been of the value found in cognition, which explains why the universe seems to be fine-tuned for life. If the cognized number is the numerical value of a physical variable, then the number will be subject to change in accordance with physical laws. Symmetry is a recurrent feature of the phenomenology. A mathematical formula conceived by the cognizer may also relate, one to one, the numerical values of quantities in one set with the numerical values of quantities of different dimensionality in another set, which suggests that physical laws are human inventions and that causality is a pure concept of the understanding.

**Cite this article:** Riley, B. (2024). The Mathematical Basis of the Phenomenal World. *Journal of Philosophical Investigations*, 18(47), 161-188. <https://doi.org/10.22034/jpiut.2024.62355.3811>



© The Author(s).

<https://doi.org/10.22034/jpiut.2024.62355.3811>

Publisher: University of Tabriz.

## Introduction

After discovering that the results of a long-term investigation into the values of all sorts of physical parameters had been subjective (Riley, 2020), the author read Immanuel Kant's *Critique of Pure Reason* (Kant, 1998) and comprehended that the value of any number, such as the numerical value of a physical parameter, is cognized in the synthesis of intuition and mathematical concept (Riley, 2022). The results of the conceptual observation of both pure and denominate numbers are shown here graphically and through mathematical equations. The essay concludes with a short discussion.

The data used include values of the fundamental physical constants (NIST, 2024), particle mass (Particle Data Group, 2024), solar system parameters (NASA, 2024) and cosmological parameters (Planck Collaboration, 2020). All data values used are stated. Central values have been used throughout; some presented values have been rounded to a smaller number of decimal places.

## 1. The Cognition of Numbers

### 1.1. The Process of Cognition

The process of cognition employed by the author makes use of a mathematical formula to map the number of interests,  $N$ , to a number,  $n$ , the value of which explains, to the satisfaction of the cognizer, the value of  $N$ . The explanation may have a physical basis but that is not essential. Commonly, the formula maps  $N$  to  $n$  so that  $N = b^n$ , where  $b$  is a number chosen by the cognizer. The quantity  $b^n$  is a power of the base  $b$ ;  $n$  is the exponent of the power. The number  $n$  takes a rational value, at least notionally, by being located within an infinitely divisible discrete numerical framework, e.g. within a framework in which numbers take integer, half-integer, quarter-integer, etc values. The formula and the numerical framework together constitute the mathematical concept that relates  $N$  and  $n$ . If  $N$  is the value of an interesting quantity,  $n$  takes an integer or low-denomination fractional value. Since rational numbers are constructed from integers, which are exact numbers, they signal the truth; in consequence of that, rational values of  $n$  communicate the truth of the explanatory mathematical concept by which they are produced, and therefore the truth of the number of interests,  $N$ . When produced from values of  $N$  that characterize interesting quantities, integer  $n$  signals the truth forcefully. The concept then often finds wider utility.

### 1.2. On Particle Scales

The first concept employed by the author, many years before the subjective nature of the investigation was realized, was that particle masses derive directly from the Planck mass,  $m_{\text{Planck}}$  – a ‘natural’ mass scale, constructed from the speed of light, the reduced Planck constant and the gravitational constant, and equal to  $2.176 \times 10^{-8}$  kg or  $1.221 \times 10^{19}$  GeV – through division<sup>1</sup> by

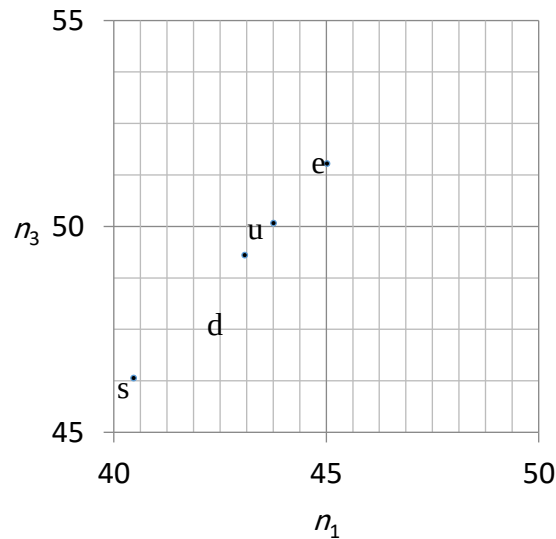
---

<sup>1</sup> Particle masses are smaller than the Planck mass.

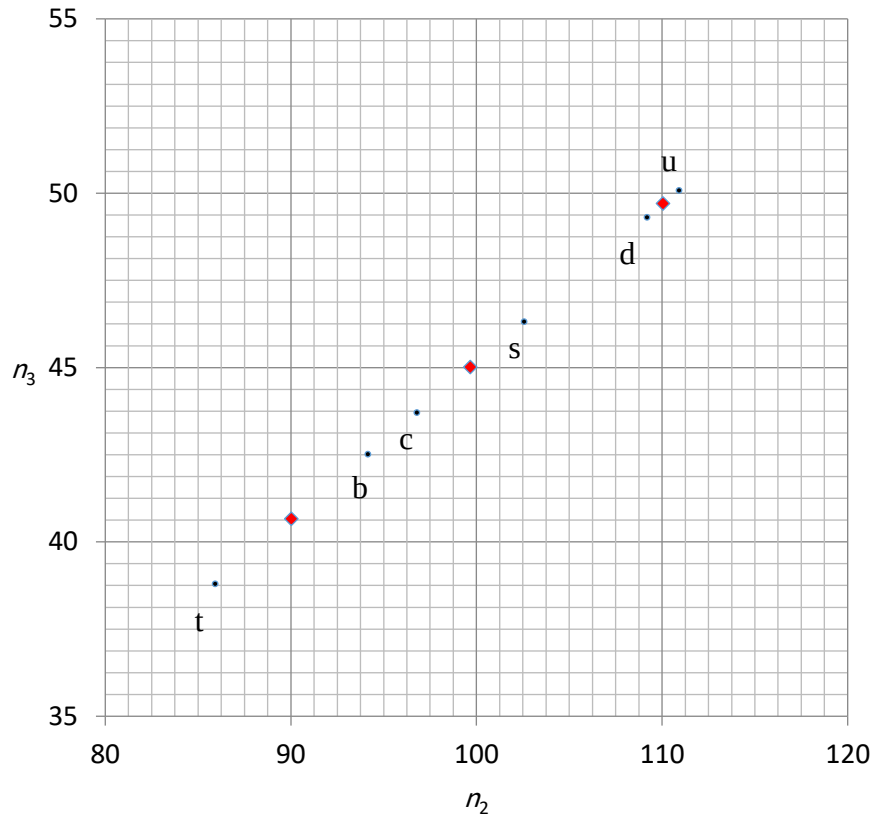
rational powers of a number, which had to be determined (Riley, 2003). In time, three such numbers were identified:  $\pi$ ,  $\pi/2$  and  $e$ , which relate to the geometry of hypothetical warped extra dimensions (Riley, 2008); two warped geometry models had been proposed a few years previously by Randall & Sundrum (1999a&b). Rational powers of  $\pi$ ,  $\pi/2$  and  $e$  were then found for all particles; integer and low-denomination fractional powers were found for interesting particles.

The masses in Planck units of the electron and the light quarks (up, down and strange) are shown as rational powers of  $\pi$  and  $e$  on the ‘levels’ of Figure 1; the level structure has been chosen by the author. Since  $n_1$  and  $n_3$  are in constant ratio the points lie on a straight line. The electron and up quark (the lightest quark) occupy ‘superlevels’, which are levels that are characterized by level-numbers that are multiples of 5; such levels are evidently reserved for very interesting particles.

The masses in Planck units of the six quarks are shown to be equal to rational powers of  $\pi$  and  $e$  in Figure 2. The quarks of each family (up and down; strange and charm; top and bottom) are symmetrically arranged about superlevels in ‘partnerships’. In general, if two objects of similar mass are judged to be closely related then they will be found to be arranged symmetrically in terms of the concept. Interestingly, when the up and down quarks had not been considered as partners they were not found to be symmetrically arranged (Figure 1).



**Figure 1.** Points  $(n_1, n_3)$  representing the masses of the electron and the light (up, down and strange) quarks in Planck units. Each particle mass had first been expressed as  $\pi^{-n_1} m_{\text{Planck}}$  and  $e^{-n_3} m_{\text{Planck}}$ .  
electron: 0.511 MeV up quark: 2.16 MeV down quark: 4.70 MeV strange quark: 93.5 MeV



**Figure 2.** Points  $(n_2, n_3)$  representing the masses of the six quarks in Planck units. Each quark mass had first been expressed as  $(\pi/2)^{-n_2} m_{\text{Planck}}$  and  $e^{-n_3} m_{\text{Planck}}$ . Diamonds mark the points of mean  $n_2$  and  $n_3$  for each pair of quarks: u-d, s-c and b-t.

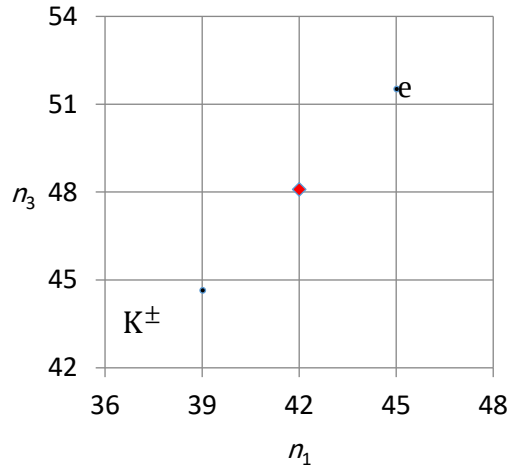
up: 2.16 MeV    down: 4.70 MeV    strange: 93.5 MeV    charm: 1.2730 GeV

bottom: 4.183 GeV    top: 172.57 GeV

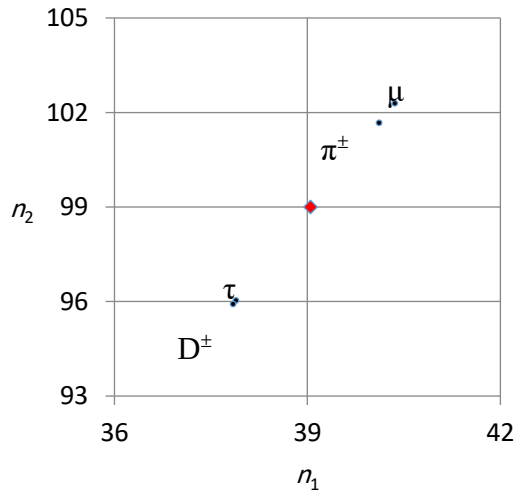
Knowing that the (spin- $\frac{1}{2}$ ) muon and the (spin-0) pseudoscalar mesons  $\pi^\pm$ , and also that the (spin- $\frac{1}{2}$ ) tau lepton and the (spin-0) pseudoscalar mesons  $D^\pm$ , are close in mass, and with supersymmetry<sup>1</sup> in mind, the author has found that each of the three charged leptons (electron, muon and tau lepton) forms a partnership with pseudoscalar mesons that is characterized by spin difference  $\frac{1}{2}$ . The three charged lepton-pseudoscalar meson partnerships are shown in Figures 3 and 4. The electron and  $K^\pm$  mesons occupy levels whose level-numbers are multiples of 3, and the particles are arranged symmetrically about an intersection of levels characterized by level-numbers that are multiples of 3. The muon and  $\pi^\pm$  mesons, and the tau lepton and  $D^\pm$  mesons, are arranged

<sup>1</sup> A symmetry between particles with integer spin and those with half-integer spin.

symmetrically about levels whose level-numbers are multiples of 3, and the two partnerships are arranged symmetrically about an intersection of levels whose level-numbers are multiples of 3. Evidently, levels characterized by level-numbers that are multiples of 3 are sometimes (in addition to levels whose level-numbers are multiples of 5) considered by the cognizer to be superlevels and as such are populated by very interesting particles.



**Figure 3.** Points  $(n_1, n_3)$  representing the masses of the electron and the  $K^\pm$  mesons (493.68 MeV) in Planck units. Each particle mass had first been expressed as  $\pi^{-n_1} m_{\text{Planck}}$  and  $e^{-n_3} m_{\text{Planck}}$ . A diamond marks the point of mean  $n_1$  and  $n_3$ .



**Figure 4.** Points  $(n_1, n_2)$  representing the masses of (i) the muon,  $\mu$  (105.66 MeV), and the  $\pi^\pm$  mesons (139.57 MeV) and (ii) the tau lepton,  $\tau$  (1776.93 MeV), and the  $D^\pm$  mesons (1869.66 MeV), in Planck units. Each particle mass had

first been expressed as  $\pi^{-n_1} m_{\text{Planck}}$  and  $(\pi/2)^{-n_2} m_{\text{Planck}}$ . A diamond marks the point of mean  $n_1$  and  $n_2$  for the four masses

Immediately on turning to length scales, the following expression was found for the Bohr radius – a physical constant that provides a measure of the size of the hydrogen atom – in terms of the Planck length,  $1.616 \times 10^{-35}$  m (Riley, 2013). Here and throughout the essay all figures in the exponents of powers, e.g. in the number 125.00 below, are significant.

$$\text{Bohr radius } (5.292 \times 10^{-11} \text{ m}) \quad (\pi/2)^{125.00} l_{\text{Planck}} \quad (1)$$

It follows from the value of the Bohr radius in Planck units that the mass of the electron is given by the following expression, where  $\alpha$  is the fine structure constant ( $=1/137.036$ ).

$$\text{electron mass } (0.511 \text{ MeV}) \quad \alpha^{-1}(\pi/2)^{-125.00} m_{\text{Planck}} \quad (2)$$

Imagining that the masses of the three up-type quarks (up, charm and top) are related to the mass of the electron through symmetry, the expressions shown below were discovered.

$$\text{up quark mass } (2.16 \text{ MeV}) \quad \alpha(\pi/2)^{-100.0} m_{\text{Planck}} \quad (3)$$

$$\text{charm quark mass } (1.27 \text{ GeV}) \quad \alpha^2(\pi/2)^{-75.0} m_{\text{Planck}} \quad (4)$$

$$\text{top quark mass } (172.57 \text{ GeV}) \quad \alpha(\pi/2)^{-75.0} m_{\text{Planck}} \quad (5)$$

The following expression has been found for the mass of the Higgs boson. The power of 2 is exact.

$$\text{Higgs boson mass } (125.20 \text{ GeV}) \quad 2^{25}(\pi/2)^{-125.00} m_{\text{Planck}} \quad (6)$$

The following expressions have been found for the masses of the up-type quarks in kilograms.

$$\text{up quark mass } (2.16 \text{ MeV}) \quad (\pi/2)^{-150.0} \text{ kg} \quad (7)$$

$$\text{charm quark mass } (1.27 \text{ GeV}) \quad \alpha(\pi/2)^{-125.0} \text{ kg} \quad (8)$$

$$\text{top quark mass } (172.57 \text{ GeV}) \quad (\pi/2)^{-125.0} \text{ kg} \quad (9)$$

The following expression for the kilogram was then known.

$$\text{Kilogram} \quad \alpha(\pi/2)^{50.0} m_{\text{Planck}} \quad (10)$$

Arresting expressions have also been found for the meter and second in Planck units.

$$\text{Meter} \quad \pi^{70.0} l_{\text{Planck}} \quad (11)$$

$$\text{Second} \quad e^{100} t_{\text{Planck}} \quad (12)$$

Arresting expressions may be found for quantities in any appropriate unit:

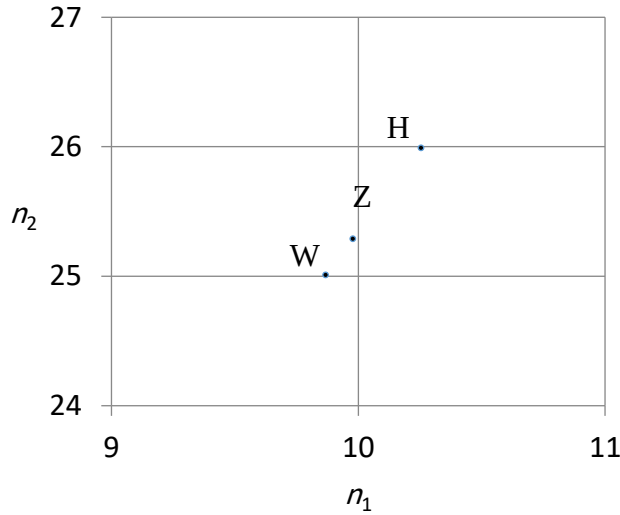
$$\text{W boson mass (80.37 GeV)} \quad (\pi/2)^{25.0} \text{ MeV} \quad (13)$$

$$\text{Z boson mass (91.19 GeV)} \quad \pi^{10.0} \text{ MeV} \quad (14)$$

$$\begin{aligned} \text{reduced Planck constant } (\hbar) \\ (6.582 \times 10^{-16} \text{ eV.s}) \end{aligned} \quad e^{-35.0} \text{ eV.s} \quad (15)$$

$$\begin{aligned} \text{speed of light } (2.998 \times 10^8 \text{ m.s}^{-1} \\ \text{or } 6.706 \times 10^8 \text{ mph}) \end{aligned} \quad (\pi/2)^{45.0} \text{ mph} \quad (16)$$

The masses in MeV of the W, Z and Higgs bosons are shown as integer powers of  $\pi$  and  $\pi/2$  in Figure 5.



**Figure 5:** Points  $(n_1, n_2)$  representing the masses of the W, Z and Higgs bosons in MeV. Each particle mass had first been expressed as  $\pi^{-n_1}$  MeV and  $(\pi/2)^{-n_2}$  MeV.

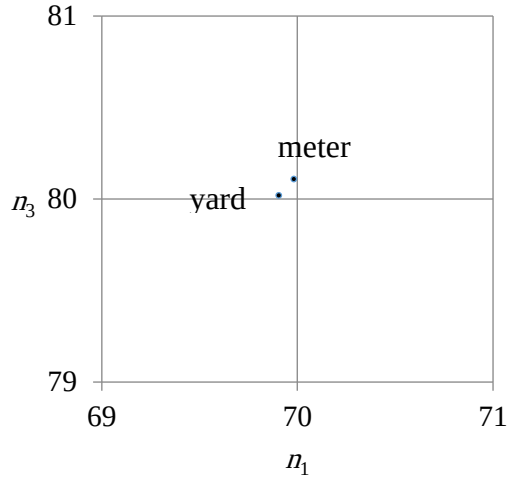
### 1.3. On Human Scales

Some common units besides the kilogram, meter and second have been found to have arresting values in Planck units, as shown in Figures 6, 7 and 8. The meter and yard (0.9144 m) lie at the intersection (70, 80) in Figure 6. The expression for the meter was given in (11). The expression for the yard is shown below.

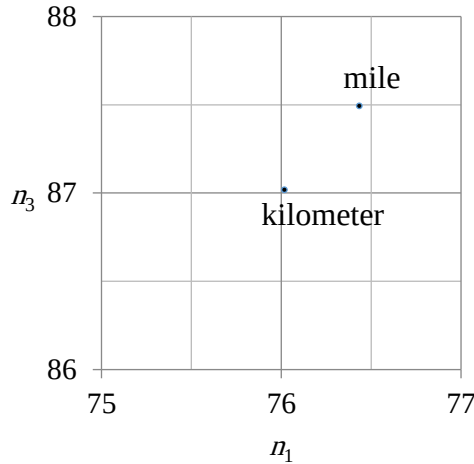
$$\text{Yard} \quad e^{80.0} l_{\text{Planck}} \quad (17)$$

The following expression has been found for the mean calendar month (30.44 days).

$$\text{Month} \quad \pi^{100} t_{\text{Planck}} \quad (18)$$

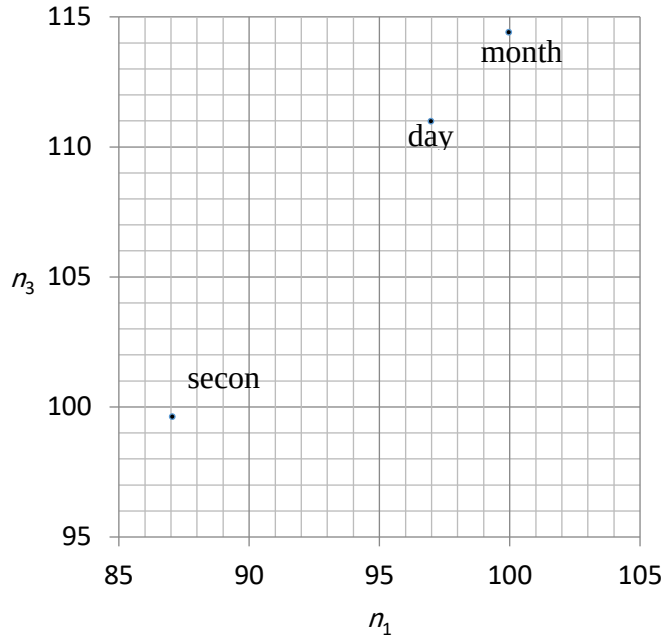


**Figure 6.** Points  $(n_1, n_3)$  representing the lengths of the meter and yard in Planck units. Each length had first been expressed as  $\pi^{n_1} l_{\text{Planck}}$  and  $e^{n_3} l_{\text{Planck}}$ .



**Figure 7.** Points  $(n_1, n_3)$  representing the lengths of the kilometer and mile (1.609 km) in Planck units. Each length had first been expressed as  $\pi^{n_1} l_{\text{Planck}}$  and  $e^{n_3} l_{\text{Planck}}$ .



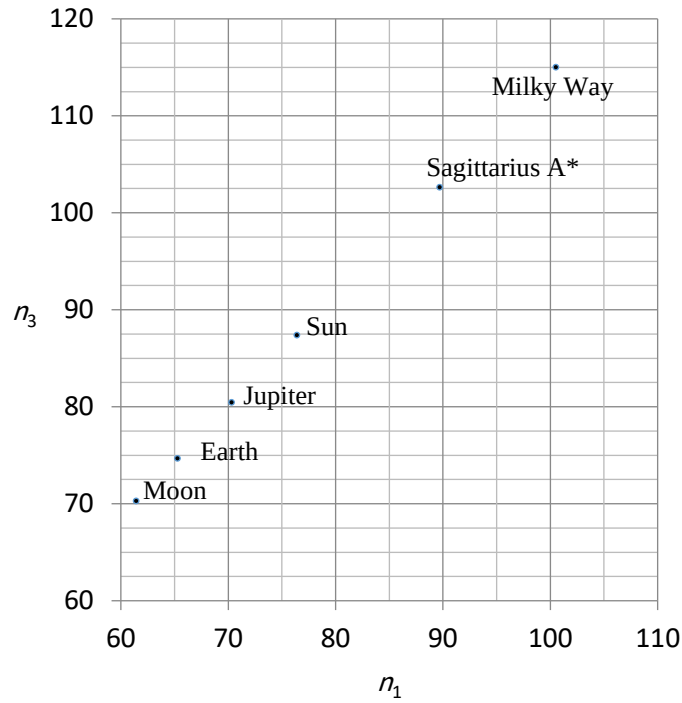


**Figure 8.** Points  $(n_1, n_3)$  representing the lengths of time of the second, day and mean calendar month (30.44 days) in Planck units. Each length of time had first been expressed as  $\pi^{n_1} t_{\text{Planck}}$  and  $e^{n_3} t_{\text{Planck}}$ .

#### 1.4. On Astronomical Scales

The masses, first in Planck units and then in kilograms, of various astronomical objects are shown as rational (integer in most cases) powers of  $\pi$  and  $e$  in Figures 9 and 10. Sagittarius A\* is the supermassive black hole located in the centre of the Milky Way. The mass of the ‘very interesting’ earth is  $\approx \pi^{50}$  kg.

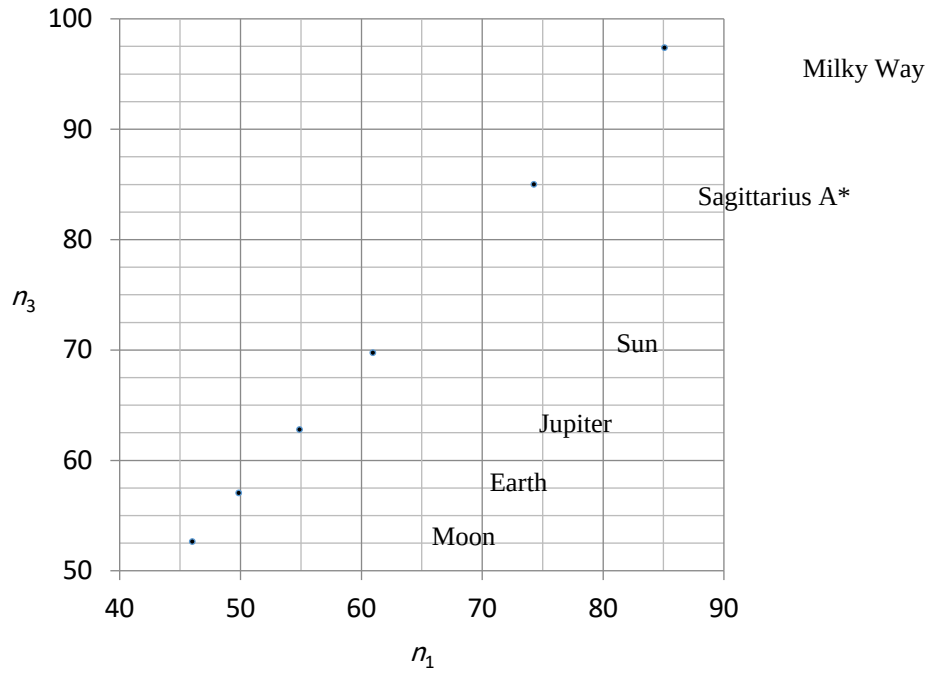
The lengths, first in Planck units and then in meters, of various astronomical distances are shown as rational (integer in most cases) powers of  $\pi$  and  $e$  in Figures 11 and 12. The Hubble length is the distance from earth at which galaxies are currently receding from us at the speed of light due to the expansion of space. Notice that the semi-major axis of the ‘very interesting’ moon’s orbit, which closely approximates to its mean distance from earth, is  $\approx e^{100} l_{\text{Planck}}$ .



**Figure 9.** Points  $(n_1, n_3)$  representing the masses of six astronomical objects in Planck units. Each mass value had first been expressed as  $\pi^{n_1} m_{\text{Planck}}$  and  $e^{n_3} m_{\text{Planck}}$ .

Moon:  $7.346 \times 10^{22}$  kg    Earth:  $5.972 \times 10^{24}$  kg    Jupiter:  $1.898 \times 10^{27}$  kg    Sun:  $1.988 \times 10^{30}$  kg

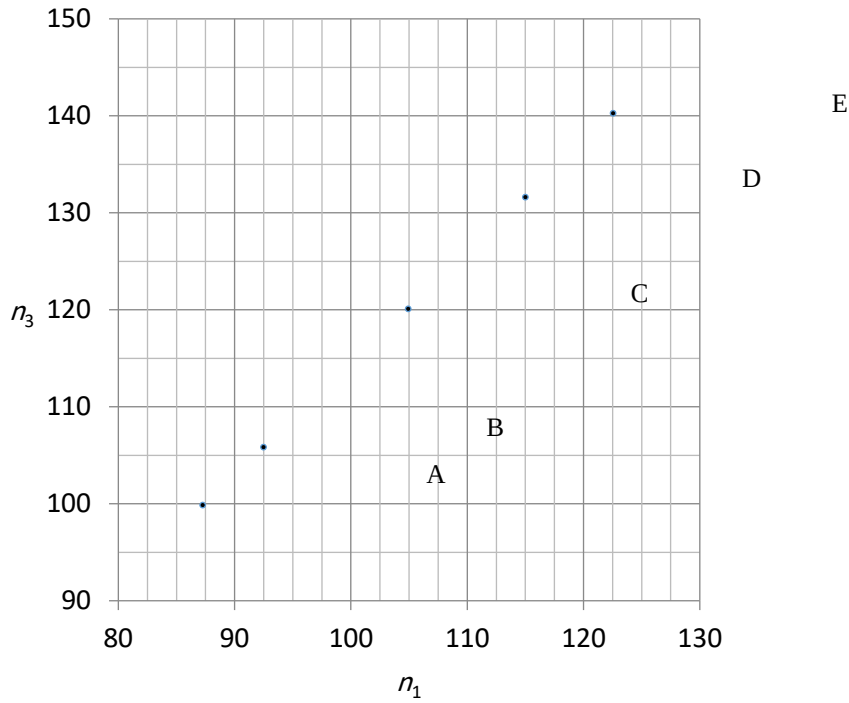
Sagittarius A\*:  $8.35 \times 10^{36}$  kg ( $4.2 \times 10^6 M_{\text{sun}}$ )    Milky Way:  $\sim 2 \times 10^{42}$  kg ( $\sim 1 \times 10^{12} M_{\text{sun}}$ )



**Figure 10.** Points  $(n_1, n_3)$  representing the masses of six astronomical objects in kilograms. Each mass value had first been expressed as  $\pi^{n_1}$  kg and  $e^{n_3}$  kg. Companion to Figure 9.

Moon:  $7.346 \times 10^{22}$  kg    Earth:  $5.972 \times 10^{24}$  kg    Jupiter:  $1.898 \times 10^{27}$  kg    Sun:  $1.988 \times 10^{30}$  kg

Sagittarius A\*:  $8.35 \times 10^{36}$  kg ( $4.2 \times 10^6 M_{\text{sun}}$ )    Milky Way:  $\sim 2 \times 10^{42}$  kg ( $\sim 1 \times 10^{12} M_{\text{sun}}$ )



**Figure 11.** Points  $(n_1, n_3)$  representing five astronomical distances in Planck units. Each distance had first been expressed as  $\pi^{n_1} l_{\text{Planck}}$  and  $e^{n_3} l_{\text{Planck}}$  (Riley, 2020).

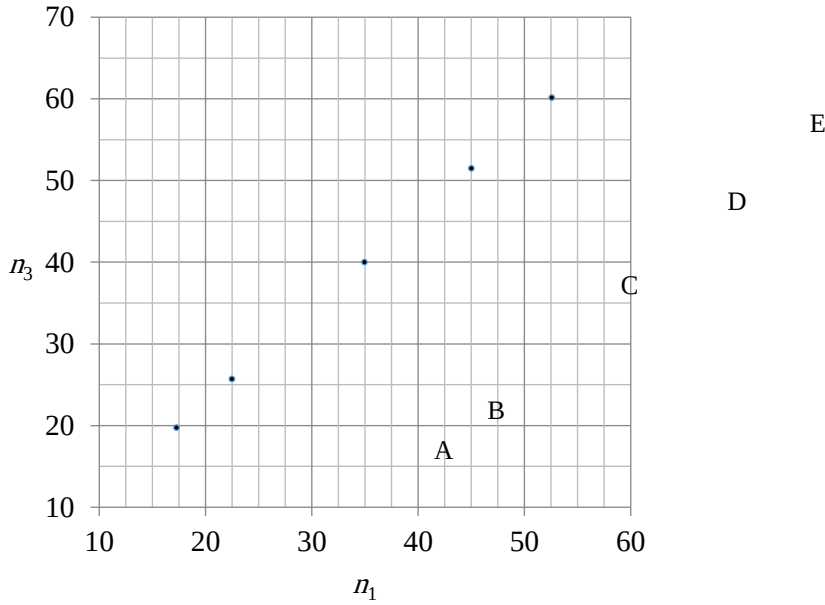
A Semi-major axis of the Moon's orbit ( $3.844 \times 10^8$  m)

B Semi-major axis of the Earth's orbit ( $1.496 \times 10^{11}$  m)

C Distance from Earth of the prominent star Vega (25.0 lyr)

D Distance from Earth of the Andromeda Galaxy (M31) (2.5 Mlyr)

E Hubble length (14.4 Glyr)



**Figure 12.:** Points  $(n_1, n_3)$  representing five astronomical distances in meters. Each distance had first been expressed as  $\pi^{n_1}$  m and  $e^{n_3}$  m. Companion to Figure 11.

ASemi-major axis of the Moon's orbit,  $3.844 \times 10^8$  m

B Semi-major axis of the Earth's orbit,  $1.496 \times 10^{11}$  m

C Distance from Earth of the prominent star Vega, 25.0 lyr

D Distance from Earth of the Andromeda Galaxy (M31), 2.5 Mlyr

E Hubble length, 14.4 Glyr

In terms of the explanatory concept used, various parameters of the sun and moon take arresting values in Planck units, as shown below.<sup>1</sup>

Sun:

$$\text{radius } (6.957 \times 10^5 \text{ km}) \quad e^{100} l_{\text{Planck}} \quad (19)$$

$$\text{rotation period } (\approx 27 \text{ days}) \quad \pi^{100} t_{\text{Planck}} \quad (20)$$

$$\text{central temperature } (1.571 \times 10^7 \text{ K}) \quad \pi^{-50} T_{\text{Planck}} \quad (21)$$

$$\begin{aligned} \text{angular momentum} \quad & 2^{250} \hbar \\ (1.93 \times 10^{41} \text{ kg} \cdot \text{m}^2 \cdot \text{s}^{-1} \text{ (Riley, 2018b)}) \end{aligned} \quad (22)$$

Moon:

<sup>1</sup> More precisely,  $R_{\text{sun}} = e^{100.47} l_{\text{Planck}}$  and  $d_{\text{moon}} = e^{99.88} l_{\text{Planck}}$

$$\text{semi-major axis } (3.844 \times 10^5 \text{ km}) \quad e^{100} l_{\text{Planck}} \quad (23)$$

$$\text{orbital period } (27.32 \text{ days}) \quad \pi^{100} t_{\text{Planck}} \quad (24)$$

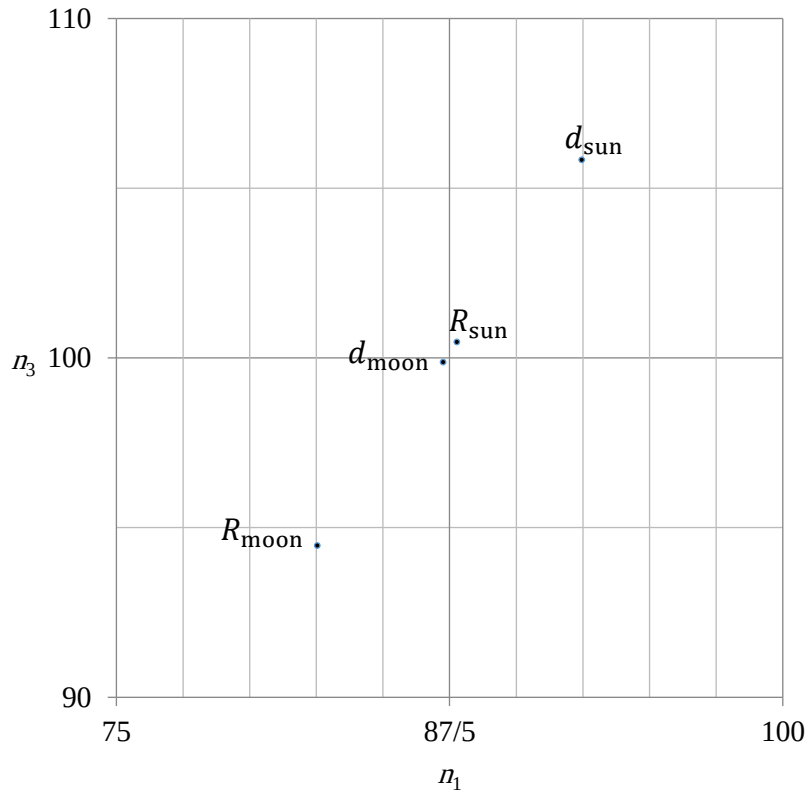
Certain parameters of the sun and moon are arranged symmetrically in terms of the concept, reflecting the fact that the sun dominates the sky during the day and the moon dominates the sky during the night, the two bodies thereby forming a partnership, at least in the mind of the cognizer. After expressing the radii of the sun and the moon,  $R_{\text{sun}}$  and  $R_{\text{moon}}$  ( $1.737 \times 10^3 \text{ km}$ ), and the distances from earth of the two bodies,  $d_{\text{sun}}$  ( $1.496 \times 10^8 \text{ km}$ ) and  $d_{\text{moon}}$ , each as  $\pi^{n_1} l_{\text{Planck}}$  and  $e^{n_3} l_{\text{Planck}}$ , each length has been represented in Figure 13 by a point  $(n_1, n_3)$ . The four points are arranged about the intersection  $(87.5, 100)$ . The values of  $n_1$  for  $R_{\text{sun}}$  and  $d_{\text{moon}}$  are symmetrically arranged about  $n_1 = 87.5$ , as are the values of  $n_1$  for  $R_{\text{moon}}$  and  $d_{\text{sun}}$ . The mean value of  $n_1$  for  $R_{\text{sun}}$  and  $d_{\text{moon}}$  is 87.51. The mean value of  $n_1$  for  $R_{\text{moon}}$  and  $d_{\text{sun}}$  is 87.50. It follows that,

$$R_{\text{sun}} d_{\text{moon}} = R_{\text{moon}} d_{\text{sun}} \quad (25)$$

and therefore,

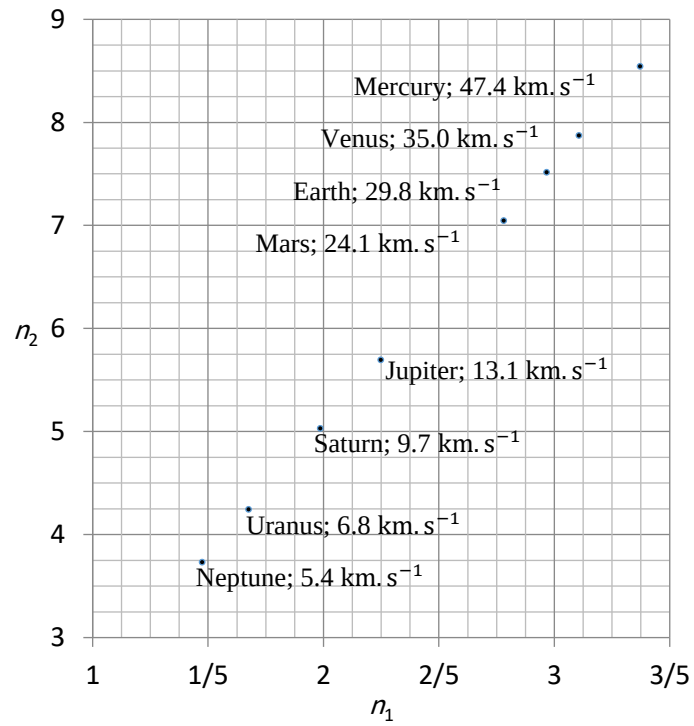
$$\frac{R_{\text{sun}}}{d_{\text{sun}}} = \frac{R_{\text{moon}}}{d_{\text{moon}}} \quad (26)$$

which indicates that the sun and moon are the same size in the sky. It is precisely because the sun and moon are regarded by the observer as a partnership that the process of cognition carried out by the observer results in them being of equal size in the sky.



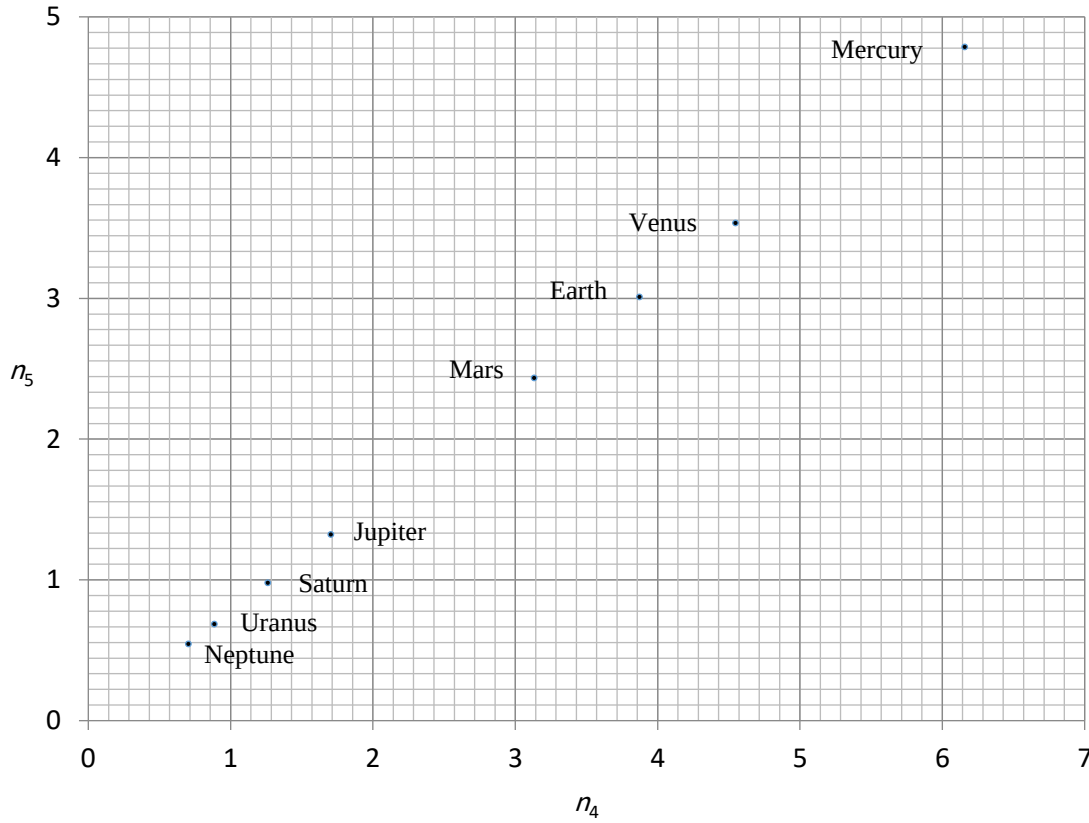
**Figure 13.** Points  $(n_1, n_3)$  representing the mean distances from earth ( $d_{\text{sun}}$  and  $d_{\text{moon}}$ ) and the radii ( $R_{\text{sun}}$  and  $R_{\text{moon}}$ ) of the sun and moon in Planck units. Each length is expressed as  $\pi^{n_1} l_{\text{Planck}}$  and  $e^{n_3} l_{\text{Planck}}$ .

Any concept can be used successfully in the cognition of a number. The mean orbital velocity of each planet of the solar system, in numbers of  $\text{km} \cdot \text{s}^{-1}$ , is shown to be a rational power of either  $\pi$  or  $\pi/2$  in Figure 14. Each velocity,  $v$  in numbers of  $\text{km} \cdot \text{s}^{-1}$ , is then shown in Figure 15 to be either a rational number  $n_4$  or a rational number  $n_5$ , where  $n_4 = v(\text{km} \cdot \text{s}^{-1})/7.7$  and  $n_5 = v(\text{km} \cdot \text{s}^{-1})/9.9$ . The points  $(n_4, n_5)$  have been located within a unique numerical framework: the minor units are sevenths and ninths.



**Figure 14.** Points  $(n_1, n_3)$  representing the mean orbital velocities of the planets in  $\text{km.s}^{-1}$ . Each velocity had first been expressed as  $\pi^{n_1} \text{ km.s}^{-1}$  and  $(\pi/2)^{n_2} \text{ km.s}^{-1}$ .





**Figure 15.** Points  $(n_4, n_5)$  representing the mean orbital velocities of the planets in  $\text{km.s}^{-1}$ . Each velocity,  $v$ , had first been expressed as  $n_4 = v(\text{km.s}^{-1})/7.7$  and  $n_5 = v(\text{km.s}^{-1})/9.9$ . Companion to Figure 14.

### 1.5. On Cosmic Scales

The total mass + energy content of the (critically dense) observable universe of radius 46.6 Glyr and Hubble constant  $67.4 \text{ km.s}^{-1} \text{ Mpc}^{-1}$  is of mass equivalent  $3.06 \times 10^{54} \text{ kg}$ . In Planck units, this quantity can be expressed as follows:

$$\begin{array}{ll} \text{mass + energy of the observable} & \pi^{125.0} m_{\text{Planck}} \\ \text{universe (mass equivalent)} & \end{array} \quad (27)$$

With a dark energy density parameter,  $\Omega_\Lambda$ , of 0.685 (Planck Collaboration, 2018) the dark energy content can be expressed as follows.

$$\begin{array}{ll} \text{dark energy of the observable} & e^{125.1} \text{ kg} \\ \text{universe (mass equivalent)} & \end{array} \quad (28)$$

Both of these ‘very interesting’ quantities (total mass + energy and dark energy content) have very interesting numerical values.

### 1.6. On Assorted Scales

Employing a new concept (distances in meters take numerical values that are rational powers of 2.2 and 3.3, chosen arbitrarily), various distances have been found (in succession and in the same order as presented in Figure 16) to take interesting values (Riley, 2021). Bursts of discovery such as this can occur with a new concept.

In Figure 16(a), the distance from earth of the Large Magellanic Cloud takes the value  $(2.2)^{62.00}$  m, while the distances of the Large and Small Magellanic Clouds – a *pair* of interesting objects (dwarf galaxies in orbit around the Milky Way) – are symmetrically arranged about the value  $(3.3)^{41.01}$  m. The two objects had been judged to form a partnership.

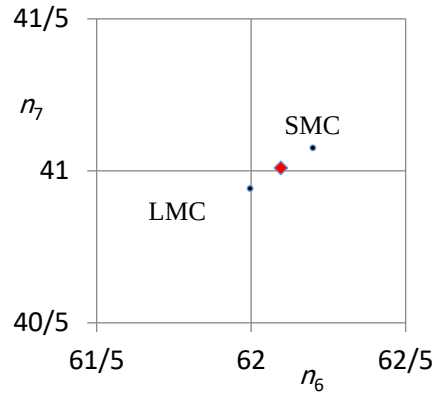
That the Bohr radius (Figure 16(f)) takes the values  $(\pi/2)^{125.00} l_{Planck}$  and  $(2.2)^{-30.01}$  m, while the meter is of value  $\pi^{69.98} l_{Planck}$ , tells one that,

$$(\pi/2)^{125} \times (2.2)^{30} \approx \pi^{70} \quad (29)$$

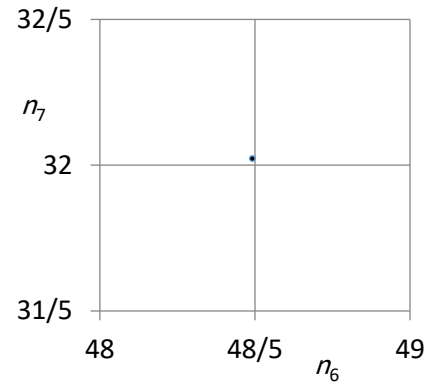
The exponent (-19.82) of the base 3.3 is also close in value to a very interesting number, i.e. -20, so that, although less precisely,

$$(\pi/2)^{125} \times (3.3)^{20} \approx \pi^{70} \quad (30)$$

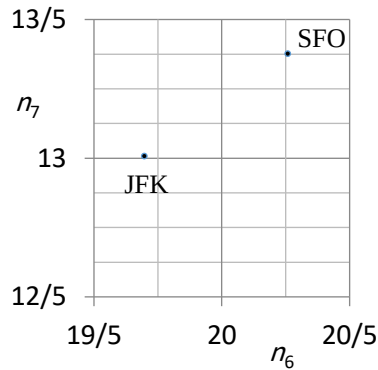
Equations (29) and (30), each of which combines three very interesting quantities, alludes to a considerable amount of iterative calculation yet the bases 2.2 and 3.3 were chosen on a whim. How the equalities in (29) and (30) might have come about will be discussed in 3.



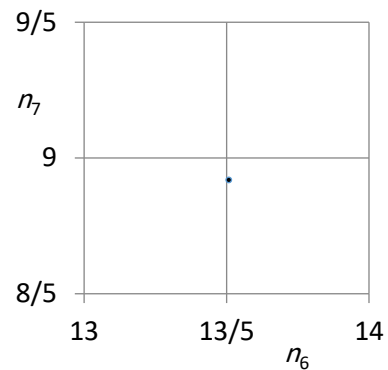
**(16a)** Distances from earth of the Large Magellanic Cloud (179, 000 lyr) and the Small Magellanic Cloud (210,000 lyr); (NASA, 2020). A diamond marks the point of mean  $n_6$  and  $n_7$ .



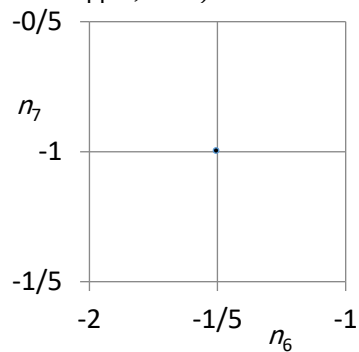
**(16b)** Distance from earth of Proxima Centauri (4.25 lyr)



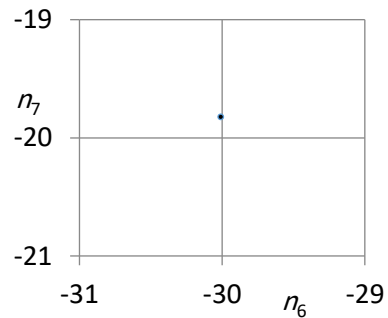
**(16c)** Distances from London LHR of New York JFK (5555 km) and San Francisco SFO (8638 km) (Great Circle Mapper, 2024)



**(16d)** The marathon distance (42.195 km)



**(16e)** The foot (0.3048 m)

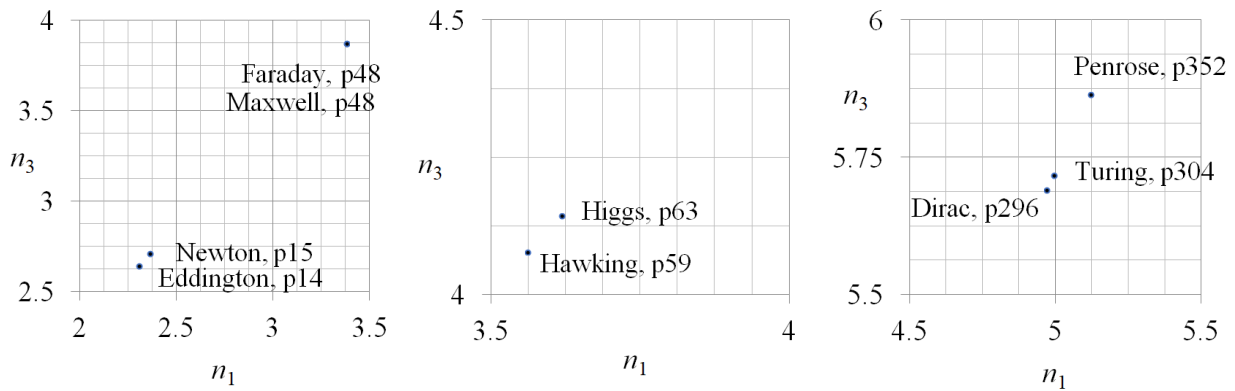


**(16f)** The Bohr radius ( $5.292 \times 10^{-11}$  m)

**Figure 16.** Points  $(n_6, n_7)$  representing assorted distances in meters. Each distance had first been expressed as  $(2.2)^{n_6}$  m and  $(3.3)^{n_7}$  m.

### 1.7. Dimensionless Numbers

A book – Brian Greene’s *The Hidden Reality* (Greene, 2011) – has been analysed to determine whether there are signs of the observer’s influence on what is observed of that publication. Here, the page numbers of the first references to British scientists whose names drew the attention of the observer are shown as powers of  $\pi$  and  $e$  in Figure 17. The exponents,  $n_1$  and  $n_3$ , of  $\pi$  and  $e$  are found to align closely with low-denomination rational numbers despite the constraints imposed by a book. At first, this result, although expected on the basis of experience, was difficult to comprehend. However, there is a Kantian explanation. What the observer sees of a book is part of the phenomenal world, constructed in the unity of human sensibility and understanding, and not reality as it is in itself: the noumenal world.



**Figure 17.** Points  $(n_1, n_3)$  representing the page numbers of the first references in the index of Brian Greene’s *The Hidden Reality* (Greene, 2011) to British scientists whose names drew the attention of the observer. Each page number has been expressed as  $\pi^{-n_1}$  and  $e^{-n_3}$ .

## 2. Physical Laws

The precise equality,

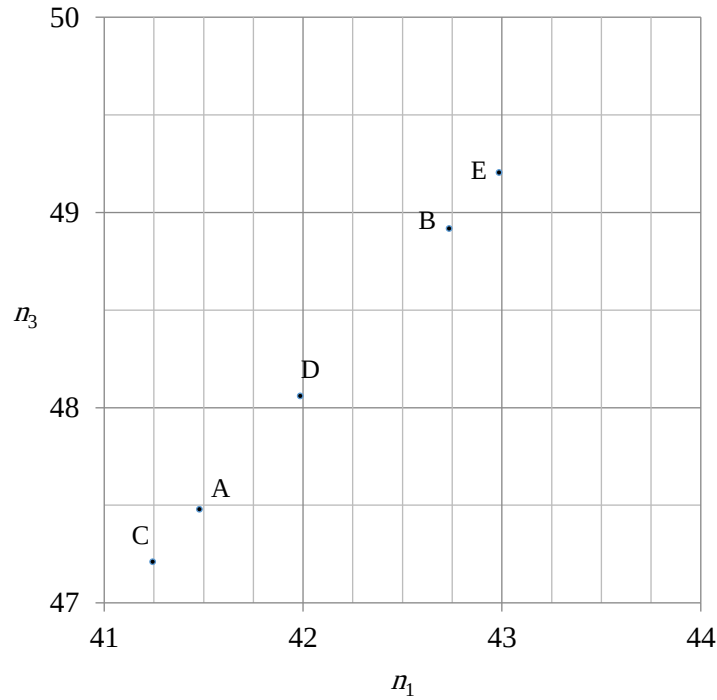
$$2a_0^5 = R_{OU}^2 \quad (31)$$

where  $a_0$  is the numerical value of the Bohr radius measured in Planck units and  $R_{OU}$  is the numerical value of the radius of the observable universe (14.3 Gpc) measured in Planck units was found by calculating the dark energy density in two different ways and then relating the two values (Riley, 2015). The equation can be balanced dimensionally with a denominate constant ( $l_{\text{Planck}}^3$ ). Equations of this form have subsequently been used to relate the numerical values of disparate physical variables, as follows.

The following equation of the form of (31) has been used to derive specific subatomic mass scales,  $m_{sa}$ , from the distances of stars from earth,  $d_*$  (Riley, 2020).

$$2m_{sa}^{-5} = d_*^2 \quad (32)$$

The negative exponent of  $m_{sa}$  arises because subatomic mass scales are smaller than Planck scale while all length scales are larger than Planck scale. The derived mass scales are shown as rational powers of  $\pi$  and  $e$  in Figure 18. Interestingly, the celebrated supergiant Betelgeuse occupies the most precise integer level intersection known,  $(n_1, n_2) = (43, 109)$ , which in the early years of this investigation – at the time, the light quark masses had not been measured with great precision – had been associated with the up and down quark masses (Riley, 2008).



**Figure 18.** Points  $(n_1, n_3)$  representing subatomic mass values,  $m_{sa}$ , in Planck units, derived from the distances from earth,  $d_*$ , measured in Planck units, of the four brightest stars in the night sky and the supergiant Betelgeuse according to the equation  $m_{d_*} = 2^{1/5} d_*^{-2/5}$ . Each derived mass value has been expressed as  $\pi^{-n_1} m_{\text{Planck}}$  and  $e^{-n_3} m_{\text{Planck}}$ . The stellar distances used:

- A Sirius, the brightest star: 2.636 pc
- B Canopus, the second brightest star: 95.95 pc
- C Alpha Centauri, the third brightest star: 1.35 pc
- D Arcturus, the fourth brightest star: 11.25 pc
- E Betelgeuse: 196.9 pc

The following equation, based on (32), relates the masses of atoms, measured in Planck units, and the radii of nearby stars of similar size to the sun, measured in Planck units.

$$2m_{atom}^{-5} = R_*^2 \quad (33)$$

The radii ( $1.223 R_{sun}$  and  $0.863 R_{sun}$ ) of Alpha Centauri A and B, stars that form a binary star system, correspond to the masses of atoms of mass number 48 and 56, respectively, as shown in Figures 19 and 20. The two derived mass values are symmetrically arranged about the level  $n_1 = 35$ , at the intersection  $(n_1, n_3) = (35, 40)$ . The level  $n_1 = 35$  is the domain of  $^{52}\text{Cr}$ , the principal isotope of naturally occurring chromium.

The radius of the sun ( $6.957 \times 10^5$  km) and that of Tau Ceti ( $0.793 R_{sun}$ ), which has been called the sun's twin, correspond to the masses of atoms of mass number 53 and 59, respectively, as shown in Figures 21 and 22. The two derived mass values are symmetrically arranged about the level  $n_3 = 40$ , at the intersection  $(35, 40)$ . The level  $n_3 = 40$  is the domain of  $^{56}\text{Fe}$ , the principal isotope of naturally occurring iron.

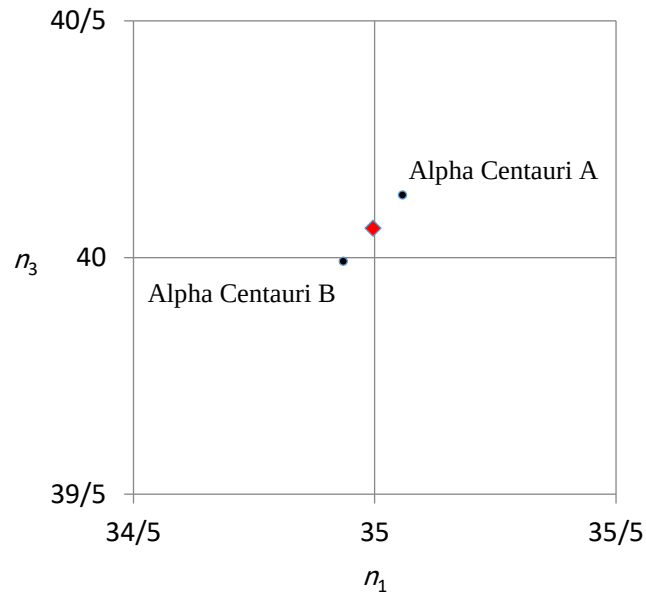
An equation was next sought that would interrelate the atomic radii of atoms (Slater, 1964) and the masses of stars, symmetry with (33) being a prerequisite. In the end, the following equation was constructed; a reason had been found for the inclusion of the factor  $(\pi/2)^{250}$  (Riley, 2018a).

$$2r_{atom}^5 = (\pi/2)^{250} M_*^2 \quad (34)$$

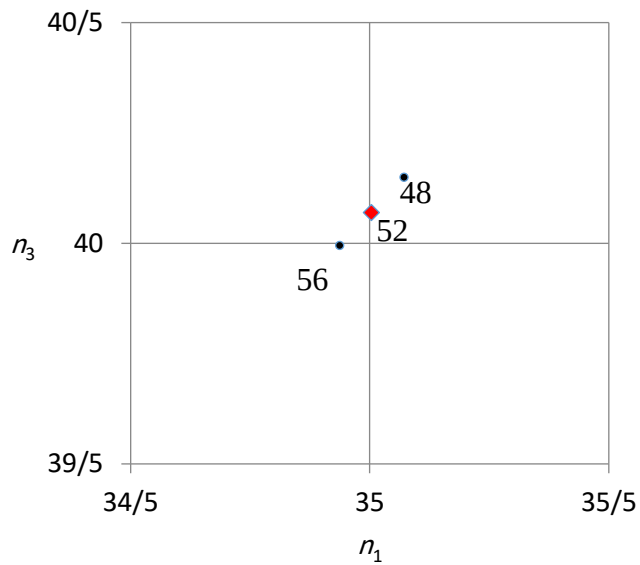
While the radius of Alpha Centauri A corresponds through (33) to the mass of an atom of mass number  $A = 48$ , e.g. the stable isotope  $^{48}\text{Ti}$ , the mass ( $1.100 M_{sun}$ ) of Alpha Centauri A has been found to correspond through (34) to an atomic radius of 143 pm; the empirical value for titanium is  $140 \pm 5$  pm. While the radius of Alpha Centauri B corresponds through (33) to the mass of an atom of mass number  $A = 56$ , e.g. the stable isotope  $^{56}\text{Fe}$ , the mass ( $0.907 M_{sun}$ ) of Alpha Centauri B has been found to correspond through (34) to an atomic radius of 132 pm; the empirical value for iron is  $140 \pm 5$  pm.

While the radius of the sun corresponds through (33) to the mass of an atom of mass number  $A = 53$ , e.g. the stable isotope  $^{53}\text{Cr}$ , the mass of the sun ( $1.988 \times 10^{30}$  kg) has been found to correspond through (34) to an atomic radius of 138 pm; the empirical value for chromium is  $140 \pm 5$  pm. While the radius of Tau Ceti corresponds through (33) to the mass of an atom of mass number  $A = 59$ , e.g. the stable isotope  $^{59}\text{Co}$ , the mass ( $0.783 M_{sun}$ ) of Tau Ceti has been found to correspond through (34) to an atomic radius of 125 pm; the empirical value for cobalt is  $135 \pm 5$  pm.

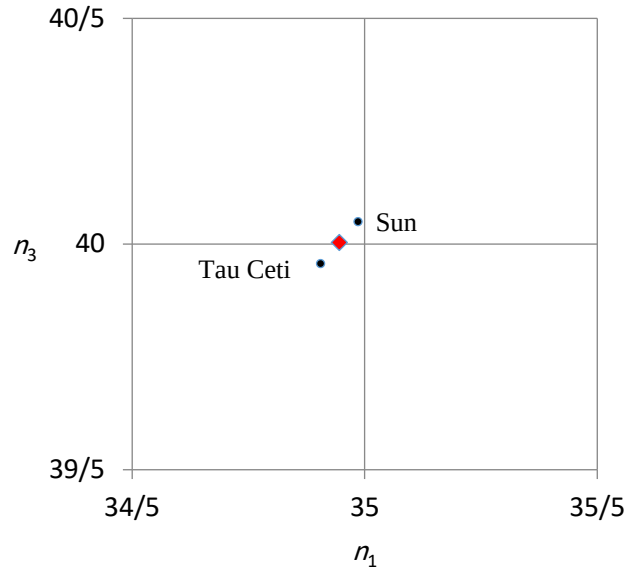
In summary, the radius in Planck units of the sun or of a nearby star of similar size to the sun corresponds through (33) to the atomic mass in Planck units of a specific period 4 transition metal, and the mass of the star corresponds, through (34), to the atomic radius typical of a period 4 transition metal.



**Figure 19.** Points  $(n_1, n_3)$  representing mass values,  $m_R$  in Planck units, derived from the radii,  $R$  measured in Planck units, of Alpha Centauri A ( $1.223 R_{\text{sun}}$ ) and B ( $0.863 R_{\text{sun}}$ ) according to the equation  $m_R = 2^{1/5} R^{-2/5}$ . Each mass value has subsequently been expressed as  $\pi^{-n_1} m_{\text{Planck}}$  and  $e^{-n_3} m_{\text{Planck}}$ . A diamond marks the point of mean  $n_1$  and  $n_3$ .

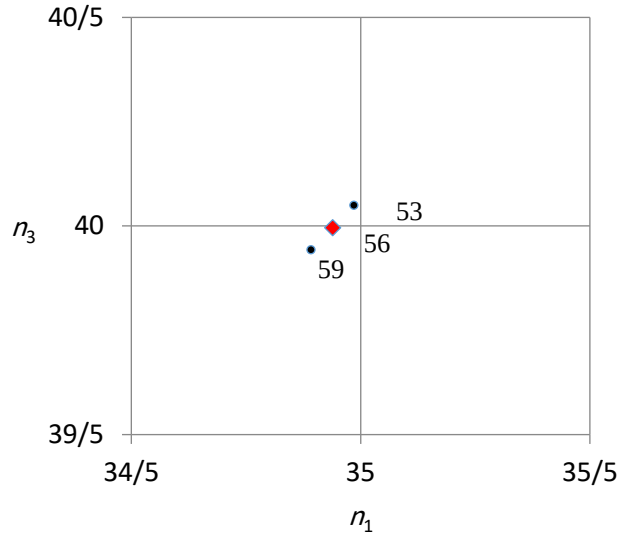


**Figure 20.** Points  $(n_1, n_3)$  representing the masses in Planck units of atoms of mass number  $A = 48$  and  $A = 56$ . A diamond marks the point corresponding to  $A = 52$ . Each mass value had first been expressed as  $\pi^{-n_1} m_{\text{Planck}}$  and  $e^{-n_3} m_{\text{Planck}}$ . For comparison with Figure 19.



**Figure 21.** Points  $(n_1, n_3)$  representing mass values,  $m_R$  in Planck units, derived from the radii,  $R$  measured in Planck units, of the Sun ( $6.957 \times 10^5$  km) and the solar analogue Tau Ceti ( $0.793 R_{\text{sun}}$ ) according to the equation  $m_R = 2^{1/5} R^{-2/5}$ . Each mass value has subsequently been expressed as  $\pi^{-n_1} m_{\text{Planck}}$  and  $e^{-n_3} m_{\text{Planck}}$ . A diamond marks the point of mean  $n_1$  and  $n_3$ .





**Figure 22.** Points  $(n_1, n_3)$  representing the masses in Planck units of atoms of mass number  $A = 53$  and  $A = 59$ . A diamond marks the point corresponding to  $A = 56$ . Each mass value had first been expressed as  $\pi^{-n_1} m_{\text{Planck}}$  and  $e^{-n_3} m_{\text{Planck}}$ . For comparison with Figure 21.

The mathematical formula of (33) relates the radius of a star and the mass of an atom. Although the range of the formula is restricted to a small number of values, were it not for the absence of any notion of causation the formula would represent a physical law; a mathematical formula relates physical variables in one set to physical variables of different dimensionality in another set. Perhaps, then, physical laws come about as the inventions of observers. If that is the case causality cannot arise from experience but is an a priori concept or category of the understanding. Equations that represent physical laws are often of elegant form because they would have been conceived in that form.

After the above results were found, the thought occurred that the atomic mass and radius of a period 4 transition metal ought to be interrelated in a recognizable way. The following equality was then found to relate the numerical values of the atomic mass,  $m_{53}$ , and radius,  $r_{53}$ , in Planck units, of  $^{53}\text{Cr}$ , the isotope that has been shown to correspond to the sun.

$$m_{53} = 2^{25.0}/r_{53} \quad (35)$$

Encouraged by this result, which applies to period 4 metals in general although not always as precisely, the thought occurred that the Bohr radius ought to be related to the mass of an interesting

object. The following equality of the form of (35) was then found, where  $m_{\text{Higgs}}$  is the mass of the Higgs boson and  $a_0$  is the Bohr radius (Riley, 2019).

$$m_{\text{Higgs}} = 2^{25.00}/a_0 \quad (36)$$

Since  $a_0 = (\pi/2)^{125.00} l_{\text{Planck}}$ , it follows that  $m_{\text{Higgs}} = 2^{25.00}(\pi/2)^{-125.00} m_{\text{Planck}}$ , which was shown as equation (6).

### 3. Discussion

Time and again, one's discoveries have been found to build on what has already been discovered. For example, after finding a highly precise and meaningful (to the cognizer) equation for the Bohr radius an equation for the mass of the electron was immediately comprehended from standard physics. This equation includes the fine structure constant,  $\alpha$ , as a factor. That factor was then found to be required in the equations for the up and top quark masses, and is present in other particle mass equations. The factor  $\alpha$  was found after being sought when a simpler expression for the quark masses could not be found that was consistent with reality (the phenomenal world). The quantity  $(\pi/2)^{25}$ , or a power of that quantity, originated with the Bohr radius and features in expressions found to relate the numerical values of 'very interesting' quantities in all sorts of contexts. After finding that the masses and radii of stable period 4 transition metals are related through a simple expression, a similar and precise expression was found to relate the mass of the Higgs boson and the Bohr radius. One thing has led to another as the knowledge of the observer has broadened. The numerical values of all sorts of quantities: quantities in any units, dimensionless quantities and differences in quantities (Riley, 2003) have all been found to map to rational numbers by the application of conceptual formulae of the observer's choice. If a particularly interesting quantity, of numerical value  $N$ , can be accommodated consistently within the phenomenal world of the observer so that the explanatory rational number  $n$  takes a conspicuous value, then it will be. Integers are particularly believable and inspire confidence in the explanatory concept; they signal to others the validity of the concept.

A mapping from the number of interests,  $N$ , to an explanatory rational number  $n$  that is an integer or even a fraction of low denomination suggests that much calculation has been performed but that has not been done, at least not within the phenomenal world. It seems the cognitive process described here is the human understanding of what has been done in reality as it is in itself: the noumenal world. The same explanation will then apply to the production of the complex equalities in equations (30) and (31) and to the invention of physical laws.

The value of a number is fixed when, in cognition, the observer has found it to be true. For the author, that is when a rational concept (a mathematical formula) explains the value of the number by the value of a rational number. Examples of such a cognitive process are readily to be found in science. If the number of interests is the numerical value of a physical quantity then, when cognized, its value will be logically consistent with the numerical values of all other physical quantities in the phenomenal world, at all times. If the quantity of interest is a physical constant, then it will always have been of the value found in cognition, which is consistent with what Kant said: that time is an a priori intuition through which the experiences of the observer are framed. That explains why the universe appears to be fine-tuned for life; the conditions for life are encapsulated in the constants of nature and those constants have been given their values by human beings.

## References

- Great Circle Mapper. (2024). <https://www.greatcirclemapper.net/en/great-circle-mapper.html>
- Greene, B. (2011). *The Hidden Reality*. Allen Lane.
- Kant, I. (1998) *The Cambridge edition of the works of Immanuel Kant*. Cambridge University Press.
- NASA. (2020). The nearest galaxies. *Imagine the Universe*. [https://imagine.gsfc.nasa.gov/features/cosmic/nearest\\_galaxy\\_info.html](https://imagine.gsfc.nasa.gov/features/cosmic/nearest_galaxy_info.html)
- NASA. (2024). Planetary fact sheet. *Lunar and Planetary Science*. <https://nssdc.gsfc.nasa.gov/planetary/factsheet/>
- NIST. (2024). CODATA internationally recommended 2022 values of the fundamental physical constants. *The NIST Reference on Constants, Units, and Uncertainty*. <https://physics.nist.gov/cuu/constants/index.html>
- Particle Data Group. (2024). The review of particle physics. *Physical Review D* 110, 030001 (2024). <https://pdg.lbl.gov>
- Planck Collaboration. (2020). Planck 2018 results. VI. Cosmological parameters. *Astronomy and Astrophysics*, 641, A6 (2020). <https://arxiv.org/1807.06209>
- Randall, L. & Sundrum, R. (1999a). A large mass hierarchy from a small extra dimension. *Physical Review Letters*, 83, 3370-3373. <https://arxiv.org/abs/hep-ph/9905221>
- Randall, L. & Sundrum, R. (1999b). An alternative to compactification. *Physical Review Letters*, 83, 4690-4693. <https://arxiv.org/abs/hep-th/9906064>
- Riley, B. F. (2008). Standard Model scales from warped extra dimensions. *Arxiv*. <https://arxiv.org/abs/0809.0111> [physics.gen-ph]
- Riley, B. F. (2013). The scales of atomic and particle physics in relation to Planck scale. *Vixra*. <https://vixra.org/abs/1305.0061> [high energy particle physics]
- Riley, B. F. (2015). The cosmological constant from a distant boundary. *Vixra*. <https://vixra.org/abs/1502.0017> [high energy particle physics]

- Riley, B. F. (2018a). The quantum/classical connection. *Vixra*. <https://vixra.org/abs/1809.0329> [quantum gravity and string theory]
- Riley, B. F. (2018b). The quantised angular momenta of astronomical bodies. *Vixra*. <https://vixra.org/abs/1811.0232> [quantum gravity and string theory]
- Riley, B. F. (2019). The mass of the Higgs boson. *Vixra*. <https://vixra.org/abs/1910.0090> [quantum gravity and string theory]
- Riley, B. F. (2020). The act of measurement I: astronomical distances. *Vixra*. <https://vixra.org/abs/2006.0247> [quantum physics]
- Riley, B. F. (2021). In the world of the observer. *Vixra*. <https://vixra.org/abs/2108.0012> [quantum physics]
- Riley, B. F. (2022). How the observer creates reality? *Vixra*. <https://vixra.org/abs/2205.0139> [quantum physics]
- Riley, B. F. (2003). A unified model of particle mass. *Arxiv*. <https://arxiv.org/abs/physics/0306098> [physics.gen-ph]
- Slater, J. C. (1964). Atomic radii in crystals. *Journal of Chemical Physics*, 41, 3199-3204.

## A Kantian Solution for the Freedom of Choice Loophole in Bell Experiments

Romeu Rossi Júnior<sup>✉ 1</sup>  | Patrícia Kauark-Leite<sup>2</sup> 

<sup>1</sup> Corresponding Author, Instituto de Ciências Exatas e Tecnológicas Universidade Federal de Viçosa Florestal, Brazil. Email: [Romeu.rossi@gmail.com](mailto:Romeu.rossi@gmail.com)

<sup>2</sup> Department of Philosophy Universidade Federal de Minas Gerais Belo Horizonte, Brazil. Email: [pkauark@gmail.com](mailto:pkauark@gmail.com)

### Article Info

#### Article type:

Research Article

#### Article history:

Received 20 May 2024

Received in revised form  
28 June 2024

Accepted 03 July 2024

Published online 07  
August 2024

#### Keywords:

Bell's inequality, Big Bell  
Test, free will, Kant,  
agency

### ABSTRACT

Bell's theorem is based on the assumptions of local causality and measurement independence. The last assumption is identified by many authors as linked to the freedom of choice hypothesis. In this sense the human free will ultimately can ensure the measurement independence assumption. The incomplete experimental conditions for supporting this assumption are known in the literature as "freedom-of-choice loophole" (FOCL). Although there is no consensus among the scientists that the measurement independence is linked to human choices, in a recent paper, published in a prestigious journal, signed by more than a hundred authors, this assumption was seriously taken for the first time in an experiment known as Big Bell Test (Abellán et al. 2018). Using photons, single atoms, atomic ensembles and superconducting devices, this experiment was performed in five continents, and involved twelve laboratories, adopting human choices to close the FOCL. Nevertheless, the possibility of human freedom of choice has been a matter of philosophical debate for more than 2000 years, and there is no consensus among philosophers on this topic. If human choice is not free, this solution would not be sufficient to close FOCL. Therefore, in order to support the basic assumption of this experiment, it is necessary to argue that human choice is indeed free. In this paper, we present a Kantian position on this topic and defend the view that this philosophical position is the best way to ensure that Big Bell Test can in fact close the loophole.

**Cite this article:** Rossi Júnior, R. & Kauark-Leite, P. (2024). A Kantian Solution for the Freedom of Choice Loophole in Bell Experiments. *Journal of Philosophical Investigations*, 18(47), 189-202. <https://doi.org/10.22034/JPIUT.2024.62277.3802>



© The Author(s).

Publisher: University of Tabriz.

<https://doi.org/10.22034/JPIUT.2024.62277.3802>

## Introduction

Performing a Bell test is not a simple task. Bell's theorem (Bell, 1964) is intended to show that local causality theories are not adequate for describing quantum phenomena. To reach an experimental result that allows such broad conclusions, it is necessary that the experiments rigorously fulfill two technical conditions: (i) speed and efficiency in the detection of measures and (ii) the permanence of entangled states that maintain their coherence even when separated at great distances. In addition, it is also necessary to presuppose a third condition, which is the unpredictable measurement settings. This third condition requires that the measurements settings imply random choices that should be statistically independent of any influence of hidden variables (Abellán et al 2018). The first experimental implementations (Aspect, Grangier and Roger 1982; Aspect, Dalibard, and Roger 1982; Weihs et al 1998; Ou and Mandel 1988; Shih and Alley 1988; Tapster et al 1994) show agreement with quantum theory, but were not able to ensure the necessary preconditions for the tests. In these cases, the Bell test goal has not been reached, since the same experimental results can be explained by classical theories using local hidden variables. These circumstances of incomplete experimental conditions are referred in the literature as "loopholes." There are different kinds of loopholes, most of them are related to technical issues. The detection efficiency loophole and the locality loophole were recently overcome in experimental settings (Hensen et al 2015; Giustina et al 2015; Shalm et al 2015).

The freedom-of-choice loophole (FOCL) is related to the assumption that the measurement setting variables ( $\mathbf{x}$  and  $\mathbf{y}$ ) are statistically independent of the hidden variables. Usually, in Bell tests, the measurement settings are determined by a device, a physical randomizer (Hensen et al 2015; Rosenfeld et al 2017; Abellán et al 2015; Fürst et al 2010; Scheidla et al 2010) that would be responsible for the required statistical independence. However, in classical terms, we can say that the randomizers have a causal past because they are physical devices, and consequently the setting variables  $\mathbf{x}$  and  $\mathbf{y}$  have a common causal past. In other words, we can say that the backward light cone of the setting variables  $\mathbf{x}$  and  $\mathbf{y}$  overlaps. The fact that the measurement setting variables  $\mathbf{x}$  and  $\mathbf{y}$  have a causal past in common, entitles one to assume that a hidden variable can be a common cause for them. To ignore this possibility would be to assume that "Bell's theorem applies only to a hybrid universe in which hidden variables determine only part of the outcomes of experiments." (Brans, C. 1988). This means that hidden variables would determine the measurable outcomes ( $\mathbf{A}$  and  $\mathbf{B}$ ), but not the setting variables  $\mathbf{x}$  and  $\mathbf{y}$ .

Efforts to close FOCL have been made by different experimental groups. Scheidla et al. 2010 and Shalm et al, 2015 claim that they have fulfilled this task by ensuring that the measurement setting choices are separated in a spacelike way from the event of the creation of the particles. They made the strong assumption that the hidden variable ( $\lambda$ ) is created together with the entangled pair of particles. Therefore, the spacelike separation between the particle pair creation event and the

measurement setting choices of the variables  $x$  and  $y$  would ensure statistical independence between  $x$  and  $y$ , on one side, and  $\lambda$ , on the other.

There are also attempts to tighten FOCL by constraining the space-time volume. In these attempts the possible causal relationship between the measurement setting variables ( $x$  and  $y$ ) and the hidden variable ( $\lambda$ ) could have occurred (Rauch et al 2018; Handsteiner et al 2017) only in a very remote past. This is done through experimental setups that allow the measurement setting choices to be made through degrees of freedom of photons emitted by distant stars. In this way, the experiment “pushes the origin of the measurement settings considerably deeper into cosmic history” (Rauch & et al 2018). According to Rauch et al 2018, only 4% of the space-time volume is left available for establishing the causal relationship between the settings variables  $x$  and  $y$ , and  $\lambda$ .

However, Abellán et al 2018 hold that even the strong restrictions established by Dominik Rauch et al. 2018 would not be sufficient to close FOCL, since in all these attempts, the measurement settings continue to be defined by physical objects, as in the case of photons emitted by distant stars. In their words, “while still requiring a physical assumption, and thus not closing the FOCL, this strategy tightens the loophole in various ways” (Abellánv & et al 2018). The assumption taken by Abellán et al 2018 is that the measurement settings can be considered legitimately independent of  $\lambda$  only if they are the result of human choices. For only in this situation is the link between the measurable variables and  $\lambda$  excluded. This idea goes back to the famous remark made by Bell in the last section of his paper “The theory of local beables” (Bell 2004, 61):

It has been assumed (in deriving Bell’s theorem) that the settings of instruments are in some sense free variables - say at the whim of experimenters - or in any case not determined in the overlap of the backward light cones.

Bell seems to have backed off from his early claim that measurement independence is ensured by human free will in another paper entitled “Free variables and local causality” (Bell 2004, chap. 12). In response to Shimony, Horne and Clauser 1976, when he was accused of having used an overly metaphysical hypothesis to substantiate his theorem, Bell changed his early position holding that the hypothesis of free will is not necessary and that a randomiser would be sufficient to guarantee the measurement independence condition (Cf. Bell 2004, 103. In view of this debate, some authors like San Pedro 2014 and Esfeld 2015 object that the premise of measurement independence has nothing to do with the issue of whether or not humans have free will and so Bell’s theorem is not concerned with human free will, but only with the independence among the set of variables in question. Although they present good arguments for that, it is not our aim here to assess who is right in this debate. Our starting point is that despite this objection there is a good part of the scientific community, as it is the case of the authors of Big Bell Test (Abellán & et al 2018), involving more than a hundred of scientists, that still maintain that only free human choices

would be able to ensure the statistical independence of the measurement settings, and suitable to close the FOCL. So, the “freedom of choice loophole” is far from being a solved issue.

By accepting the free will choice as the only way to close the FOCL, the authors of Big Bell test implicitly introduce into the process a different kind of causality, aka causality through freedom, that cannot be assimilated to the natural causal chain (natural causality). In philosophical terms, they tacitly admit that physical events can be caused either by natural causes or by human freedom. In the first case, the effect has a cause that was caused by another cause. In the second case, the effect has a cause that was not caused by any empirical cause and, in this sense, one can say that this cause is free. Thus, according to this view, the variable  $\lambda$  could be considered as a possible cause of any physical event, e.g., processes in random number generators, emission of photons in stars, etc., but not a cause of human choices. This philosophical thesis was admitted by Bell and by the authors of Big Bell test to be sufficiently close to the FOCL, and so to ensure the conditions of validity of Bell’s theorem. At this point, it becomes clear that the initial expectation of transforming the philosophical debate between Einstein and Bohr on the incompleteness of quantum theory into a purely experimental investigation was not accomplished by Bell.

The philosophical debate on free will can be traced back to Augustine 2011, for whom the possibility of human freedom in presence of the foreknowledge of God was a central issue. In the XVIII century, with the advent of Newtonian physics, this debate acquired new contours. The free will thesis seemed to be opposed to the determinism of the laws of physics. In this sense, human freedom would not be possible if every event is determined by the previous state of things according to strict laws of nature. The philosophers who hold that universal natural determinism and freedom are mutually metaphysically inconsistent are classified as incompatibilists, whereas those who deny this possibility and assert the consistency of universal natural determinism and freedom are known as compatibilists.

However, the simple negation of determinism is not enough to hold a coherent position in favor of the free will assumption. If every event were to result from a random process, then human choices would have to result from a random process too, and thus they would have no agential causal source. If human choices have no agential causal source, they cannot be effect of human free will. Thus, in a universally indeterministic world, decisions and choices would not be caused by human free will because they would be random events and not have an agential source. Therefore, indeterminism is also a threat to human free will. Nevertheless, the case of “superindeterminism,” i.e., universal natural indeterminism, which deny both the determinism and the causality, is excluded in the context of Bell experiments, which try to show precisely the inadequacy of causal theories applied to quantum phenomena, whether they are deterministic or probabilistic.

Conversely, if human choices were to be completely determined by natural causality, then free will would be an illusion. However, it is not necessary to assume that this natural causality is



deterministic. We can postulate a kind of “compatibilistic incompatibilism” between natural probabilistic causality (not deterministic) and free will. The possibility of human choice being determined by a natural causality (not necessarily deterministic) is often identified as a “superdeterministic” position (Abellán & et al 2018). Nevertheless, Bell and the authors of Big Bell test hold that any kind of superdeterministic position cannot be experimentally tested. In their words: “the theory that the entire experiment, including choices and outcomes, is pre-determined by initial conditions is known as superdeterminism and superdeterminism cannot be tested” (Abellán et al 2018). Thus, they assume in the experiment, although not explicitly, a philosophical position according to which freedom and nature belongs to two different domains. The first one is essentially human and agential, and the second one essentially physical and non-agential. It is only in the domain of essentially physical events that the concept of natural causality can be applied. In the domain of human action, the choices are free.

In this paper, we show that the tacit philosophical commitment of Big Bell experiment is better understood in Kantian terms. We follow here the authors (as for instance Bitbol 2010, Bitbol et al. 2009, Bitbol et Osnaghi 2013, Falkenburg forthcoming, Kauark-Leite 2012, 2017) to whom Kantian philosophy is a key to understand philosophical problems in quantum domain. In this work, in particular, we therefore make explicit the Kantian philosophical basis of the central hypothesis of the experiment carried out by Abellán et al 2018, according to which human choices are free. For that, we will take into account a certain interpretation of Kant’s theory of human agency defended among others by Henry Allison 1990 and Maria Borges 2019. It is not the purpose of this work to go into detail concerning the discussions between Kantian scholars on the problem of the free will. Our aim is just to show that Big Bell experiment’s account is consistent to the Kantian interpretation that attributes to the agent the power to start a causal series spontaneously, free from the influences of any previous events. In Kant’s philosophy, natural causality allows us to understand the nature as subject to laws which hold universally in nature and give it a predictable character. In this sense, every event is preceded by a cause which generates a causal chain in accordance with a rule. However, that is not the only kind of causality that is operating in the empirical world. We need to presuppose a causality through freedom according to which the power of choice of the human agent has empirical causal efficacy and can initiate a series as a spontaneous cause that is uncaused by any earlier event. Therefore, the idea that human choices are free and have causal efficiency with respect to natural objects of nature is built into the central hypothesis of Abellán et al 2018 and it also finds support in Allison’s interpretation of Kant’s theory of human free agency.

## **1. The Theorem**

In a traditional Bell experiment a pair of entangled particles are spatially separated and measured by two observers, conventionally called Alice and Bob. The variables  $x$  and  $y$  are the setting

variables that describe the possible measurement settings that can be performed by them respectively. The variables **a** and **b** represent the measurement outcomes of the observables associated to **x** and **y**.

Bell's theorem imposes restrictions on the probability  $\mathbf{P}(\mathbf{a}, \mathbf{b}|\mathbf{x}, \mathbf{y})$  through assumptions about the causal structure that represents the physical system in a Bell experiment. The conditions are (i) local causality, and (ii) measurement independence. "Local causality" refers to the assumption that the value of the variable **a** (**b**) is the joint effect of a hidden variable  $\lambda$  and the setting variable **x** (**y**).

The experiments performed by two observers – Alice and Bob – are space-like separated events, therefore **a** and **b** are statistically independent given  $\lambda$ , **x** and **y**. Also, **a** (**b**) is statistically independent of **y**(**x**). Or in mathematical terms:  $\mathbf{P}(\mathbf{a}, \mathbf{b}|\mathbf{x}, \mathbf{y}, \lambda) = \mathbf{P}(\mathbf{a}|\mathbf{x}, \lambda) \mathbf{P}(\mathbf{b}|\mathbf{y}, \lambda)$ .

The measurement independence condition states that the setting variables are independent of  $\lambda$ . It can be written as  $\mathbf{P}(\mathbf{x}, \mathbf{y}|\lambda) = \mathbf{P}(\mathbf{x}, \mathbf{y})$  which is equivalent to  $\mathbf{P}(\lambda|\mathbf{x}, \mathbf{y}) = \mathbf{P}(\lambda)$ .

Using the assumptions of local causality and measurement independence one can write the conditional probability  $\mathbf{P}(\mathbf{a}, \mathbf{b}|\mathbf{x}, \mathbf{y})$  as:

$$\mathbf{P}(\mathbf{a}, \mathbf{b}|\mathbf{x}, \mathbf{y}) = \int d\lambda \mathbf{P}(\mathbf{a}, \mathbf{b}|\mathbf{x}, \mathbf{y}, \lambda) \mathbf{P}(\lambda|\mathbf{x}, \mathbf{y}) = \int d\lambda \mathbf{P}(\mathbf{a}|\mathbf{x}, \lambda) \mathbf{P}(\mathbf{b}|\mathbf{y}, \lambda) \mathbf{P}(\lambda) \quad (1)$$

If  $\mathbf{P}(\mathbf{a}|\mathbf{x}, \lambda)$  and  $\mathbf{P}(\mathbf{b}|\mathbf{y}, \lambda)$  can assume only the values 0 or 1, the hidden variable theory is deterministic. In this case, the value of the variable **a**(**b**) can be precisely known when **x**(**y**),  $\lambda$  are given. However, this is not a necessary condition for Bell's theorem. One may consider a non-deterministic causal hidden variable theory which allows that  $\mathbf{P}(\mathbf{a}|\mathbf{x}, \lambda)$  and  $\mathbf{P}(\mathbf{b}|\mathbf{y}, \lambda)$  assume intermediate values. The assumptions will give the same result shown in equation (1). According to Bell's theorem, therefore, deterministic and non-deterministic hidden variables theories (HVT) are equivalent. Since deterministic HVT are more restrictive, in this essay we will focus on random HVT, and consequently, the causal relations among variables are also in the domain of probabilistic natural causality.

The predictions of quantum mechanics are not compatible with the separable form of  $\mathbf{P}(\mathbf{a}, \mathbf{b}|\mathbf{x}, \mathbf{y})$  in Equation (1). This incompatibility can be experimentally tested. However, the assumptions of locality and measurement independence must be guaranteed in the experiment. It has been shown (Hall 2011; Degorre et al 2005; Hall 2010; Brans 1988) that if one of the assumptions is relaxed, it is possible to reproduce statistics of quantum states within a hidden variable model.

On the one hand, in order to ensure locality, one must deal with technical issues in the experimental setup. The measurements **a** and the setting variable **x** have to be space-like separated from **b**. On the other hand, there is no agreement among physicists on the necessary conditions for ensuring the measurement independence condition (MIC). The problems involved are beyond the technical issues. Additional properties of the hidden variables are considered in experimental setups that aim to ensure MIC. As mentioned above, this situation of incomplete experimental

conditions related to measurement independence is known as the “freedom-of-choice loophole” (FOCL).

The equivalence between the conditional probabilities  $P(x, y|\lambda) = P(x, y)$  and  $P(\lambda|x, y) = P(\lambda)$  given in the MIC points to two different causal relations among  $\lambda$ ,  $x$  and  $y$ . In the first,  $\lambda$  cannot be the cause of the setting variables  $x$  and  $y$ . In the second,  $x$  and  $y$  cannot be the cause of  $\lambda$ . In this essay, we follow Abellán et al. 2018 and hold that FOCL is related only to the first relation. The second applies to the locality loophole.

Even though the three assumptions are all necessary for a Bell test, the first experimental implementations were performed without any concern for MIC. According to Aspect, Dalibard, and Roger 1982, the setting variables  $x$  and  $y$  were chosen into the backward light cone of the entangled pair creation event. This allows for the possibility of causal relations among the setting variables and  $\lambda$ . For Weihs et al. 1998, the setting variables were chosen in the future light cone of the emission event. In this case,  $x$  and  $y$  could have been caused by  $\lambda$ . In both situations, it was assumed that  $\lambda$  must have been created at least at the same time as the entangled pair. Thus, in these earlier experiments the FOCL was left open.

In recent Bell’s inequality experiments, the authors claim that FOCL was closed. Their presuppositions can be roughly divided into two groups. In the first, Scheidl et al. 2010 and Shalm et al. 2015 presuppose that the hidden variables are generated together with the entangled particles. Thus, since the spacelike separation between the particle’s generation event and the choice of the measurement settings is ensured, one can as a consequence assume that  $\lambda$  is not a cause of  $x$  and  $y$ . In the second, Hensen et al. 2015; 2016 presuppose that the measurement setting variables  $x$  and  $y$  have no causal past.

In the first group, the assumption that  $\lambda$  is generated together with the entangled particles is a strong and additional assumption about hidden variables that is not present in the derivation of Bell’s inequality. The problem is that this additional hypothesis imposes a kind of restriction that makes these experiments not instances of violations of Bell inequalities in a general way, as they are supposed to be, but only in a very specific way, that is, in the case in which  $\lambda$  is generated only in a particular moment. The problem with the second kind of experiments is that the choices of settings  $x$  and  $y$  are done by a physical device (a randomizer). Therefore, as a physical device, it is submitted to the causality principle. So, this central assumption in the second kind of experiments leaves the FOCL open.

Abellán et al 2018 argue that it is not possible to close FOCL “while still requiring a physical assumption” about the choice of  $x$  and  $y$ . Following the famous argument by Bell in (Bell, 2004), they hold that FOCL can be closed only if measurement setting variables are defined by human choices. Based on this argument, they performed the Big Bell Test in five continents, and twelve laboratories, using photons, single atoms, atomic ensembles and superconducting devices. A significant part of the scientific community specializing in the field of nonlocality was involved in

this test. For the first time, human choices (from 100,000 volunteers) were used to select measurement settings in a Bell test. The magnitude of this collaboration (more than a hundred authors signed a single paper) shows that the scientific community is indeed concerned about the impossibility of closing FOCL using randomizers (physical devices) or any other kind of physical object.

An important question arises: why are human choices preferable to any kind of physical process for close FOCL? Although this question is not explicitly answered in (Abellán et al 2018), the authors explicitly say that events “requiring a physical assumption” cannot close FOCL. The Bigbell Test requires thus two strong assumptions. First, since every physical event has a causal past, one cannot ensure that  $\lambda$  does not belong to it. And second, since human choices are not an empirical causal process, and therefore  $\lambda$  cannot be a cause of them, they are the only alternative for closing FOCL.

## 2. Empirical Causality vs. Human Causality Through Freedom

In this section, we present a philosophical argument based in Kant’s Critical philosophy in order to support the implicit assumption in (Abellán et al 2018), according to which human choices are independent of  $\lambda$ . For this purpose, it is necessary to distinguish between the causality that regulates physical processes (which we will call merely physical causality) and the causality associated with human decisions (which we will call human causality through freedom). The first presupposes that the connection between cause and effect is constituted as a change in which “the apprehension of one thing (that which happens) follows that of the other (which precedes) in accordance with a rule” (Kant, KrV, A193/B238). Cause and effect are both merely physical events and consequently we can think of  $\lambda$  as a merely physical cause of a later effect. The second presupposes that the cause of a physical effect has its origin not in a merely physical event but in the human *arbitrium*, i.e., the human will, or “power of choice” (*Willkür*). In this case, since  $\lambda$  is a physical variable, it is not a part of the causal chain triggered by the human *arbitrium*. This is because the cause of a physical effect in such a case is not determined by any physical variable but by a human action.

In his famous Critical program, these two kinds of causality emerge from the basic distinction adopted by Kant between nature and freedom. In Kant’s theory, the term “nature” refers to the set of spatiotemporal objects of experience that are subject to rules given *a priori*. Physical causality is one of these rules. Contrary to common sense, causality in Kantian terms is not a condition of a mind independent world, but instead a condition imposed by our faculty of understanding. Therefore, physical causality is one of the rules we use to represent appearances, and not a property of a mind independent world. Experience is not an event independent of the subject but instead constituted by the subject. The revolution brought about by this new approach in the field of metaphysics was of course compared by Kant, in the Preface to the second edition or B of his *Critique of Pure Reason*, to the Copernican revolution in the natural sciences. But whereas

Copernicus moved the Earth from the center of the universe to the periphery of our solar system, circling the Sun, Kant's theory moves the subject from the periphery of knowledge to the epistemic center, a place previously occupied by the object.

The second kind of causality, that which is not empirical, does not belong to nature but instead is grounded in human freedom. Contrary to the realm of nature, freedom is not conditioned by spatiotemporal objects of experience and therefore it is not restricted to the laws that strictly govern appearances. Causality through freedom, as spontaneous, is able to start a causal series without being itself caused by previous physical events. Kant establishes a cognitive boundary that limits the merely physical causality to the empirical objects which are mechanically conceived. This allows the creation of a conceptual space unrestricted by such causality. Because this conceptual space does not belong to the realm of appearances, or phenomena, Kant needs to postulate another realm, which he characterizes as "noumenal." So, causality through freedom belongs not to appearances but instead to the conceptual space of noumena. Nevertheless, causality through freedom can start a causal series in a physical space. Therefore, according to Kant, the distinction between nature as phenomenal and freedom as noumenal is necessary for guaranteeing the possibility of the second kind of causality without also denying the existence of physical causality.

When applied to the rational human agent, the distinction between nature and freedom seems to result in an inconsistency. On the one hand, the rational human agent belongs to the set of appearances because, as a human animal, it is a spatiotemporal object of experience. Therefore, it is subject to empirical causality. On the other hand, the rational human agent is taken to have the ability to initiate a causal series spontaneously, without being determined by empirical causality. To overcome this apparent inconsistency, Kant presents a detailed theory about the dual "character" of rational human agency. The rational human agent has an "empirical character," which is subject to natural causality, and also an "intelligible character," which is able to spontaneously choose the course of its actions.

Thus, dual character theory applies specifically to the rational human agent's will. The empirical character of the will supports a psychological causal explanation for the agent's performance, according to which the agent's choices and actions are, to a certain degree, predictable. The agent's desires, inclinations, and beliefs are psychological factors that allow a certain degree of predictability; hence one can take these factors to be as the empirical causes of choices and actions. This empirical notion of causality underlying the relation between psychological factors and human behaviors can be thought as being not of the same kind as the physical causality but as a kind of natural causality that has a causal history. From this perspective, one can conceive in terms of a natural causality this empirical character of the will. If human choices were all defined only by this empirical character of the will, we might think that ultimately human choices would be determined by a physical variable like  $\lambda$ . This is because psychological causality being empirical presupposes has a causal history. In this case one cannot ensure that a physical cause, such as  $\lambda$ , is independent

of psychological factors. If that were the case, then the experiment proposed by Abellán et al 2018 test based on empirical character of human choice would not guarantee the closing of the loophole (FOCL).

In Kantian terms, the causality associated with the empirical character is not sufficient to determine the will of the rational agent. According to Kant, the choice is a deliberative process of reasoning in which the association of ideas and concepts that guide the human action requires a degree of independence from empirical causes. This presupposes a spontaneity of reason for starting a causal series which, in turn, was not determined by any previous natural cause. In this sense, it is necessary to assume a freedom of the rational agent in determining his choices and actions triggered by these choices. Kant characterizes this free character of the human will, which is independent of empirical causality, as “intelligible.” By means of the intelligible character of the will, one can explain the choices and acts of a rational agent as effects of a different kind of cause that is not itself subject to the rules of empirical causality, whether merely physical or empirically psychological. Kant calls this “causality through freedom.”

Henry Allison uses what he calls the “Incorporation Thesis” to support the presence of the dual character of the will in Kant’s philosophy. According to Allison’s interpretation of Kant’s theory of agency, the choices and actions of a rational agent are yielded by the incorporation of motives, like desires, inclinations, moral principles, etc., as reasons. The motives are presented to the rational human agent, who deliberately, self-consciously, and spontaneously decides to incorporate them as reasons or choice and action. The incorporation thesis allows us to identify the human power of choice as “*arbitrium liberum*” and not as “*brutum*” (Kant, KrV, A535/B562). In the animal power of choice, based on *arbitrium brutum*, sensible stimuli are causes of actions. If the human power of choice were an *arbitrium brutum*, then the choices and actions of a rational human agent would be caused by empirical psychological factors. But, for Kant, this is not the case. In Allison’s words, “such a subject is, therefore, more properly characterized as a patient rather than an agent”. (Allison 1996, 130). To think of human rational agency as an *arbitrium liberum* means that inclinations and desires are not direct causes of actions. The rational human agent deliberately and spontaneously decides to incorporate motives, even though the agent is aware of desires and inclinations. Being aware of desires and inclinations does not mean that they have a causal function in the deliberative and decision-making process. The agent can freely incorporate or not incorporate a motive as a maxim or a rule and, and choose or act according to them. Allison understands the act of incorporation as a genuine causal factor for any rational human action (Allison 1990). Thus, the incorporation is an efficient cause of the choice or action for a rational human agent. Therefore, it must take place in a certain period of time which starts a causal series. But the incorporation itself is not the result of previous causes. It initiates by itself (as a spontaneous or uncaused cause) a new causal series. This type of causality, causality through freedom, can be thought only as an intelligible character of the human power of choice.

We must emphasize that the Incorporation Thesis is based on the absolute spontaneity of the *arbitrium*, without any empirical causal influence. Therefore, it requires an absolute kind of freedom, which does not presuppose any previous merely physical or empirically psychological cause, which Kant calls “transcendental freedom.” The incorporation of a motive as a rule for rational human choice and action starts a new causal series. This condition of absolute freedom of the agent is more restrictive than the condition required to the solution of FOCL in Bell experiments with human choices. As was shown by Hall 2010, the complete independence between the hidden variables  $\lambda$  and settings variables  $\mathbf{x}$  and  $\mathbf{y}$  are not required. The level of tolerance for the measurement independence condition is low (indeed, it was even calculated by Hall 2010) but it is not null. Therefore, when one accepts that the settings variables are defined by human choices, these choices would not necessarily have to be absolutely free. One could consider a degree (very low) of dependence between the hidden variables and human choices and still close the FOCL. In this sense, the Kantian position on the freedom of rational human agents is more restrictive than the condition required to close FOCL with human choices.

As regards merely physical causality, the Kantian position is also more restrictive than Bell’s position. Kant holds that merely physical causality is deterministic, while in Bell’s theorem the causality can be taken as probabilistic. Kant’s theory of rational human agency requires the compatibility of absolute noumenal freedom and merely physical phenomenal causal determinism. These Kantian conditions are stronger than those associated with the application of human choice to close FOCL. In the latter, what is required is compatibility of a quasi-absolute freedom (as demonstrated in the works of Hall 2011; Degorre et al 2005; Hall 2010) and probabilistic or indeterministic causality. Therefore, Kant’s theory of agency is better suited to support the free will assumption in (Abellán et al 2018), because it is able to resolve more serious inconsistencies.

The Kantian solution that allows the compatibility between absolute noumenal freedom and merely physical phenomenal causal determinism cannot be easily classified as either “compatibilist” or “incompatibilist” in the traditional sense. This is a matter of debate among interpreters (Hanna 2006; Wood 1984; Allison 1990). The peculiarity of Kantian position is noted by Robert Hanna as follows: “Kant’s libertarian theory of freedom of the will is philosophically significant precisely because it is neither hard determinist, nor soft determinist, nor causal indeterminist, nor hard indeterminist, nor compatibilist, nor incompatibilist.” (Hanna, 2006, 419). This because, on the one hand, the absolute noumenal freedom required by Kant’s solution is not compatible with deterministic merely physical phenomenal causality, and, on the other hand, the introduction of a conceptual space—i.e., noumenal space--free from empirical causality allows Kant to claim that causality through freedom is not in direct causal competition with empirical causality. In this sense, nature and freedom are not in direct opposition but can be thought as yielding causal series which are compatible but not in any way reducible or otherwise assimilable to each other. Therefore, this view is sometimes called Kant’s “incompatibilistic compatibilism.”

In conclusion, we argue that the alternative for closing the FOCL through the hypothesis of free will can find philosophical support in Kant's theory of rational human agency. By promoting an "incompatibilistically compatibilist" approach to the relation between freedom and nature, Kant's theory provides philosophical support for the central hypothesis of the experiment carried out in (Abellán et al. 2018). The requirement that the capacity of the rational human agent for choosing the directions of the measurement setup free from any causal influence, is fulfilled by Kant's theory of rational human agency. According to Allison's interpretation, the incorporation of a motive as a rule for choice and an action has efficient causal power even though it is not merely physically or empirically psychologically caused by previous events. In other words, it has no phenomenal causal history. It is a spontaneous cause that starts a new series of phenomenal events. This kind of causality through freedom is exactly what validates the use of human choices, as free choices, to decide the directions of the measures in the experiment. Therefore, his Kantian philosophical point of view enables us to conclude that rational human agency is the right candidate for closing the FOCL.

### Funding Information

This work has received funding from CNPQ, from the *Brazilian Federal Agency for Support and Evaluation of Graduate Education* (CAPES) and is part of the activities of *Kant in South America* (KANTINSA) Project, founding by *Minas Gerais Research Foundation* (FAPEMIG) and the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 777786.

### Acknowledges

We would like to thank Michel Bitbol, Stefano Osnaghi and Robert Hanna for their critical comments and helpful suggestions.

### References

- Abellán, C. & et al. (2018). Challenging local realism with human choices, *Nature*, 557, pp. 212-216.
- Abellán, C. & et al. (2015). Generation of fresh and pure random numbers for loophole-free Bell tests, *Phys. Rev. Lett.* 115 (25), 250403. <http://dx.doi.org/10.1103/PhysRevLett.115.250403>
- Allison, H. E. (1990). *Kant's theory of freedom*, Cambridge University Press.
- Allison, H. E. (1996). *Idealism and freedom: essays on Kant's theoretical and practical philosophy*, Cambridge University Press.
- Aspect, A.; Dalibard, J. & Roger, G. (1982). Experimental test of Bell's inequalities using time-varying analyzers, *Phys Rev Lett.*, 49 (25), 1804–1807.
- Aspect, A.; Grangier, P. & Roger, G. (1982). Experimental realization of Einstein-Podolsky-Rosen-Bohm Gedankenexperiment: A new violation of Bell's inequalities. *Phys Rev.*, 49. 91-94.
- Augustine. (2011). *On Grace and Free Choice, and Other Writings*. Cambridge University Press.
- Bell, J. S. (1964). On the Einstein-Podolsky-Rosen Paradox, *Physics*, 1 (3), 195-200.



- Bell, J. S. (2004). *Speakable and Unspeakable in Quantum Mechanics*, Cambridge University Press.
- Bitbol, M. (2010). Reflective metaphysics: understanding quantum mechanics from a Kantian standpoint. *Philosophica*, 83 (1), 53-83. <http://dx.doi.org/10.21825/philosophica.82161>
- Bitbol, M.; Kerszberg, P. & Petitot, J. (Eds.). (2009). *Constituting Objectivity: Transcendental Perspectives on Modern Physics*. Springer.
- Bitbol, M. & Osnaghi, S. (2013). Bohr's Complementarity and Kant's Epistemology. In *Bohr, 1913-2013, Séminaire Poincaré XVII*, 145-166. [http://dx.doi.org/10.1007/978-3-319-14316-3\\_8](http://dx.doi.org/10.1007/978-3-319-14316-3_8)
- Borges, M. (2019). *Emotion, Reason and Action in Kant*. Bloomsbury Academic.
- Brans, C. (1988). Bell's theorem does not eliminate fully causal hidden variables, *Int. J. Theor. Phys.*, 27 (2), pp. 219-226. <http://dx.doi.org/10.1007/BF00670750>
- Degorre, J.; Laplante, S. & Roland, J. (2005). Simulating quantum correlations as a distributed sampling problem, *Phys. Rev. A*, 72 (6), 062314. <http://dx.doi.org/10.1103/PhysRevA.72.062314>
- Dominik, R. & et al. (2018). Cosmic Bell Test Using Random Measurement Settings from High-Redshift Quasars, *Phys. Rev. Lett.*, 121, p. 080403. <https://doi.org/10.48550/arXiv.1808.05966>
- Falkenburg, B. (forthcoming). *From Heisenberg to Mittelstaedt and Scheibe: Kant's Impact on the "German Copenhagen School" of Quantum Mechanics*. In Biagioli, F. & Giovanelli, M. (eds.): *Neo-Kantian Perspectives on the Exact Sciences*.
- Fürst, M. & et al. (2010). High speed optical quantum random number generation, *Opt. Express*, 18, p. 13029.
- Esfeld, M. (2015). Bell's theorem and the issue of determinism and indeterminism. *Foundations of Physics*, 45(5), p. 471-482.
- Giustina, M. & et al. (2015). Significant-loophole-free test of Bell's theorem with entangled photons, *Phys. Rev. Lett.*, 115, p. 250401.
- Hall, M. J. W. (2010). Relaxed Bell inequalities and Kochen-Specker theorems, *Phys. Rev. Lett.*, 105, p. 250404.
- Hall, M. J. W. (2011). Local Deterministic Model of Singlet State Correlations Based on Relaxing Measurement Independence', *Phys. Rev. A*, 84, p. 022102.
- Handsteiner, J. & et al. (2017). Cosmic Bell Test: Measurement Settings from Milky Way Stars, *Phys. Rev. Lett.*, 118, p. 060401.
- Hanna, R. (2006). *Kant, Science and Human Nature*, Oxford University Press.
- Hensen, B. & et al. (2016). Loophole-free Bell test using electron spins in diamond: second experiment and additional analysis, *Sci. Rep.*, 6, p. 30289.
- Hensen, B. & et al. (2015). Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres, *Nature*, 526, pp. 682-686.
- Kant, I. (1998). [KrV], *Critique of pure reason*, translated by P. Guyer and A. Wood, Cambridge University Press.
- Kauark-Leite, P. (2012). *Théorie quantique et philosophie transcendantale: dialogues possibles*. Hermann,
- Kauark-Leite, P. (2017). Transcendental Versus Quantitative Meanings of Bohr's Complementarity Principle. In H. Folse & J. Faye (eds.), *Niels Bohr and Philosophy of Physics: Twenty-First Century Perspectives*. London: Bloomsbury Publishing, pp. 67-89.

- Ou, Z. Y. & Mandel, L. (1988). Violation of Bell's inequality and classical probability in a two-photon correlation experiment, *Phys Rev Lett.*, 61, 50-53.
- Rosenfeld, W. & et al. (2017). Event-ready Bell test using entangled atoms simultaneously closing detection and locality loopholes, *Phys. Rev. Lett.*, 119, p. 010402.
- San Pedro, I. (2014). Measurement independence, parameter independence and non-locality. *European Journal for Philosophy of Science*, 4(3), 369-374. <http://dx.doi.org/10.1007/s13194-014-0091-8>
- Shalm, L. K. & et al. (2015). Strong loophole-free test of local realism, *Phys. Rev. Lett.*, 115, p. 250402.
- Shih, Y. H. & Alley, C. O. (1988). New type of Einstein-Podolsky-Rosen-Bohm experiment using pairs of light quanta produced by optical parametric down conversion, *Phys Rev Lett.*, 61, 2921–2924.
- Scheidla, T. & et al. (2010). Violation of local realism with freedom of choice, *Proc. Natl. Acad. Sci. U.S.A.*, 107, p. 19708-19713.
- Shimony, A.; Horne, M. A. & Clauser, J. F. (1976). Epistemological Letters. Reprinted in *Dialectica*, 39
- Tapster, P. R., Rarity, J. G. & Owens, P. C. M. (1994). Violation of Bell's inequality over 4 km of optical fiber, *Phys Rev Lett.*, 73, pp. 1923–1926.
- Weihs, G. & et al. (1998). Violation of Bell's inequality under strict Einstein locality conditions, *Phys. Rev. Lett.*, 81, 5039.
- Wood, A. (eds.) (1984). Kant's Compatibilism, in *Self and Nature in Kant's Philosophy*, pp. 73–101.

## On the Architectonic Idea of Mathematics

Edgar Valdez

Associate Professor of Philosophy Department, Saint Peter's University, USA. Email: [evaldez1@saintpeters.edu](mailto:evaldez1@saintpeters.edu)

---

### Article Info

#### Article type:

Research Article

#### Article history:

Received 25 July 2024

Received in revised form  
29 July 2024

Accepted 29 July 2024

Published online 07  
August 2024

#### Keywords:

architectonic, geometry,  
mathematics, arithmetic,  
construction.

### ABSTRACT

The architectonic is key for situating Kant's understanding of science in the coming century. For Kant the faculty of reason turns to ideas to form a complete system. The coherence of the system rests on these ideas. In contrast to technical unity which can be abstracted a posteriori, architectonic ideas are the source of a priori unity for the system of reason because they connect our reasonable pursuit to essential human ends. Given Kant's focus on mathematics, in the architectonic and his critical philosophy more generally, we must have some sense of the architectonic idea of mathematics. In this paper, I argue for the key principles of the architectonic idea of mathematics: 1) because mathematics is grounded in a priori intuition, it is a peculiarly human activity; 2) the method of mathematics is one of a priori construction, a method only mathematics can employ and; 3) the objects of mathematics are extensive magnitudes. Given these principles, we can use the architectonic idea to have some clarity about how mathematics has dealt with historical development.

---

**Cite this article:** Valdez, E. (2024). On the Architectonic Idea of Mathematics. *Journal of Philosophical Investigations*, 18(47), 203-218. <https://doi.org/10.22034/jpiut.2024.18293>



© The Author(s).

<https://doi.org/10.22034/jpiut.2024.18293>

Publisher: University of Tabriz.

---

## **Introduction**

One way to consider the role and significance of Kant's philosophy for the future is to examine how Kant anticipated future developments in human investigation. This is best exhibited in Kant's account of an architectonic of reason. Kant's Doctrine of Method aims to determine "the formal conditions of a complete system of pure reason," where reason is the faculty of inferring and concluding from concepts (B, 735). The coherence of such a system rests on ideas that are grounded in an empirical consciousness but have no given or empirical object as referent. Among those formal conditions is systematicity, which Kant characterizes in the Architectonic. Given the importance of mathematics for the ends of reason and the many developments in mathematics since Kant, we should be able to express some of the principles of the architectonic idea of mathematics. In what follows I aim to give some exposition of what, for Kant, must be the architectonic idea of mathematics.

For Kant, mathematics concerns how we construct magnitudes a priori. First, I discuss architectonic unity, distinguishing it from technical unity and emphasizing its importance for the interests of reason. Reason requires that an architectonic idea is regulative. Second, I consider how mathematics is a peculiarly human science and why that peculiarity is tied to the essential ends of human reason. Mathematics takes as its substratum the a priori intuitions of space and time, the conditions for human sensibility, and they form the basis of nature. Third, I outline Kant's account of construction in mathematics. I then consider what for Kant are the objects of the mathematical domain, namely extensive magnitudes. Finally, I consider briefly how the architectonic idea helps Kant incorporate some developments in mathematics.

### **1. Architectonic Ideas**

Reason is the faculty of inferring from principles. The understanding is constrained by rules, but reason proceeds through principles. Reason seeks a coherence that is not given by empirical objects. Whereas the understanding serves to understand perceptions, reason seeks to comprehend possible experience, of which actual experience is only a part (B, 367). To do this reason turns to ideas, a conjunction of concepts that complete a system. These ideas are still only possible in an empirical consciousness that brings coherence to the manifold of what is given in experience. Ideas are a composite of concepts that we infer in order to give coherence to possible experience.

Systematicity is the fundamental feature of the architectonic because "systematic unity is that which first makes ordinary cognition into science, i.e., makes a system out of a mere aggregate of it, architectonic is the doctrine of that which is scientific in our cognition in general, and therefore necessarily belongs to the doctrine of method" (B, 860). Only through this systematicity is more knowledge not merely an aggregate but an integral part of our science. Systematicity anticipates more knowledge to be incorporated. It expects growth and development from within over time. For

Kant, architectonic structure is the source of science's rigor. In architectonic unity, method and structure form the unity of knowledge a priori, while contingent purposes and experiences form a technical unity a posteriori. In investigating the bounds of metaphysics, we can grasp an architectonic idea of metaphysics whose structure precedes its content.

A science's unity must come from its a priori methodology, not from already experienced answers or affections. This architectonic structure is distinct from a mere technical or coincidental unity, "[t]he whole is therefore articulated; and not aggregated: it can, to be sure, grow internally, but not externally, like an animal body, whose growth does not add a limb but rather makes each limb stronger and fitter for its end without any alteration of proportion" (B, 861). Kant expects the sciences to add to the cognition of the articulated whole of sciences so that the system of truths that unifies them should grow to be a fuller system. This articulated whole is not to be understood as finished, but complete as in forming a system. The system serves as ground for completion from within.

Technical unity arises from the contingent coincidence of cognition. In this case, the unity is determined by the parts but also exhausted by it. There is nothing from within in furtherance of which the unity can grow. Its completion stems from a terminus rather than a purpose. Growth must come from the a posteriori addition of further coincidence. Technical unity looks back while architectonic unity looks forward in anticipation. Architectonic unity stems from "a single supreme and inner end, which first makes possible the whole" (B, 861). In this case the unity precedes any of the parts unified and growth occurs from within. The science only grows if that growth is a part of that same idea of reason.

On Kant's view, philosophy admits of the same architectonic structure, though it has been poorly expressed at times, leading to confusion about its method and aims. Kant is working to recognize the architectonic idea of philosophy and its reasonable ends. It is insufficient merely to historically study the work of previous philosophers or adopt their truths. One must philosophize in accord with the architectonic idea of philosophy. While Kant hopes to bring about a revolution in philosophy, the revolution is not one that changes the architectonic idea of philosophy but works to realize it and bring philosophy closer to scientific status. Philosophy, for Kant, admits of a systematic unity that is the basis of its scientific status. Thus far, philosophers have been unable to set philosophy on the path to science because they have misread the end of philosophy.

In an architectonic idea we find a systematic unity that understands cognitions in light of the same purpose or function. Science requires systematicity because it must yield knowledge of necessary laws. Isolated moments of cognition can only yield knowledge of laws when they are united under a larger explanatory function. The architectonic of reason structures reason and in turn all systematic reasonable pursuits. Thus, it is the location within this a priori structure that indicates the function of a science. As such, functions are embedded in other functions and unite other functions under them. Much like concepts, these functions have both extension and intension. The

organic metaphor Kant uses to describe architectonic structure maps onto the intension and extension of functions within the architectonic. As an example, we think of the digestive system as having a purpose. This gives purpose to the disparate elements of the digestive system each with their own subordinate functions. The function of the digestive system itself is also an element of a larger function, i.e. the function of the organism. Particular sciences have explanatory functions. Those functions can be understood as being in service of a higher reasonable end. But also, within those sciences there are questions or issues that require a further determined idea. Kant uses this relationship to argue that the work of the Critique can be situated in a larger framework of science that alludes to a taxonomy of the sciences. In this larger framework there are subordinate sciences that require metaphysics but there are also broader interests of reason of which philosophy and metaphysics are only a part. As such, the architectonic serves its highest function, the essential ends of human reason. Kant uses the architectonic to argue that philosophers have thus far only dealt with the scholastic concept of philosophy, philosophy as concerned with rigor and precision. What he postulates as the ideal is philosophy in its cosmopolitan concept, as advancing the essential ends of humanity, calling on philosophers to see their work as tied to essentially practical ends.

Gabriele Gava argues that if we adhere strictly to the link to the essential ends of human reason, some of the disciplines we recognize as science today would fall short of the architectonic unity that Kant prescribes since humanity's essential ends are necessarily practical and physics and chemistry and the like are theoretical. Gava sees this tension due to a double function of the relationship between reason and its ends. One function concerns the immediate explanatory ends of a science and the other the mediate ends that point to the priority of practical philosophy.

Kant also claims that architectonic unity is only possible by making reference to the essential ends of reason. Essential ends are practical and are so treated by practical philosophy. This would imply the quite implausible consequence that only philosophy, insofar as it considers the systematic relationship between our knowledge and these essential practical ends, can become a science (Gava, 2005, 373).

If architectonic unity requires a science to recognize the essential ends of humanity, many seemingly united or systematic realms of inquiry would fall short of scientific status given their agnosticism with regard to the essential ends of humanity. Kant grants that we may be mistaken about the architectonic unity of some investigations, but he also seems committed to the idea of physics and chemistry as sciences. Gava proposes eliminating the requirement that architectonic unity necessarily be linked to the essential ends of human reason, holding that an a priori internal unity is enough to ensure scientific status while unity in philosophy would still prioritize the practical ends of reason over the theoretical. Gava's point is that systematicity concerns a certain structural feature that allows for the unity of cognitions. We can keep this feature without having

to turn to the essential ends of humanity. Such ends would seem to denote content of the function rather than a strictly structural feature. I doubt, however, that this tension rises to the level of problem or contradiction.

The architectonic idea is not part of the psychology of the scientist, or even philosophers or historians of science. That is, someone inquiring need not always be conscious of the mediate purpose of a science for that science to work toward that purpose or function.

Nobody attempts to establish a science without grounding it on an idea. But in its elaboration the schema, indeed even the definition of the science which is given right at the outset, seldom corresponds to the idea; for this lies in reason like a seed, all of whose parts still lie very involuted and are hardly recognizable even under microscopic observation. For this reason, sciences, since they have all been thought out from the viewpoint of a certain general interest, must not be explained and determined in accordance with the description given by their founder, but rather in accordance with the idea, grounded in reason itself, of the natural unity of the parts that have been brought together. For the founder and even his most recent successors often fumble around with an idea that they have not even made distinct to themselves and that therefore cannot determine the special content, the articulation (systematic unity) and boundaries of the science (B, 862).

Those of us working in service of scientific inquiry are likely to have an operative definition of a science that does not give full expression to the idea that situates it within the architectonic of reason. The biologist may not comprehend how biology serves the essential ends of reason even though biology as a science maybe be necessary for a coherent view of such ends. That might give us cause to question our definition or anticipate development but not to abandon the presumption of the scientific status of our idea. The cogs in a machine, the bees in a swarm, the organs in an organism need not be aware of the whole of the system for the system to be its intelligible organizing principle.<sup>1</sup> Though the explanations within a science may be ultimately in service of the essential ends of humanity, the ends of humanity need not be expressed in every cognition of that science. The duality of function that Gava points to is not one that sees these functions as in conflict with one another. For Kant, we cannot simply talk about the structure of reason in the abstract. We must talk of the structure of human reason; the structure then cannot take its form without situating the essential ends of humanity as that in service of which all science ultimately gets its function. To acknowledge the essential practical ends of humanity is not to abandon the ends of our science. Rather these functions should be understood as ordered.

---

<sup>1</sup> I would also hold that the collection or cohesion of scientific investigation is not enough to meet the standard for architectonic unity absent a connection to the ends of humanity but such disagreement is not central to the concern I am addressing concerning mathematics.

Kant rejects a Platonist notion of idea but uses the term because of its metaphysical baggage. As Onora O'Neill, (1992, 285) points out these ideas are to be thought of as precepts not given realities,

He defends his appropriation of this misleading Platonic term, not because but *in spite* of its metaphysical resonance. The term suits not because Kant too wants to endorse a classical, theoretical conception of reason, as correspondence of thought to its real archetypes, but because Plato's Ideas are potent symbols of striving for the most encompassing unity. The Platonic Ideas are an image of the unity of the highest principles that guide a quest for the Good and the Beautiful as well as the True. Kant allows himself this borrowing, which parallels his own three fundamental questions, but rejects the entire Platonic account of the metaphysical basis of unity and success in these quests. He firmly rejects all thought that his Ideas of Reason correspond to any real archetypes, and adopts a position that is irreconcilable with any form of the Platonic vision of Ideas as patterns for knowledge and mathematics (Oneill, 1992, 285).

Ideas for Kant are inferred for the sake of coherence so they do not depend on any one agent thinking them, but they are not the foundations of any metaphysical claims. There is a distinction between an idea within the architectonic framework and our contingent recognition of it. The scientific status of particular fields of inquiry is conditioned by an architectonic idea but not by our knowledge of the architectonic idea. Our recognition of an architectonic idea may be muddled or even absent. Mathematicians and physicists need not be thinking of the essential practical ends of human reason when engaged in mathematics in order for mathematics to be a part of the framework of human reason. In this sense, the unity Gava prescribes is the one we most concern ourselves with in identifying sciences.

Thus, the metaphysics of nature as well as morals, but above all the preparatory (propaedeutic) critique of reason that dares to fly with its own wings, alone constitute that which we can call philosophy in a general sense. This relates everything to wisdom, but through the path of science, the only one which, once cleared, is never overgrown, and never leads to error. Mathematics, natural science, even the empirical knowledge of humankind, have a high value as means, for the most part to contingent but yet ultimately to necessary and essential ends of humanity, but only through the mediation of a rational cognition from mere concepts, which, call it what one will, is really nothing but metaphysics (B, 878).

The value that these sciences have must be examined through rigorous metaphysics, but this need not be a project for the physicist or mathematician according to Kant.



Philosophy, however, must at some point address the essential ends of human reason directly and systematically because the internal a priori relation science has to the essential ends of humanity are to be understood by the philosopher. Part of the task of philosophy is to understand such ends and situate them in relation to the essential ends of humanity, something the ideal philosopher is tasked with investigating. The question of the essential ends of humanity falls within the immediate ends of philosophy. Gava cites<sup>1</sup> Kant's remark of the logician and mathematician being only artists of reason—in contrast to the philosopher as legislator—as evidence that the possibility of only philosophy meeting the standard for science is one Kant entertains. But though this is perhaps a careless remark on Kant's part it is important to note the context in which he makes it. When Kant distinguishes between the scholastic concept and the cosmopolitan concept of philosophy, he claims that only the latter relates all science to the essential ends of humanity. It is only in this respect that the mathematician and logician would not be up to the task because the task is a philosophical one. As such it is one for the ideal philosopher,

From this point of view [that of the ideal philosopher] philosophy is the science of the relation of all cognition to the essential ends of human reason (*teleologia rationis humanae*), and the philosopher is not an artist of reason but the legislator of human reason. It would be very boastful to call oneself a philosopher in this sense and to pretend to have equaled the archetype, which lies only in the idea. (B, 867).

By this Kant means that we have not yet come to realize the relation between all our cognition and the essential ends of humanity. The mathematician may be versed in the rational cognition of geometry but that does not equip her to relate even mathematical cognition to the essential ends of humanity. As Guyer notes only an architectonic can hope to bridge the seeming divide between the determinism of a system of nature and the fundamental freedom that must be postulated for practical human ends,

Kant's basic idea is that to think rationally is to think systematically: to think about nature rationally is to think about it systematically; to think about our own conduct rationally is to think about it systematically; and ultimately, we must think about how mankind could collectively achieve a systematic union of ends within the system of nature (Guyer, 2005, 3).

This occurs only when philosophy turns its attention to these essential ends. For Kant, any attempt to offer an assessment of reason requires making sense of mathematics. This has been a

---

<sup>1</sup> The scholastic philosopher cannot thus be considered a real philosopher, but only an “artist of reason” [Vernunftkünstler], along with the mathematician, the logician and the naturalist. By contrast, the true philosopher is a “legislator of reason” [Gesetzgeber der menschlichen Vernunft; KrV, A 839/B 867], even though this idea of the philosopher as a “knower of wisdom” [Kenner der Weisheit; V-Met-L2/Pölitiz; AA 28, 534] remains only an archetype. P.378

source of confusion for philosophers because metaphysicians have continually *misused* mathematics in hopes of advancing philosophy,

But what obscured the fundamental idea of metaphysics from yet another side was that, as a priori cognition, it shows a certain homogeneity with mathematics, to which, as far as a priori origin is concerned, it is no doubt related; but the comparison between the kind of cognition from concepts in the former with the manner of judging a priori through the mere construction of concepts in the latter requires a difference between philosophical and mathematical cognition—thus a decided heterogeneity is revealed, which was always felt, as it were, but was never able to be brought to distinct criteria. Thus, it has been the case until now that since philosophers themselves erred in the development of the idea of their science, its elaboration could have no determinate end and no secure guideline.... (B, 872).

Kant distinguishes between the cognition from concepts of the philosophical method and the cognition from the construction of concepts of the mathematical method. The method of mathematics is not available to philosophy, but too many philosophers have attempted to use it. In distinguishing these methods Kant offers some principles of the architectonic idea of mathematics.

## 2. Mathematics as Human Endeavor

The critical standpoint shifts metaphysics from a concern of how cognition conforms to objects to a concern for how objects conform to cognition, a reversal of the priority of matter over form. Kant views the rationalists before him as investigating the matter of the world but his Copernican Revolution in philosophy turns the attention to the form of the objects of cognition. That is, the form of our experience of the world is embedded in our epistemic condition and not in things in themselves. Thus, an investigation of nature must turn to an investigation of the a priori structure of our experience. Rather than consider what the world is like from a perspective outside of it, we can ask only from a perspective within the world. It follows then that mathematics is also a distinctly human investigation. Mathematics draws on the formal conditions of space and time but space and time are a priori intuitions *for us*. Thus, mathematics is wholly a priori but only an investigation of what we have ourselves put into the objects of investigation. Other rational beings may be subject to the same logical constraints but nothing requires that such beings have the same intuitive frameworks for experiencing the world. Rational beings with different intuitive frameworks—including computers that can adhere to logical constraints and make large computations very quickly—would mathematize differently, “all these principles, and the representation of the object with which [mathematics] occupies itself, are generated in the mind completely *a priori*...” (B, 299) This is not then an idiosyncratic or solipsistic activity. It remains wholly a priori but its intelligibility has its seat in the structure of the human subject. The coherence

of our empirical consciousness requires an a priori spatiotemporal manifold and mathematics is the science through which we can treat this manifold as object.

Imagination is the faculty of representing an object in intuition that is not present. This faculty is divided in two. The reproductive imagination offers reproductions and distortions of objects that were previously given while the productive imagination presents objects that were not previously given. The productive imagination is a figurative synthesis. It is spontaneous and determines the sense a priori in accordance with the unity of apperception (B, 151). It yields only a priori determinations of space and time. An object is constructed when an intuitive manifold is united with a concept. That is, we only have a geometrical object when we have brought the spatial manifold under a conceptual unity. To cognize a line, one must draw it, but the drawing that is required is the a priori construction of it, regardless of whether or not it is accompanied by an empirical pencil and paper drawing. The unity of this act is the unity of consciousness in the concept of a line. Space as the form of outer intuition guides the construction. While the reproductive imagination is passive and governed by sense and empirical laws, the productive imagination yields the manifold of sense in accordance with a concept.

Imagination is the faculty of representing an object even without its presence in intuition. Now since all of our intuition is sensible, the imagination, because of the subjective condition under which alone it can give a corresponding intuition to the concepts of understanding, belongs to sensibility; but insofar as its synthesis is still an exercise of spontaneity, which is determinative and not, like sense, merely determinable, and can thus determine the form of sense a priori in accordance with the unity of apperception, the imagination is to that extent a faculty for determining the sensibility a priori, and its synthesis of intuitions, in accordance with the categories, must be the transcendental synthesis of the imagination, which is an effect of the understanding on sensibility and its first application (at the same time the ground of all other applications) to objects of the intuition that is possible for us (B, 151-2).

Since it is the figurative synthesis that serves as the mathematical object and not the empirical image left behind, mathematics occurs only in the subject. It concerns the outside world, but the activity is in the subject.

Mathematics is connected to essential human ends because it is necessary for the various natural sciences that allow us to understand nature. As Kant puts it in the *Metaphysical Foundations*, “in any special doctrine of nature there can be only as much proper science as there is mathematics therein.” (4, 470) A robust understanding of nature is necessary for the essential ends of human reason because though such ends may be moral, they take place against the backdrop of the natural world. Thus, the full synthesis that brings together the determinism of nature and the freedom

necessary for morality is only possible with a robust mathematical device, the same subjective condition for reasoning about the essential ends of humanity. We need good mathematical knowledge of nature to make good moral decisions about peace, climate change, agricultural policy etc. Devices that can compute quickly may produce data but data alone is not cognizable. More importantly, although being precise about those connections may prove elusive, we need not see it as a necessary step in identifying the human element of the architectonic idea of mathematics.

### 3. Construction

The mathematical method is integral for the architectonic idea of mathematics. Kant distinguishes between the philosophical and mathematical method.

Mathematics proceeds through the construction of concepts. This method always treats its objects in concreto. [Philosophy] confine itself solely to general concepts, [mathematics] cannot do anything with the mere concepts but hurries at once to intuition, in which it considers the concept *in concreto*, although not empirically, but rather only as one in which it has exhibited a priori, i.e., constructed, and in which that which follows from the general conditions of the construction must also hold generally of the object of the constructed concept (B, 743-4).

In mathematical judgments, there is always an object before us. But mathematics is not an a posteriori endeavor in which we keep mathematical objects before us by relying on empirical objects or sensory affection. Rather we construct the objects of investigation, we produce an object in intuition that exhibits the formal conditions or features we are investigating.

For Kant, mathematics proceeds by definitions, axioms and demonstrations. A definition presents completely the concept of something while delineating its limits and contexts. The only concepts that allow for real definitions are those with an arbitrary synthesis with a priori construction because other attempts at definitions fall short of an exhaustive delineation of constitutive properties. Axioms are synthetic in that they appeal immediately to the intuition of space and are a priori in that they are necessarily true. Mathematics can synthetically combine two constructed concepts to yield another that can be likewise constructed in intuition and this is what Kant calls a demonstration. More precisely a demonstration that proceeds through the construction of concepts he calls a *Mathema* (B, 764). This constructed intuition is a priori. Given their immediacy and construction in intuition, axioms are intuitive principles to be distinguished from discursive principles. Mathematical construction also requires postulates, immediately certain intuitive principles that are necessary for construction,

Now a postulate in mathematics is the practical proposition that contains nothing but the synthesis through which we first give ourselves an object and generate its concept, e.g. to describe a circle with a given line from a point on a plane; and such

a proposition cannot be proved, since the procedure it demands is precisely the procedure through which we first generate the concept of such a figure (B, 287).

These postulates, much like the postulates for God, freedom, and immortality are practical in that they are necessary for activity. The metaphysical objects are necessary for moral behavior and the mathematical postulates are necessary for geometrical construction. Unlike the postulates of the metaphysical objects, mathematical postulates concern a procedure or activity through which we construct something necessary for a mathematical demonstration<sup>1</sup>.

For Kant, space does not give us geometrical objects, rather it conditions their construction. These objects are constructed through a conceptual synthesis. But the formal conditions of space give us the content that we synthesize to construct geometrical objects. This highlights the distinction between space and geometry. As Kant writes in his response to Kästner, metaphysical space “is considered in the way it is given, before all determination of it in conformity with a certain concept of object” (20, 419), Geometrical space is constructed through a determinate synthesis which distinguishes it from the a priori intuition of space. Space conditions the geometrical representations that are possible but not because it already contains such representations. Rather it does so by conditioning the synthetic construction of geometrical spaces and objects. Space provides the manifold that is in turn limited for our construction.

The activity of the productive imagination seems to presume or presuppose a perceptual analog to the precise mathematical properties of continuity and homogeneity. The condition for phenomenological or perceptual activities in intuitive spatial representation, prior to any conceptual synthesis, is distinct from the mathematical properties that require synthesis. The latter are not merely read off from our perception of our spatial representation. Quite the contrary, from the mere a priori representation of space, we cannot establish or discover properties beyond those considered in the Metaphysical Exposition. The space of geometrical activities is itself the product of an a priori synthesis.

In geometry, the activities of drawing a line, extending a line, circumscribing a circle—often referred to as straight-edge and compass—are foundational in that they make all other constructions possible. These elementary activities correspond to a priori phenomenological activities that have been united with a concept to produce an a priori synthesis. For Kant, the possibility of such constructions is uniquely conditioned by the a priori intuition of space. As he puts it in *On Kästner’s Treatises*:

That, however, the possibility of a straight line and a circle cannot be demonstrated mediately through inferences, but only immediately through the construction of

<sup>1</sup> In “Kant on Parallel Lines: Definitions, Postulates, and Axioms,” Jeremy Heis offers a detailed account of Kant’s characterization of mathematical postulates and the difficulty of considering the parallel postulate to be a genuine postulate. In this paper, I am taking Kant at his word about the role postulates play in pure mathematics, without such a detailed examination.

these concepts (which is not at all empirical), is due to the fact that among all constructions (exhibition which are determined according to a rule in intuition *a priori*) some must be the first, such as the drawing (*Ziehen*) or the describing (in thought) of a straight line and the rotation thereof around a fixed point, where neither the latter can be derived from the former, nor these from any other construction of the concept of magnitude (20, 411).

The geometrical activities in question are inseparable from the magnitudes provided by intuition. To construct mathematically is to remain in immediate relation to intuition. Onof and Schulting (2014) show that this distinction between space as form of intuition and as formal intuition is present throughout the Critical period as evidenced in the Transcendental Deduction

Space represented as object (as actually required in geometry), contains more than the mere form of intuition, it also contains the composition of the manifold given in accordance with the form of sensibility in an intuitive representation, so that the form of intuition gives the manifold, but the formal intuition gives the unity of the representation. I ascribed this unity in the Aesthetic merely to sensibility, only to note that it precedes all concepts, although it does presuppose a synthesis, which does not belong to the senses but through which all concepts of space and time first become possible. For since through it (as the understanding determines the sensibility) space or time are first given as intuitions, the unity of this *a priori* intuition belongs to space and time, and not to the concept of the understanding (J24) (B160).<sup>1</sup>

Cognition of objects requires a synthesis. Space as the form of intuition is given, but it cannot be cognized. While singular, it merely forms human receptivity to the world and cannot be made a unity with any concepts. The formal intuition of space has been subjected to a kind of pre-conceptual unity so that rather than perceive a mere manifold there can be a representation of space as object. But once geometrical activity has begun, the manifold has been made a unity and synthesized with concepts. It has been subjected to a series of conceptual syntheses, among which might be continuity and homogeneity.<sup>2</sup>

---

<sup>1</sup> This footnote is key to the debate in Kant scholarship between conceptualism and non-conceptualism. In the former, all unity, even the form of sensibility, is owed to concepts. In the latter, sensibility to concepts requires a unity that is not owed to the understanding. I side with non-conceptualists in this regard although I think a conceptualist can recognize the same elements of the architectonic idea of mathematics. Where the conceptualist might take issue is in my characterization of construction.

<sup>2</sup> Part of the case that I elsewhere make in considering Kant's views on geometry is that geometrical construction requires many syntheses and many of these we implicitly assume are necessary for all geometrical activity although they are merely a subset of all the geometrical syntheses that are possible.

#### 4. Magnitudes

Mathematics constructs magnitudes. For Kant, a magnitude is given as continuous and thus extensive. Magnitudes draw on the formal properties of the sensible condition—space and time—and are not the empirical images left behind. As Daniel Sutherland points out, Kant's view of magnitude draws on the Greek theory of proportions which focuses on homogenous continuous magnitudes, irrespective of numbers.

Kant thinks of mathematical cognitions as cognitions of magnitudes. Since his view of magnitudes derives from the Euclidean tradition, his account of mathematical cognition turns on the cognitions that make the theory of proportions possible. Those cognitions include the cognition of comparative size relations by means of the cognition of equality and part-whole relations. They also include the cognition of the part-whole composition relations of magnitudes (Sutherland, 2006, 539).

Only extensive magnitudes can be constructed and put into relations with one another in this way. Because all magnitudes are limitations of space and time, they are all situated within the same whole.

Since the mere intuition in all appearances is either space or time, every appearance as intuition is an extensive magnitude, as it can only be cognized through successive synthesis (from part to part) in apprehension. All appearances are accordingly already intuited as aggregates (multitudes of antecedently given parts), which is not the case with every kind of magnitude, but rather only with those that are represented and apprehended by us as extensive (B, 203-4).

With extensive magnitudes, the representation of parts requires and presupposes the representation of the whole.<sup>1</sup> Mathematical construction is a matter of constructing the parts by delimiting the whole.

For Kant investigations are only properly scientific insofar as they admit of the application of pure mathematics. But such application is not possible if all we can do is enumerate the elements of a domain rather than construct extensive magnitudes. It is not simply that mathematics is concerned with *quanta*, but rather that it constructs magnitudes. These magnitudes are continuous and intuitive. Intuition conditions mathematical construction by providing intuitive content but that content alone does not make mathematics, it must be made a synthetic unity to form concepts and judgments,

---

<sup>1</sup> In "Infinity and Kant's Conception of the 'Possibility of Experience,'" *Philosophical Review* Volume 73. No. 2, Charles Parsons argues that there is a tension in Kant holding there to be certain infinite features to our intuitive representations and characterizing intuitions as singular and immediate. Parsons notes that Kant does not think we can have an intuitive grasp of an actual infinity. So, the infinite features of intuition require a going over and processing of our experience. A kind of zooming in and out but such processing then is not singular and immediate.

In fact, it is not images of objects but schemata that ground our pure sensible concepts. No image of a triangle would ever be adequate to the concept of it. For it would not attain the generality of the concept, which makes this valid for all triangles, right or acute, etc., but would always be limited to one part of this sphere. The schema of the triangle can never exist anywhere except in thought, and signifies a rule of the synthesis of the imagination with regard to pure shapes in space (B, 180).

Mathematical knowledge is always expressed in concepts. The *in concreto* method allows mathematics to attain universal conceptual knowledge while still being tethered to an *a priori* intuitive representation,

Philosophical cognition thus considers the particular only in the universal, but mathematical cognition considers the universal in the particular, indeed even in the individual, yet nonetheless *a priori* and by means of reason, so that just as this individual is determined under certain general conditions of construction, the object of the concept, to which this individual corresponds only as its schema, must likewise be thought as universally determined (B, 742).

Mathematics is not merely a matter of turning to intuition. For Kant, all genuine cognition must turn to intuition. But in the case of mathematics there is a particular mode of tending to intuition that denotes the particular content that intuition provides. To be clear, I have distinguished three features of Kant's view concerning what denotes mathematics—who constructs, how construction occurs, and what construction produces—but these features are not separable. The way I have characterized construction requires that we view construction as a peculiarly human activity and as necessarily concerned with extensive magnitudes.

## 5. Consequences

The architectonic idea of mathematics brings coherence to the mathematical domain for interpreting Kant in light of developments in mathematics. This is useful both with respect to developments since Kant's day as well as those developments that are still to come. The architectonic idea of mathematics allows us to identify something as mathematical that Kant originally excluded and something as not mathematical that Kant originally included. Kant denies the intuitive possibility of certain non-Euclidean objects. NonEuclidean manifolds were not considered a part of geometrical activity. Kant may have viewed geometry unconventionally, but he did not include non-Euclidean geometry in the mathematical domain. The architectonic idea calls for constructing magnitudes in space with certain delimiting concepts. Coherent with that idea are various kinds of spaces, geometrical in virtue of the same thing, their reliance on the *a priori* intuition of space. Kant's error in rejecting objects of non-Euclidean geometry stems either from



letting a founder like Euclid or Newton denote the idea of mathematics or from considering the empirical image associated with mathematical construction to be indicative of the a priori form of intuition. The methods and objects of non-Euclidean geometries arise from the same architectonic idea and are necessarily a part of maintaining the unity of the idea of geometry. While Kant was committed to certain strictly Euclidean principles of geometry, the architectonic idea allows us to recognize and rectify this error. What for contingent historical reasons seemed outside the scope of geometrical activity can through the architectonic idea of mathematics be understood as grounded in the same internal relations. Non-Euclidean geometry is a process of taking precise conceptual analogs to phenomenological activity and synthesizing them with the content provided by the a priori intuition of space. Non-Euclidean construction requires space for its concepts to have an intuitive referent and not be a matter of mere play.

Kant also holds that arithmetic is constructive in the same manner as geometry, but arithmetical procedures do not construct magnitudes in the way geometry does. This is not a concern about syntheticity<sup>1</sup> but about mathematical construction. Kant acknowledges a difference between geometry and arithmetic in a letter to Schulz when he says that arithmetic ‘has no axioms, because it actually does not have a quantum, i.e., an object of intuition as magnitude, for its object, but merely quantity, i.e., a concept of a thing in general by determination of magnitude’ (10:555-6). Kant grants that while particular numbers when applied to intuition have spatiotemporal referents, the mere consideration of the relationships of the concepts of magnitudes is itself prior to any application to objects. Thus, the consideration of arithmetical magnitudes would seem to require a mediating concept for the application to objects.

Arithmetic’s reliance on intuition is quite different from the turn to intuition in geometry. Whereas geometry requires the formal conditions of space to provide content in order to construct any of its objects, arithmetic relies on intuition for the one after another (*hintereinander*) succession that is a part of the arithmetical process. As such, arithmetic does not construct mathematically or deal with extensive magnitudes. In this way, we can also see that it does not have the feature of being peculiarly human in the same way since it does not draw on the form of human subjectivity in the same way. Again, the technical unity or merely contingent contiguity of arithmetic regularly being employed in conjunction with mathematical activity contributes to seeing it as united with the mathematical. Within the architectonic idea, we must recognize arithmetic as having a different internal relation to human reason.

---

<sup>1</sup> There are accounts of analysis that require considering the activity of arithmetic to be synthetic. I have in mind something like R. Lanier Anderson’s, “It Adds Up After All: Kant’s Philosophy of Arithmetic in Light of the Traditional Logic,” *Philosophy and Phenomenological Research*, 69 (3), in which he argues that operator functions cannot be expressed through containment relations. But though certainly related the question of mathematical construction is distinct from the question of syntheticity and my point is that arithmetic does not construct magnitudes.

These errors in Kant are often used to argue that he either grossly misunderstood mathematics or that mathematics has changed so fundamentally since Kant that we cannot reconcile the claims he makes about mathematics with contemporary perspectives. Architectonic ideas should give us some pause in taking either of those two positions to an extreme. Even in laying out the framework of the architectonic Kant leaves room for progress and development, not simply so that others can continue what he began but also as a way of recognizing that he may have only a limited or contingent understanding of the a priori internal relations of scientific inquiry. As Paula Manchester argues, in the doctrine of method the architectonic precedes the history of reason precisely because Kant is leaving open a place in the schema for development, a place to be filled in by other rational investigators (Manchester, 2008, 147). If architectonic ideas are regulative with regard to functions and the essential ends of humanity, it follows that we ought to work toward understanding what serves as some of the content of the internal ideas of our scientific investigations. These ideas serve to direct the sciences and discovering or expressing them is a difficult task for us but a necessary one. Given that mathematics is devoid of empirical content and that it is the first evidence we have of the possibility of synthetic a priori principles, it should be among the first sciences whose architectonic idea we work to discover and express clearly. I have focused on questions of pure mathematics but for Kant this architectonic framework is for making sense of all scientific inquiry. We do not have to take Kant's word that chemistry must be understood a certain way. Rather, we must take developments in the sciences as opportunities to identify and further refine our understanding of the internal systematic unity of the sciences.

## References

- Gava, G. (2014). Kant's Definition of Science in the Architectonic of Pure Reason and the Essential Ends of Reason, *Kant-Studien*, 105 (3), 372-393. <http://dx.doi.org/10.1515/kant-2014-0016>
- Guyer, P. (2005). *Kant's System of Nature and Freedom: Selected Essays*, Oxford University Press.
- Kant, I. (1999). *Correspondence*, Edited by A. Zweig, Cambridge University Press.
- Kant, I. (1998). *The Critique of Pure Reason*, Edited by P. Guyer & A. Wood, Cambridge University Press.
- Kant, I. (2014). On Kästner's Treatises," translated by Ch. Onof & D. Schulting, *Kantian Review*. 19 (2), 305-313. <https://doi.org/10.1017/S1369415414000077>
- Manchester, P. (2008). Kant's Conception of Architectonic in its Philosophical Context, *KantStudien*, 99 (2), 133-151. <http://dx.doi.org/10.1515/KANT.2008.010>
- Oneill, O. (1992). Vindicating Reason in *Cambridge Companion to Kant*, Edited by P. Guyer, Cambridge University Press.
- Onof, Ch. & Schulting, D. (2015). Space as Form of Intuition and as Formal Intuition: On the Note to B160 in Kant's Critique of Pure Reason, *Philosophical Review*. 124 (1), 1-58. <http://dx.doi.org/10.1215/00318108-2812650>
- Sutherland, D. (2006). Kant on Arithmetic, Algebra and the Theory of Proportions, *Journal of the History of Philosophy*. 44 (4), 533-558. <http://dx.doi.org/10.1353/hph.2006.0072>

## Many Shades of Love in Kant

Maria Borges 

University of Santa Catarina, Brazil. Email: [maria.borges@ufsc.br](mailto:maria.borges@ufsc.br)

| Article Info                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | ABSTRACT                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p><b>Article type:</b><br/>Research Article</p> <p><b>Article history:</b><br/>Received 21 May 2024<br/>Received in revised form<br/>21 July 2024<br/>Accepted 24 July 2024<br/>Published online 07<br/>August 2024</p> <p><b>Keywords:</b><br/>love, sympathy, affect,<br/>passion, sex</p>                                                                                                                                                                                                                                                                 | <p>Kant is usually considered a cold moralist who does not give any importance to feeling and emotions. In this paper I show that Kant has a place for love although he uses this word in different meanings through his work. I will analyze the use and meaning of five different uses of the term love: self-love, practical love, love as affect, love as passion, and sexual love. I show that Kant has a place for love in his theory, in a plurality of shades and meanings, going from the practical love to romantic love. Some of their expressions are meaningful for the moral life, such as practical love, some are not. Kant portrait romantic love as a silly affect, sometimes as a dangerous passion who can even call for medication. And about sex, he claims that it is nothing but the use of the other as a means, which may obtain a higher juridical status if this use is reciprocal, in the case of marriage.</p> |
| <p><b>Cite this article:</b> Borges, M. (2024). Many Shades of Love in Kant. <i>Journal of Philosophical Investigations</i>, 18(47), 219-232. <a href="https://doi.org/10.22034/jpiut.2024.62599.3829">https://doi.org/10.22034/jpiut.2024.62599.3829</a></p> <div><div></div><div>© The Author(s).<br/><a href="https://doi.org/10.22034/jpiut.2024.62599.3829">https://doi.org/10.22034/jpiut.2024.62599.3829</a></div><div>Publisher: University of Tabriz.</div></div> |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |

## Self-love

We all have in mind the aim of the *Groundwork* and the *Critique of practical reason*: to prove that we can have morality without feelings of pleasure and displeasure. More than that, that the true moral worth of an action is at least revealed more clearly if we don't have any emotion, not even sympathy for the fate of others. Since the Henson's article "What Kant might have said: moral worth and the overdetermination of a dutiful action" (HENSON, 1979), much has been said about that the presence of feelings and emotions, and now the majority of commentators agree that emotions do not really make the action without moral worth. However, I still think that feelings cannot be the motivation or the incentive of a moral action, although they can be present, and the mere presence does not make the action morally unworthy. Then the cold philanthropist could be not that cold and still perform a moral action.<sup>1</sup>

In the *Critique of Practical Reason*, Kant links the two faculties that are important for action. Although these faculties have been absent in the *First Critique* and in the *Groundwork*, Kant here brings back the psychology of faculties, that he inherited from Baumgarten. But how could he reconcile his moral theory with a psychology of desire and pleasure? The solution was to relate the faculty of pleasure and displeasure with the inferior faculty of desire.

Kant claims in theorem II of the KpV:

Material practical rules place the determining ground of choice in the pleasure or displeasure to be felt in the reality of some object, and they are not related with the superior faculty of desire or to morality. They are all principles of the same kind that belong to the *principle of self-love or happiness* (KpV, AA 5, 22).

And in the Corollary of § 3 of the KpV Kant claims that

All *material* practical rules put the determining ground of the will in the *lower faculty of desire*, and were there no merely formal laws of the will sufficient to determine it, then neither could *any higher faculty of desire* be admitted (KpV, AA, 5, 22).

If determination of the will is based on pleasure and displeasure, only the inferior faculty of desire is concerned and we do not have any practical law, only the principle of self-love. Then, to do good because you feel pleasure in someone's happiness is not a moral action but belongs to the principle of self-love.

## Practical love

In the *Doctrine of virtue*, love comes back again, not in a depreciative way, but as the practical love of beneficence, not related to the feeling of pleasure. This idea that Kantian practical love is not

---

<sup>1</sup> For the relation between emotions and action, see Shermann (1990), Borges (2004). Cohen (2014), Morrison (2008).

related to delight is shown in the *Doctrine of Virtue* in the Section XII of the Introduction, named “Concepts of what is presupposed on the part of feeling by the mind’s receptivity to concepts of duty as such”. One of these concepts is love of human beings. This love is not a feeling but a conduct:

Love is a matter of feeling, not of willing, and I cannot love because I will to, still less because I ought to (I cannot be constrained to love); so, a duty to love is an absurdity. But benevolence (*amor benevolentiae*), as conduct (*als ein Tun*), can be subject to a law of duty. However, unselfish benevolence toward human beings is often (though very inappropriately) also called love (TL, AA 6, 401).

By love of human beings, Kant means, not the love of delight (*complacentia*), but the love of benevolence (*benevolentia*), since the latter could be demanded from someone, but not the former, given that it would be a contradiction that somebody should have the obligation to feel pleasure:

So, the saying “you ought to love your neighbor as yourself” does not mean that you ought immediately (first) to love him and (afterwards) by means of this love do good to him. It means, rather, *do good* to your fellow human beings, and your beneficence will produce love of them in you (as an aptitude of the inclination to beneficence in general). Hence the love that is delight/ *Liebe des Wohlgefallens* (*Amor complacentia*) is direct. But to have a duty to this (which is pleasure joined immediately to the representation of an object’s existence), that is, to have to be constrained in take pleasure is a contradiction (TL, 6, 402).

Love of human beings, as a concept necessary for the mind’s receptivity to concepts of duty as such, is not a feeling based on the delight for helping other people, but an inclination to beneficence. The same holds for the duties of virtue to others.

Kant enumerates two ends that should be considered as duties: self- perfection and other people's happiness. These two ends lead to two different kinds of duties: the duties of man related to him, and duties related to others, among which we find the *duty of love to other human beings*, which consists in promoting the happiness of others. This love is not a feeling but a maxim of benevolence that leads to beneficent actions:

In this context, however, love is not to be understood as *feeling*, that is, as a pleasure in the perfection of others, love is not to be understood as *delight* in them (since others cannot put one under obligation to have feelings). It must rather be thought as the maxim of *benevolence* (practical love), which results in beneficence (TL, 6, 449).

There is no delight in practical love because it is impossible to have a duty to have a feeling, even if it is a contentment with the happiness of others. Practical love is not properly speaking a love-feeling, but a maxim to perform beneficent actions.

### **The revival of Sympathy**

The duties of love are divided in duties of beneficence, gratitude, and sympathy. The duty of beneficence is the duty to promote according to one's means the happiness of others in need. While benevolence is satisfaction in the happiness (well-being) of others, beneficence is to make the happiness of others one's end. The duty to gratitude is honoring a person because of a benefit he has rendered us.

And then sympathy comes back as a duty, after its clear disapproval as a feeling with no moral value in the *Groundwork*. How can we conceive a duty to have a feeling? Is sympathy a feeling or it is now a propensity to perform a benevolent action?

The difference between sympathy in the *Groundwork* and in the *Doctrine of Virtue* is that in the latter there is a division between the capacity and the will to share in others' feelings (*Humanitas practica*) and the receptivity to the feeling of joy and sadness of the other (*Humanitas aesthetica*). Kant claims that

the first is free, and is therefore called sympathetic (*communio sentiendi liberalis*); it is based on practical reason. The second is unfree (*communio sentiendi illiberalis, servilis*); it can be called communicable (since it is like receptivity to warmth or contagious diseases), and also compassion, since it spreads naturally among human beings living near one another. There is obligation only to the first (TL, AA 6, 456).

Sympathy related to *humanitas practica* is a duty, but compassion, related to *humanitas aesthetica* is not a duty, because it does not lead to beneficence. To feel the pain of other human being and do nothing to alleviate it will only increase the suffering in this world.

The § 34 of the *Doctrine of Virtue* shows that the only sympathy we could accept is the practical one, not really a feeling, but a maxim of helping people in distress. There is a difference between sympathy and compassion, a feeling that only increases the suffering of the world. Kant even claims that this is an insulting kind of beneficence:

But there cannot possibly be a duty to increase the ills in the world and so to do good from compassion. This would be an insulting kind of beneficence, since it expresses the kind of benevolence one has toward someone unworthy, called pity (TL, AA, 6, 457).

However, the § 35 of the *Doctrine of Virtue* surprises us by stressing the importance of feelings in the accomplish of beneficence. If sympathy in order to be moral, should be active, now come

into play compassionate natural feelings. And to cultivate these feelings are said to be an “indirect duty”.

And then comes the famous and ambiguous passage:

It is therefore a duty not to avoid the places where the poor who lack the most basic necessities are to be found but rather to seek them out, and not to shun sickrooms or debtors’ prison and so forth in order to avoid sharing painful feelings one may be not able to resist. For this is still one of the impulses that nature has implanted in us to do what the representation of duty alone might not accomplish (TL, AA 6, 457).

What is the meaning of that? Should we stimulate the development of this kind of feelings? Many commentators consider that the role that Kant attributes to sympathy is, therefore, of a provisory moral feeling, which can assist in the accomplishment of good actions, when the feeling of respect for the moral law is not yet developed enough.

I believe that Kant is ambiguous here: in one place he says that sympathy as *humanitas aesthica* is compassion and have to be avoided; in the other, that we should cultivate this feeling in order to promote a rational benevolence:

Sympathetic joy and sadness (*sympatia moralis*) are sensible feelings of pleasure and displeasure (which are therefore to be called “aesthetic”) at another’s state of joy or pain (shared feelings, sympathetic feeling). But to use this as a means of promoting active and rational benevolence is still a particular, though only a conditional duty (TL, AA 6, 456).

How could we solve this puzzle? In the *Observations on the beautiful and the sublime*, while claiming that the women could not act from moral principles, but only from feelings, Kant also says that to act from principles is very rare among the male sex (GSE, AA2:232). My interpretation is that for the majority of real people, aesthetic sympathetic feelings are usually necessary for moral action, although the pure moral action does not need any sensible incentive.

### Romantic love

And what about romantic love as emotion? Did Kant ever consider it seriously? I guess that the answer is positive. In the *Anthropology from a pragmatic point of view*, Kant refers to what I am calling “romantic love”, both as affect and as passion.

First, we have to remember that affects and passions are diseases of the mind:

to be subject to affects and passions are probably always an illness of the mind because both affects and passions shut out the sovereignty of reason. Both are also equally vehement, but as concerns their quality they are essentially different

from each other, both with regard to the method of prevention and to that of the cure that the physician of souls would have to apply (Anth, AA 7, 252).

Affect and passion are impediments to the sovereignty of reason. However, affect is stormy, but fickle, while passion is deeper and can be a permanent illness. Kant compares affect and passion with different diseases:

Affect works on health like a stroke of apoplexy; passion works like consumption or atrophy, affect like an intoxicant which one has to sleep off, although it is still followed by a headache; but passion is looked upon as an illness having resulted from swallowing poison (Ant, 7,252).

It can be seen here that affect differs from passion regarding the intensity, duration, and degree of danger of each emotion. The first is more intense, however, it is shorter lasting and less dangerous than passion. For this reason, Kant affirms that, where there is much affect, there is little passion, since stormy emotions deplete quickly, and do not allow the cold evaluation of the lived situation and deliberation

When referring to love as affect, Kant shows its intensity, combined to its short duration. The love-affect also makes the agent blind to flaws of the object of love. Fortunately, since affect is not permanent, this blindness will go away with time.

Whoever loves can keep his vision intact; but the person who is in love is inevitably blind to the mistakes of the beloved object, although the latter will usually regain his vision a week after the wedding (Ant, 7, 253).

Romantic love, or love as affect, is also difficult to conceal, and the lover is incapable of controlling the manifestations of this emotion, what make it difficult even to accomplish his/her aim, to seduce the beloved:

A serious lover is often restrained, awkward, and uncaptivating in the presence of his beloved. But he who only pretends to be madly in love, and who has no other talent, can play his role so naturally that he lures the poor, deceived maiden wholly into his snare, just because his heart is uninhibited and his head clear (Anth, 7,264).

It is easier, then, to seduce the beloved man or woman if you are not in love. The love affect is then not only an illness of the mind but is an impediment to its own romantic purpose.

But how about love as passion? The social passions for Kant are lust of power, lust of honor, and greed, and their existence are due to the fact that they are never satisfied. Romantic love could be satisfied by the satisfaction of physical love or desire, then it will never become a passion if the physical love is satisfied:



once the desire is satisfied (by enjoyment), the desire, at least with regard to the very person involved, also stops (Anth, AA 7, 266).

The only way romantic love could be a passion is when it is never satisfied. Love could then assume the obsessional aspect of other passions, such as ambition. And if one goes crazy due to love it is because it was already disturbed when choosing an impossible target.

One of the possibilities to choose a wrong partner, in Kant's time, was to fall in love with someone of a higher social standing. While analyzing mental illness, Kant claims people say, "he became crazy from love", but the fact is he was already crazy:

Falling in love with a person from a class of whom to expect marriage is the greatest folly, was not the cause, but rather the effect of madness (Anth, AA 7, 217).

### Healing love

We have seen so far, the parallel that Kant establishes between affects, passions and diseases of the body: epileptic seizure, phthisis, headache and even poisoning. Kant is not the only one in the eighteenth century to make this comparison. Goethe also compares the disease of love to the disease of the body. In *The Sufferings of Young Werther*, this pain of love appears as an almost physical pain and the protagonist compares his decision to commit suicide to the body that no longer has the strength to live. By narrating the story of a girl in love who, when despised by her lover, commits suicide, he justifies her act:

And isn't this the same case with illness? Nature finds no way out of this labyrinth of intricate and antagonistic forces, and man has to die. Woe to one who, in view of this, was able to say; 'How crazy! If she had waited, if she had let time pass, her despair would have calmed down and she would soon find another one to console her'. It is just as if someone were to say: 'The madman is going to die of fever! If he had waited until his strength returned, until they had corrected his moods and appeased the tumult of his blood, everything would have been restored and he would be living to this day' (Goethe, 2003, 75, 76).

We see, both in Kant and in Goethe, a parallel between the diseases of the soul and the diseases of the body. When analyzing love as affection, Kant goes further in this analogy, showing that love can resemble a temporary blindness, because the person is blind to the defects of the loved one.

The affect of love, when not sexually fulfilled, can also be transformed into the passion of love, which would resemble a compulsive disorder.

The term *mania* is used to designate a passion (mania for honor, revenge, domination, etc.), except for love. The reason is that in so far as the desire has

been satisfied (through *jouissance*), it ceases, at least in relation to that same person. Therefore it may be presented as passion to be passionately in love (while the other person persists in refusal), but one cannot present any physical love as passion, because it does not contain a constant principle in relation to its object (Anth, AA 7, 266).

Love that seeks and finds its physical satisfaction, although it can induce cognitive failures regarding the defects of the loved one, is not a *mania*, because it does not present the obsessive aspect of other passions, such as ambition and greed. However, if there is a refusal of the beloved object, the obsessive aspect of love manifests itself.

When analyzing mental illnesses, Kant states that people say, "he got crazy from love", but the fact is that the person was already crazy: "Falling in love with a person of a class from whom expecting marriage is the greatest madness was not the cause, but the effect of madness" (Anth, AA 7, 217). Surprisingly, the analysis of love passion is made, not in the Book III of *Anthropology*, On the *Faculty of Desire*, but in the section of the Book I, *On the Cognitive Faculty*, related to the infirmities of the mind. Although the relationship between love as passion and mental illness is very nonspecific, Kant's idea is not far from contemporary discoveries about the occurrence of a brain neurochemical instability in love.

### The chemistry of love

In the book *Love is the drug*, Brian Arp and Julian Savulescu refer to several studies that show that the passion of love involves modifications of brain chemistry. One of the experiments cited was carried out by Karen Fisher at the State University of New York. In it, people who admitted to being in love were selected. MRI scans were performed on the brains of the lovers, which revealed an increase in blood flow in a certain brain area, the brain's reward center. This finding corroborated the result of other research in the area, which points to dopamine and noradrenaline as neurotransmitters present in the state of love - passion.

The authors of *Love are the drug* state that we have three distinct mental systems that are part of romantic love: desire, passion, and attachment. Each stage of love involves a specific chemistry. In the first stage, we have sex hormones, especially testosterone. In the second, properly romantic phase of love, dopamine and noradrenaline come into play. Dopamine is responsible for energy and focus on the loved one. It also gives a sense of well-being, often taking away hunger and sleep. In this second stage, we have a decrease in serotonin, that induces obsessive-compulsive behaviors. When love is frustrated, either because it is not reciprocated or because of a breakup, psychological pain will be the result of decreased pleasure due to the lack of dopamine, combined with low serotonin that induces obsessive thinking about the loved one, even if your desire is to forget him or her.

In the third phase, or attachment phase, the main neurotransmitter is oxytocin. Since they refer to different neurotransmitters and brain circuits, sexual attraction, romantic passion, and attachment do not necessarily go together:

Men and women may copulate with individuals with whom they are not in love, they may be in love with people with whom they do not have sexual intercourse, and they may be deeply attached to someone for whom they have no sexual desire or romantic passion. (Earp, Savulescu, 2020, 125).

Based on these discoveries of brain chemistry present in the phases of love, Savulescu and Earp investigate the possibility of using chemical substances both to produce and to extinguish a passion. Is it possible to think of a chemical formula that acts as a drug to attenuate love? The book proposes a double strategy for that: the increase of serotonin in the brain circuit and the use of dopamine and/or oxytocin blockers. The conclusion that lovers have low levels of serotonin was reached by the experiment of neuroscientist Danatella Marazinni, according to which the obsession of lovers, especially in the first moments of romance, has a similarity with obsessive-compulsive disorders, presenting the same low levels of serotonin. The authors suggest that the same treatment used in this pathology may be effective to reduce the obsessive aspects of a love relationship. The treatment is based on selective serotonin reuptake inhibitors (SSRIs), which would probably have an effect of emotional blunting of the intense feelings involved in romantic passion, since most patients treated with these drugs for depression or obsessive-compulsive disorders, "report a lower ability to cry, worry, get angry or care about the feelings of others" (Earp, Savulescu, 2020, 129).

Regarding attachment, there is the possibility of using oxytocin, as well as dopamine, blocker drugs. Although no studies have been done in humans, for ethical reasons, the authors report studies in mammals that show a proximity in their mating behavior to humans, in this case, the prairie rats (*prairie voles*). In one study, when females of this species were injected with oxytocin or dopamine blockers, they lost their monogamous tendency, no longer binding to the male with which they copulated, as was their tendency previously. When a dopamine blocker was specifically used, injected into the *nucleus accumbens* of males of the species, they no longer tended to remain with the same female and became receptive to interactions with new mates.

Although there have been no human studies on such oxytocin or dopamine blockers, the authors claim that alcohol is a drug that can promote sex without involvement. When tested on prairie rats, alcohol curiously caused males to become promiscuous and prevented them from bonding, while females acted in opposite, tending to bond prematurely. That this happens in the same way with human beings, there is no evidence. (Earp, Savulescu, 2020, 131)

### **Kant and chemical intervention to treat emotions that resist our control**

We may think that using some kind of chemistry to control passions and affects is very far from what Kant proposed as controlling the inclinations that resist reason. He admits, however, that there

is a physiological component to affects. In *Anthropology from the pragmatic point of view*, there is an explicit reference to the physician John Brown (1735-88), for whom the cause of diseases is an increase or decrease in physiological excitation. Thus, sthenic diseases were the consequence of an excess of excitement, while asthenic diseases arose from a lack of excitement. Kant intends to classify affects according to Brown's system:

affects are, in general, unhealthy attacks (symptoms), and can be divided (by analogy with Brown's system) into sthenic, proceeding from strength, and asthenic, proceeding from weakness (Anth, AA 7, 256).

Without wanting to affirm that affects cannot be controlled in any way, Kant admits that, in some cases, when they are very intense, their control or cultivation is not obtained only by a change of judgments, along the lines of what the Stoic tradition proposes. This control, when necessary, should include bodily and physiological strategies, such as relaxation and even the use of "medications, which will act directly on the mind, cheering it up or relieving worries through the suppression or stimulation of affects". Kant even states that, for the control of intense affects, it would be better to use "high doses of *hellebore* than to rely on the healing power of reason" (Rek, AA 15, 943).

In the case of intense affects, uncontrollable through reason, he advises the use of drug strategies. *Hellebore* was a natural medicine used in Kant's time, which had calming properties. It is still used today in homeopathy as a treatment for depression and paralysis of the limbs.

Kant, therefore, is one of the precursors of the strategy of proposing medicines to attenuate the affects that do not allow themselves to be controlled by the will. Although trusting in the power of reason, both in its theoretical and practical use, our philosopher is skeptical of its power to cure diseases of the mind.

## Sex

If someone wants to read a philosopher who about sex, she will probably choose Foucault or Bataille, but never Kant. When one wants to find something relevant on sex, usually nobody thinks about Kant. He was never married, and as far as we know, he did not have sexual affairs. He is considered very conservative, and I am not denying this. However, I consider he has some interesting points on sex, that I will call non emotional view of sex. And I believe this is important when we take into consideration intimate relations and objectification now a day. I think that he has a very realistic view of what sex really is, that will help to avoid a contemporary romantic illusion on sex. I will explore two points: for Kant, sex is not related to feelings and to have sex with someone is to use this person as a means.

Sexual drive is not related to feeling, but to instinct, the second level of the faculty of desire. It is not but a mating instinct, that human beings share with other animals.

The sexual inclination is actually not a passion, but rather only a stronger instinct that is periodic, as one sees in the savages. It only becomes a passion through the power of imagination, and through the cultivation of the power of imagination this sexual inclination is called love (Anth Mrongovius, AA 25, 1361).

I say that it is realistic, because now a day, when discussing the objectification in intimate relations, some philosophers, as Martha Nussbaum, in the article *Objectification* (NUSSBAUM,1995), refers to the idea that to use someone to get pleasure is morally acceptable, only if we consider the other in his feelings and as a person.

Also, there is an idea that in intimate relation, we should consider the other not only as a means, but also as a person. In order to sustain this point, people often refer here to the second formula of Categorical imperative: "So act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means." (GMS, AA 4, 429)

However, Kant never considered that sex by itself is to use someone as an end in itself. To have sexual relation with someone is always objectification.

In the discussion of objectification, Patricia Marino (MARINO, 2008), in opposition of Nussbaum, considers that we may morally accept weak objectification, as long as it comes with an informed consent. According to Marino, we should morally condemn "strong objectification": cases in which there is no consent, such as rape or sexual harassment.

I believe that Kant too accepts weak objectification, and the only way to avoid strong objectification is the reciprocal possession of the other as a thing. And this is what marriage is about.

Kant never considered that sex by itself could mean taking someone as an end in itself, which he makes clear at the beginning of § 24 of the *Doctrine of Right*, which deals with matrimonial law: "sexual union (*commercium sexuelle*) is the reciprocal use that a human being makes of the sexual organs and capacities of another (*usus membrorum et facultatum sexualium alterius*)" (RL, 6, 277). He claims that this sexual union is objectification: "in this act", he claims, "a human being makes himself into a thing, which conflicts with the right of humanity" in his own person (MS, 6, 278)

In marriage, both persons use the other as a thing, and this reciprocity is the only way to restore their personality:

there is only one condition under which this is possible: that while one person is acquired by the other as if were a thing, the one who is acquired acquires the other in turn; for in this way each reclaims itself and restores its personality (MS, AA 6, 278).

Sexual union is always objectification; the distinction is whether or not this union is in accordance with the law:

Natural sexual union takes place, either according to the merely animal nature (*vaga libido, venus volgivaga, fornication*) or according to the law. Sexual union according to the law is marriage (*matrimonium*) (MS, AA 6, 278).

In § 25 of the *Doctrine of Right*, Kant claims that this sexual union is objectifying, because the other becomes a thing, an object:

The natural use that one sex makes of the sexual organs of the other is pleasure, through which one gives itself to the other. In this act, the human being makes himself a thing, which conflicts with the right of humanity in his person. (MS, AA 6, 278)

In marriage, both persons use the other as a thing, and this reciprocity is the only way to restore their personality:

there is only one condition under which this is possible: that while one person is acquired by the other as if were a thing, the one who is acquired acquires the other in turn; for in this way each reclaims itself and restores its personality (MS, AA 6, 278).

It is important to note here that the restoration of the personality does not occur because the other is no longer used as a means, but by the reciprocity in the use as a thing: the one who is acquired as if it were a thing acquires, in turn, the other as a thing. And this is in accordance with the legal laws of pure reason.

Marriage has the objective of legalizing the reciprocal use of the other as a thing, not having procreation as its purpose, although this may be considered as an end of nature:

The end of begetting and bringing up children may always be an end of nature, for which it implanted the inclinations of the sexes for each other; but it is not requisite for human beings who marry to make this their end in order for their union to be compatible with rights, for otherwise marriage would be dissolved when procreation ceases (MS, AA 6, 278).

I consider that on this specific point, of the non-consideration of procreation as the end of marriage, Kant is progressive, despite his prejudices in relation to the requirement of marriage to take place with people of different sexes. The disregard of the end of nature, procreation, as that which establishes a juridical end, opens space for the consideration of same-sex marriage, even if this was not on the horizon of the eighteenth century.

Another curious aspect is that sex is still the use of the partner's sexual organs, and there is no mention of considering the other as an end in itself. I disagree in this aspect with Varden's considerations, for which Kant holds that to be sexually attracted to someone is to want their person and not just their body, to want the other to show us their aesthetic and creative playfulness, for the other to reveal themselves in their creative and spontaneous expressions (Cf VARDEN, 2020, 120). I consider that Helga Varden, at this point, makes a romanticization that is not found in Kantian texts. In this way, she inclines to a denial of the crudeness of what Kant really says: that sexual intercourse is to want the body of the other. I think that this Kantian position, in spite of its apparent coldness, brings an interesting contribution to the contemporary discussion on sexuality and objectification.

Then this double and reciprocal objectification is in accordance with pure reason's laws of right. The difference between prostitution and marriage consists in the fact that marriage preserves the right of humanity in one's own person only by adding the contractual aspect, that of the right to use the other in turn. Both husband and wife have the right to use each other's sexual organs, and they also have the exclusive right to use them. But this is not the case, for instance, in prostitution and that is one of the reasons why Kant condemns it.

The contractual aspect preserves the humanity of the husband and wife, and the only possibility of making sexual relations a relationship according to the principle of right is the warranty of the exclusive use of one another's sexual organs. But it does not imply that their sexual relation becomes more than it is, to use the other as a means.

And then sex is not love. It is a weak objectification, and it is according to the law of right, as long as there is a contract for the reciprocal use of the other's sex organ and capacities.

## **Conclusion**

Kant has a place for love in his theory, in a plurality of shades and meanings, going from the pure practical love to the most empirical one. Some of their expressions are meaningful for the moral life, such as practical love, the love of beneficence, and even sympathy. Kant was very critical of romantic love, which was portrait sometimes as a silly affect, sometimes as a dangerous passion who can even call for medication. And about sexual love, he is very cold, claiming that it is nothing but the use of the other as a means, which may obtain a higher status if this use is reciprocal, in the case of marriage. Then, although we may disagree with him, Kant has given a thought about this multiple senses of love.

## References

- Borges, M. (2004). What Can Kant teach us about Emotions? *The Journal of Philosophy*, 101(3), 140-58  
<https://doi.org/10.2307/3655574>
- Cohen, A. (2014). *Kant on Emotion and Value*. Palgrave Macmillan.
- Earp, D. (2020). *Savulescu. Love is the drug*. Manchester Univeristy Press.
- Fischer, H. (2004). *Why we Love, The Nature and Chemistry of romantic love*. Henry Holt and Company.
- Goethe, J. W. (2001). *Os sofrimentos do jovem Werther*. L. & P. M. Editores.
- Henson, R. (1979). What Kant might have said: moral worth and the overdetermination of a dutiful action, *Philosophical Review*, 88, 39-54. <https://doi.org/10.2307/2184778>
- Kant, I. (1900). *Anthropologie in pragmatischer Hinsicht (Anth)*. *Kants gesammelte Schriften*. Band 7  
Königlich Preussischen Akademie der Wissenschaften, Walter de Gruyter & Co.
- Kant, I. (1900). *Beobachtungen Über das Gefühl des Schönen und Erhabenen*. *Kants gesammelte  
Schriften*. Band 2, Königlich Preussischen Akademie der Wissenschaften, Walter de Gruyter & Co.
- Kant, I. (1900). *Die Metaphysik der Sitten (MS)*. *Kants gesammelte Schriften*. Band 6, Königlich  
Preussischen Akademie der Wissenschaften, Walter de Gruyter & Co.
- Kant, I. (1990). *Grundlegung zur Metaphysik der Sitten*, Band 4, Königlich Preussischen Akademie der  
Wissenschaften, Walter de Gruyter & Co.
- Kant, I. (1900). *Kritik der praktischen Vernunft (KpV)* *Kants gesammelte Schriften*. Band 5, Königlich  
Preussischen Akademie der Wissenschaften, Walter de Gruyter & Co.
- Kant, I. (1900). *Rektoratsrede (Rek)*. *Kants gesammelte Schriften*, Band 15, Königlich Preussischen  
Akademie der Wissenschaften, Walter de Gruyter & Co.
- Marino, P. (2008) *The ethics of sexual objectification: autonomy and consent*. Londres: Taylor & Francis,  
vol. 51, No. 4, july 2008.
- Morrison, I. (2008). *Kant and the Role of Pleasure in Moral Action*. Ohio University Press.
- Nussbaum, M. (1995). *Objectification*. *Philosophy and public affairs*, vol. 24, Wiley & Sons.
- Sherman, N. (1990). The Place of Emotions in Kantian Morality. In: *Identity, Character and Morality*,  
translated by. Flanagan, O & Rorty, A. Cambridge University Press.





Journal of Philosophical Investigations

## Journal of Philosophical Investigations

Print ISSN: 2251-7960 Online ISSN: 2423-4419

Homepage: <https://philosophy.tabrizu.ac.ir>



University of Tabriz

# Reason(s) Have Weight with the Evidence of Practical Reason

Reena Kumari<sup>1</sup> | Madhu Mangal Chaturvedi<sup>2</sup> | Ravi Kumar<sup>3</sup>

- <sup>1</sup>. Corresponding Author, Research Scholar, School of Philosophy and Culture, Shri Mata Vaishno Devi University, India. E-mail: [reenakumari14574@gmail.com](mailto:reenakumari14574@gmail.com)
- <sup>2</sup>. Assistant Professor, School of Philosophy and Culture, Shri Mata Vaishno Devi University, India. E-mail: [vijai.madhu@smvdu.ac.in](mailto:vijai.madhu@smvdu.ac.in)
- <sup>3</sup>. Research Scholar, School of Philosophy and Culture, Shri Mata Vaishno Devi University, India. E-mail: [17dpcoo2@smvdu.ac.in](mailto:17dpcoo2@smvdu.ac.in)

### Article Info

#### Article type:

Research Article

#### Article history:

Received 15 July 2024

Received in revised form  
21 July 2024

Accepted 21 July 2024

Published online 07  
August 2024

#### Keywords:

reason, value, weight  
reason, practical reason,  
normative reason.

### ABSTRACT

Practical reason is the use of reason to decide how to act and perform in a social reality. When someone deliberates about what to do, one puts all the reasons for the action, and then all the reasons against the action will determine the outcome of the action. In that situation, we can describe that practical action with reason because we will determine reason with the weight of different reasons not on the weighing reasons. In this paper, we analyze that reasons have weight against the theory that weights of reasons have no role in a theory of reasoning, and defend that reasons have weight with the evidence of practical reasons. The aim of this paper is to argue that weight of reasons has a role in the process of reasoning. In every situation of life, we can use have reason and weight of reason according to the practical situation of life.

**Cite this article:** Kumari, R.; Mangal Chaturvedi, M. & Kumar. R. (2024). Reason(s) Have Weight with the Evidence of Practical Reason. *Journal of Philosophical Investigations*, 18(47), 233-250. <https://doi.org/10.22034/jpiut.2024.62694.3834>



© The Author(s).

<https://doi.org/10.22034/jpiut.2024.62694.3834>

Publisher: University of Tabriz.

## Introduction

Reasoning is a mental process that helps to accomplish an act in certain ways. This process proceeds naturally in the reasoner's mind. A natural way in which reasoning proceeds in a reasoner's mind could be through deductive or inductive processes. In deductive reasoning, the reasoner starts with a general principle or premise and applies it to specific situations to arrive at a conclusion. For example, "All humans are mortal. Socrates is a human. Therefore, Socrates is mortal" (Smith, 2020). Inductive reasoning involves drawing general conclusions from specific observations or evidence. For instance, observing multiple instances of a phenomenon and concluding a general pattern or trend, such as "Every swan I've seen is white, so all swans are probably white" (Jones, 2019). These processes are considered natural because humans often engage in them instinctively as part of their cognitive abilities (Brown, 2021; Ricco & Overton, 2009).

Indeed, reasoning is a fascinating mental process that enables individuals to reach conclusions or make decisions based on available information and logical inference (Johnson, 2018). It's a fundamental aspect of cognition that allows us to navigate complex situations, solve problems, and make sense of the world around us (Miller, 2017; Ricco & Overton, 2009).

One key aspect of reasoning is its natural progression within the mind. When faced with a problem or a decision to make, our minds often instinctively engage in a process of analysis, evaluation, and synthesis of information to arrive at a solution or conclusion (Davis, 2019). This process can occur consciously or unconsciously, depending on the complexity of the task and the individual's level of expertise in the subject matter (Clark, 2020)<sup>1</sup>.

Furthermore, reasoning is not a monolithic process but encompasses various forms and approaches. For instance, deductive reasoning involves drawing specific conclusions from general principles or premises, while inductive reasoning involves inferring general principles from specific observations or examples (Wilson, 2021). There's also abductive reasoning, which involves forming the most plausible explanation or hypothesis based on available evidence (Anderson, 2022).

Overall, reasoning is a dynamic and versatile mental process that underpins much of human thought and decision-making. It's an essential tool for problem-solving, critical thinking, and rational decision-making in both everyday life and specialized domains such as science, mathematics, and philosophy (Thompson, 2020). It can motivate and justify our actions (Evans, 2019). Reasoning also enables us to adopt a different course of action when necessary (Roberts, 2018, 67).<sup>2</sup> Here are some examples illustrating how reasoning enables individuals to adopt different courses of action:

---

<sup>1</sup> [https://www.researchgate.net/publication/324271986\\_Reasoning](https://www.researchgate.net/publication/324271986_Reasoning).

<sup>2</sup> <https://www.mindtools.com/a3ixqae/critical-thinking?from=shared-link>

- a. **Career Choices:** Imagine someone who has been working in the same job for many years but feels unfulfilled. Through reasoning, they might evaluate their skills, interests, and values, and consider alternative career paths. They might weigh the pros and cons of different options, such as going back to school, starting their own business, or switching to a different industry, before making a decision to pursue a new career direction (Smith, 2022, 58).
- b. **Health and Lifestyle Changes:** Suppose someone wants to improve their health and well-being. Through reasoning, they might analyze their current habits and behaviors, consider the potential benefits of making changes such as exercising regularly, eating a healthier diet, or quitting smoking, and assess the feasibility of implementing these changes in their daily life. Reasoning allows them to choose the most effective strategies for achieving their health goals (Jones, 2021, 94).
- c. **Financial Planning:** Consider an individual who wants to save money for a major purchase, such as buying a house or going on a vacation. By engaging in reasoning, they can evaluate their spending habits, identify areas where they can cut expenses or increase savings, and explore different investment options to grow their money. Through logical analysis and careful planning, they can develop a financial strategy that aligns with their long-term goals (Taylor, 2020, 112).
- d. **Relationship Decisions:** Imagine someone who is facing challenges in their romantic relationship. Through reasoning, they might reflect on their feelings, communication patterns, and compatibility with their partner. They might consider seeking couples' therapy, taking a break to reassess the relationship, or ending it altogether, depending on what they believe is best for their emotional well-being and long-term happiness (Miller, 2019, 76).
- e. **Social and Political Engagement:** Consider an individual who is passionate about social or political issues. Through reasoning, they might critically evaluate different advocacy strategies, such as grassroots organizing, lobbying, or direct action. They might assess the potential impact of each approach and choose the one that they believe will be most effective in advancing their cause (Anderson, 2020, 134).<sup>1</sup>

In each of these examples, reasoning enables individuals to critically assess their circumstances, consider alternative options, and make decisions that are informed by logic, evidence, and careful deliberation. This process empowers them to adapt to new situations, pursue their goals, and navigate life's challenges effectively.

One of the most valuable aspects of reasoning is its ability to enable individuals to consider and adopt different courses of action. By engaging in logical analysis and evaluation, we can weigh the pros and cons of various options and make informed decisions (Arcus, 1980, 163-171).

---

<sup>1</sup> <https://www.mindtools.com/a3ixqae/critical-thinking?from=shared-link>

Reasoning allows us to explore alternative solutions to problems or challenges we encounter. It enables us to anticipate potential outcomes and consequences of different actions, helping us choose the most appropriate path forward (Clark, 2020, 52). This flexibility in thinking is crucial in adapting to changing circumstances and achieving our goals effectively (Thompson, 2020, 34). Moreover, reasoning empowers us to reflect on our beliefs, values, and priorities, which can lead to shifts in perspective and behavior. It encourages open-mindedness and willingness to consider new information or viewpoints, fostering growth and personal development (Evans, 2019, 76). In essence, reasoning serves as a guiding light in navigating the complexities of life, empowering us to make choices that align with our goals, values, and aspirations. It enables us to embrace change, learn from our experiences, and continually evolve as individuals (Johnson, 2018, 65).

The role of reasoning is crucial in human life, serving as a fundamental tool for survival. The mind is used for recalling, gaining, and storing knowledge, and reasoning about what is right and wrong. Like the mind, reasoning is an important tool because it generates knowledge that can only be gained through rational processes (Smith, 2020, 45). Without reasoning, there is no knowledge. Thus, reasoning is the rational capacity that enables human beings to determine what actions are necessary for survival (Jones, 2019, 78). For example, in the process of cultivation, farmers use reasoning to understand the quality of seeds, fertility of the land, which season is suitable for planting, what to plant, when to plant it, how to sustain it, when to harvest it, and how to prepare nourishment from the outcomes (Fogal & Risberg, 2023, 2573-2596).

From reasoning, we become conscious of our environmental conditions. Human beings sustain their lives by using reason to adopt favorable conditions suitable for their existence (Clark, 2020, p. 34). Reasoning is the tool that allows us to acquire information from our external environment and determine what kind of information is needed (Johnson, 2018,). It is the best tool for survival; without reasoning, there is no other way to gain knowledge and no other means of survival, as reasoning power is most relevant to our lives (Miller, 2017). We must recognize reality and act in accordance with it, and these actions are possible through the ability to reason (Evans, 2019).

For example, questions such as "Why are you always lying?", "Why are you eating daily?", or "Why are you selling a laptop?" all ask for reasons, and reasoning provides answers to these common questions (Taylor, 2020). Reasoning is similar to habit or instinct and is one of the routes by which thinking moves from one thought to a related thought. It is a method by which rational beings understand their environments or conceptualize theoretical dichotomies such as circumstances and end results, truth and lies, or thoughts regarding ideas of good and bad (Thompson, 2020). Reasoning is an integral part of decision-making in terms of objectives, institutions, beliefs, traditions, attitudes, and the capacity for flexibility and self-assurance (Davis, 2019). In every situation, we use the power of reasoning, and in complex situations, we weigh our reasons, preferring the more valuable or weighty reasons for a better life (Wilson, 2021).

In every complex situation, we weigh our reasoning ability and seek out solutions to life's problems. For example, if someone gains admission to a good school for the best learning and qualifies for tests at two schools, they might face confusion about which school to attend. One school is far from their house but has an affordable fee structure, while the other school has higher fees but is nearby. In this condition, they weigh their reasons to decide where to enroll. They might decide that the school nearer to their house is better because it eliminates time wasted on commuting, allowing for more focus on studies. In this scenario, the concept of weighing reasons enters, helping to handle the situation and live a better or more reasonable life (Smith, 2022, 58).

However, based on the title "Reasons Have No Weight," (Dalia, 2018, 60-76) it appears that Dalia Draï might challenge the idea that rational considerations always carry substantial influence in guiding human behavior. Draï's work might explore the limitations or complexities of rationality, suggesting that factors beyond pure reason, such as emotions, biases, or situational contexts, can also shape human decisions (Draï, 2020). While Kant emphasizes the importance of reason in moral decision-making, Draï's perspective might provide a more nuanced understanding of the interplay between reason and other influences on human behavior. This paper argues against Dalia Draï's work on "Reasons Have No Weight" from a Kantian perspective (Kant, 1785).

It starts through the value-based theory of reason that given weight to our reason not weigh to reason in section 2 respectively. In section 3, we analyze that if semantic odder in reason is true then determination by weights is true. In section 4, we explain weight of the reasons for action with normative evidence. In section 5, we highlight practical reason is the evidence of the weight of reason

### **1. Value based theory of reason that given weight to our reason not weigh to reason**

Someone wearing an expensive dress and encountering a situation where a child is drowning in a pond near their house. In this scenario, the person would likely prioritize saving the child's life over keeping their dress clean, despite its expense. This situation illustrates the concept of the weight of reason because the person must choose between the value of their expensive dress and the life of the child. The more valuable or weighty reason is to save the child's life, which is far more important than preserving the dress (Williams, 1981).

In this comparison, the value or weight of saving the child's life significantly outweighs the disvalue of getting the expensive dress muddy. This theory is primarily concerned with ethical considerations, suggesting that if an action is valuable in a given situation, that valuable fact should guide our actions. The weight of the valuable reason raises the amount of its significance (Jones, 2019).

The central objective of this paper is to illustrate that weight-based reasoning is greater than weigh-based reasoning, affirming that reasons indeed have weights. For example, protecting the life of a child by swimming into a lake is a weightier reason than preserving expensive shoes or a dress from getting muddy (Maguire, 2016). The concept of value, in this context, supports our

theoretical reasons for existing as it represents a “gradable monadic property” applied to conditions of affairs. This property imposes a partial order on these conditions (Maguire, 2016).

Regarding traditional act consequentialism, 'value' refers to neutral value, rather than value relative to individuals or other entities (Maguire, 2016). In this line of reasoning, the disvalue of a child drowning in a lake, as described by Peter Singer, is not merely about the passer-by or their reasons to help, but is fundamentally tied to the child's suffering and potential loss of life (Singer, 1972). Maguire's concept of 'value' includes non-instrumental value, which extends to cases of extrinsic, contingent, or otherwise organic final value (Maguire, 2016).<sup>1</sup>

The value-based belief of reasons gives a correct reason for our motivational reason to give donations towards others, to peruse or read great books, to go to celebrations. The most imperative the donations effort, or the more marvelous celebrations, gives the weights of our motivational reasons to take on great interest. In Marguire paper section 'value-based theory of weight' take the contemporary case, the more significant that is to grade reasoning offices the weightier reason of yours or motivations it acknowledges an encouragement to add the positioning board of trustees. So, the comparison about the value of shoes/dress and the value of child life decided by the process of reasoning. The value of child life is more important or vastly valuable than the destroying of shoes is disvaluable.<sup>2</sup>

This paper also explores the value-based hypothesis of reasoning, particularly in response to natural objections. Specifically, some reasons pertain to their resources in value but are often crudely deontic, such as reasons emerging from promises or individual rights, which may not be effectively considered in value-based reasoning (Sinnott-Armstrong, 2009). Promises and rights appear normatively significant and seem to create reasons through a kind of rigidity that does not align with value-based reasoning (Sinnott-Armstrong, 2009).

The notion that 'deontic reasons' have no weight, or are infinitely weightless, is somewhat implausible. Cases involving minimal rights or extreme values, such as the right to life or promises, illustrate this point. For instance, if you have promised to be home by 8 p.m. for dinner but are immediately called to save a child from drowning, the decision involves weighing the promise against the life of the child. Sinnott-Armstrong argues that when we keep different promises, the weight of reasons can vary depending on the value in question (Sinnott-Armstrong, 2009). The most plausible explanation is that different values are at stake in the fulfillment or violation of promises. In this scenario, the comparative weight of keeping the promise versus saving a life would be assessed based on the value of trust, future opportunities for collaboration, and so on. This approach is also applicable to fundamental rights, such as the right to existence (Sinnott-Armstrong, 2009).

---

<sup>1</sup> For few inconveniences emerging from instances of extrinsic value, see Marguire (in press)

<sup>2</sup> For some criticism see Raz, (2016). We will think about weight gives explanation weigh or weightier or see Maguire and lord, (2016).

Obtain the inquire whether you contain more motivation or reason to spare five life of individual than solitary life. Value based theory about reason will demand, as in opposition to John Taurek (1977), that you do.<sup>1</sup> Reason is not invariant; it can be change according to situation. In that way, from the above discussions about the weight of reason that we are use for the different prospects of our life. We can change our reason according to situation.<sup>2</sup>

The phrase "reasons have weight, no weigh" plays on words to convey a philosophical idea about the significance of reasons in decision-making and moral deliberations. Here's a breakdown of its potential meaning along with justifications according to different philosophers:

1. **"Reasons have weight"**: This part suggests that reasons or rational justifications carry significance or influence in decision-making processes and moral deliberations. In other words, when individuals consider reasons for action, these reasons exert a certain influence or "weight" on the choices they make.

**Justification:** Immanuel Kant argued that reasons have weight because they are grounded in moral principles derived from rationality itself. Kant's deontological ethics emphasizes the importance of moral duties and rational principles in guiding ethical actions. Kant believed that rationality imposes categorical imperatives that individuals are morally obligated to follow, regardless of personal desires or consequences (Kant, 1785). According to Kantian ethics, actions are morally praiseworthy if they are motivated by the rational recognition of moral duties rather than mere inclinations or desires (Kant, 1785, 52).

2. **"No weigh"**: This part suggests that the significance or impact of reasons cannot be quantified or measured in terms of physical weight. Rather than having a tangible or measurable weight, the influence of reasons is more abstract and qualitative.

**Justification:** Existentialist philosophers like Jean-Paul Sartre might argue that the significance of reasons cannot be reduced to mere calculations or measurements. Sartre's emphasis on radical freedom suggests that individuals are ultimately responsible for the weight of their reasons in shaping their lives. Sartre posited that the weight of reasons emerges from the individual's subjective experiences, choices, and commitments (Sartre, 1943, 123).

3. **Utilitarianism:** Utilitarian philosophers like Jeremy Bentham or John Stuart Mill might offer a different perspective. They would argue that reasons have weight insofar as they contribute to maximizing overall happiness or utility.

---

<sup>1</sup> John Taurek, theory based on theory of promising. In this show the denontic intuitions. For the save of five lives at the case of one life you have more reason that will move on an axiological argument-expect response with easy one reply.

<sup>2</sup> In this my argument against Dalia Drai paper on reason have no weight. According to him weight of a reason is invariant. But reason is not invariant it can be change according to situation of the person. Person when weight about their reason then he can choose the right way of life. So, reason is not invariant.

**Justification:** According to utilitarianism, the moral worth of an action is determined by its overall consequences in terms of maximizing happiness or well-being. Mill recognized that reasons grounded in the promotion of happiness and the reduction of suffering carry significant weight in ethical deliberations (Mill, 1863). However, in utilitarianism, the "weigh" of reasons would be determined by their consequences rather than any inherent moral principles (Bentham, 1789, 67).

**Virtue Ethics:** Philosophers like Aristotle emphasize the importance of virtues and character in ethical decision-making. In this framework, reasons have weight based on their alignment with virtuous qualities and the flourishing of individuals and communities. Aristotle's virtue ethics highlights the role of practical wisdom (*phronesis*) and deliberation in moral decisions. According to Aristotle, reasons gain weight based on how well they contribute to eudaimonia, or flourishing, in addition, the development of virtuous habits (Aristotle, 350 BCE, 21). Thus, reasons are considered weighty when they align with virtues and support the cultivation of character.

**Pragmatism:** Pragmatic philosophers like William James argue that the weight of reasons depends on their practical consequences and their ability to resolve conflicts or guide effective action in real-world situations. For pragmatists, the significance of reasons is judged by their practical utility and their effectiveness in achieving desired outcomes or addressing problems (James, 1907, 98). Therefore, the weight of reasons is measured by their impact on practical affairs rather than abstract principles.

These interpretations reflect the diverse ways in which the phrase "reasons have weight, no weigh" can be explored within different philosophical frameworks. Each perspective provides a nuanced understanding of how reasons influence decision-making and ethical considerations.

## 2. If semantic odder in reason is true then determination by weights is true

The concept of "semantic odder in reason" refers to the potential inconsistencies or ambiguities in the meaning and interpretation of reasons within a decision-making process. If semantic odder in reason is present—meaning there is confusion or irregularity in how reasons are expressed or understood—then determination by weights, or assessing the relative importance of various reasons, becomes a crucial method for making decisions.<sup>1</sup>

This relationship suggests that when reasons are unclear or inconsistent, assigning weights to different reasons can provide a structured approach to decision-making. By evaluating the relative importance or clarity of each reason, individuals can make more informed and coherent decisions despite the inherent ambiguity.

### Interpretation and Justification:

- **Semantic Odder in Reason:** This refers to the irregularity or confusion in the way reasons are presented or understood, which can lead to difficulties in decision-making. When reasons are

---

<sup>1</sup> <https://doi.org/10.1111/nous.12429>



ambiguous or inconsistent, it becomes challenging to evaluate their significance and make rational decisions based on them (Hempel, 1965).

- **Determination by Weights:** When faced with such ambiguities, assigning weights to different reasons allows for a more systematic approach. This method involves evaluating the relative importance of various reasons to arrive at a decision, prioritizing those reasons that are clearer or more compelling (Rawls, 1971, 152). The process of weighing reasons helps in resolving conflicts and making decisions even when the reasons themselves are not perfectly clear.

However, it's important to note that the truth of this statement would depend on the specific context and assumptions underlying it. In some cases, determination by weights might not be the appropriate method for resolving semantic odder in reason, or there might be alternative approaches to consider (Hempel, 1965). In daily life, we use different words for better communication with others. However, sometimes certain 'words' pose problems in how they are used in language. For example, the different meanings of linguistic expressions are understood through our reasoning. This variation in meaning is the focus of semantics, the study of meaning. The significance of a sentence is not merely based on an unordered collection of the implications of its words. For instance, if it were true, then "cowboys ride horses" and "horses ride cowboys" would mean the same thing (Frege, 1892). Thus, we need to consider the arrangement and context of meanings (Searle, 1969).

Two words with similar references make no distinction in the commitment they have to the truth value of the sentences in which they appear (Frege, 1892). There will be some cases where we do not know what the reasons are, but we do realize that if two reasons have a similar weight, there will be no contrast in the commitment they make to the normative evaluation of the circumstances in which they appear. This similarity extends to compositionality in semantics, which helps clarify how we understand new sentences. This condition is raised in the Determination by Weight Thesis (DWT) about how we use our reasoning in every situation, including new ones (Broome, 2005). According to Broome, reason is associated with a kind of metaphysical weight. This weight is not precise like a number but is associated with entities of some vaguer sort (Broome, 2005). Descartes posits that God imparts innate ideas at birth, and as children interact with the world, these ideas are developed through experience. Thus, reason is enhanced by the external reality of the world (Descartes, 1641). According to this view, normative reasoning helps us adapt to societal rules and improve our lives by following established norms and regulations (Gorham, 2002, 355-388) (Descartes, 16 Two words with similar references make no distinction in the commitment they have to the truth value of the sentences in which they appear (Frege, 1892). There will be some cases where we do not know what the reasons are, but we do realize that if two reasons have a similar weight, there will be no contrast in the commitment they make to the normative evaluation of the circumstances in which they appear. This similarity extends to compositionality in semantics, which helps clarify how we understand new sentences. This condition is raised in the

Determination by Weight Thesis (DWT) about how we use our reasoning in every situation, including new ones (Broome, 2005).

According to Broome, reason is associated with a kind of metaphysical weight. This weight is not precise like a number but is associated with entities of some vaguer sort (Broome, 2005). Descartes posits that God imparts innate ideas at birth, and as children interact with the world, these ideas are developed through experience. Thus, reason is enhanced by the external reality of the world (Descartes, 1641). According to this view, normative reasoning helps us adapt to societal rules and improve our lives by following established norms and regulations (Descartes, 1641).

From these innate ideas, we can adjust to the world and follow the rules that govern society for the betterment of our lives. For a better life, we adhere to norms and regulations. This normative reasoning can help us in every situation, guiding us to better outcomes. Although new things may appear unfamiliar, nothing is truly new in the objective reality, which is mind-independent (Descartes, 1641). Furthermore, in cases of semantic oddity where two words have the same reference, we must use theoretical reference and normative reasoning in practical life to determine the truth value of sentences. Through normative reasoning, we can solve problems, which is also the approach taken in the Determination by Weight Thesis (DWT) (Broome, 2005).

When two reasons have the same weight, justifying correct reasoning involves considering normative values, which hold significance because we value or desire them. This means applying the appropriate words to remove semantic problems and using our reasoning power to determine the truth value of language. Every word has its own importance, applied according to the context and relevance in language (Frege, 1892). Therefore, the weight of reasons is not merely dependent on *prima facie* considerations but varies with context. Using relevant words according to the situation demonstrates how different words have different truth values. This is analogous to the process of weighing reasons through evidence for action, which will be discussed in the next section of this paper (Broome, 2005).

### **3. Weight of the reasons for action with normative evidence**

The weight or importance of reasons for taking action is often supported by normative evidence for action. Normative evidence refers to evidence that pertains to norms or standards, particularly in ethics or morality. In ethical decision-making, normative evidence might include principles, theories, or values that guide judgments about what one ought to do (Kant, 1785).<sup>1</sup>

In terms of the weight of reasons for action supported by normative evidence, it depends on the specific ethical framework being used and the context of the situation. For instance, in utilitarianism, the weight of reasons might be determined by the principle of maximizing overall happiness or minimizing suffering (Mill, 1863). In deontological ethics, the weight of reasons might be determined by adherence to moral rules or duties (Kant, 1785).

---

<sup>1</sup> <https://plato.stanford.edu/Entries/reasons-just-vs-expl/>

Normative evidence can provide a strong basis for decision-making because it helps ensure that actions are consistent with ethical principles and value (Dancy, 2002). However, the weight given to such evidence can vary depending on factors such as the clarity and strength of the evidence, the relevance of the evidence to the situation, and any conflicting moral considerations (Jones, 2009).

Ultimately, the weight of reasons for action supported by normative evidence should be carefully considered in ethical deliberation, but it may not be the only factor at play. Other considerations, such as practical constraints, individual preferences, and the potential consequences of actions, may also need to be taken into account (Williams, 1985).

The weight of reasons for action with normative evidence refers to the strength or significance of the reasons that support a particular course of action, especially when grounded in normative (ethical or moral) principles or evidence. Here are some examples: (Gregory, 2016. 2291-2310)

**a. Ethical Dilemmas:** Consider a situation where you witness someone stealing from a store. Your reasons for intervening might be weighed against the normative evidence provided by ethical principles like honesty, fairness, and the well-being of others. The weight of your reasons for action could be influenced by how strongly these principles apply in the given context.

**b. Environmental Conservation:** Suppose you're deciding whether to support a policy aimed at protecting a local ecosystem from industrial pollution. Your reasons for action might be bolstered by normative evidence from environmental ethics, scientific data about the ecosystem's health, and the potential long-term consequences of pollution on human and ecological well-being.

**c. Social Justice:** Imagine you're advocating for policies to address income inequality. Your reasons for action might be supported by normative evidence from theories of distributive justice, empirical data on income disparities, and the moral imperative to promote fairness and equality in society.

**d. Medical Ethics:** In the realm of healthcare, the weight of reasons for action could be assessed when making decisions about patient care. Normative evidence from medical ethics, such as principles of beneficence, non-maleficence, autonomy, and justice, would inform the weight of reasons for particular medical interventions or treatment plans.

**e. Legal and Political Decision-making:** When crafting laws or policies, legislators and policymakers weigh reasons for action supported by normative evidence from legal principles, constitutional values, empirical research, and public opinion. For example, in debates about freedom of speech, the weight of reasons might be assessed based on principles of free expression and the harm principle.

In each of these examples, the weight of reasons for action is determined by the strength of normative evidence and ethical principles relevant to the particular context.

In other words, "normative reasons" are called "reasons as good bases (RGB)," according to which a normative reason for you is something that is a good basis for doing an action (Gregory, 2016, 2291-2310). In RGB, this makes good sense of the weight of reason. For example, if you will save

someone's life by throwing a rope, according to RGB, throwing a rope is a normative fact from which you will save someone's life (Gregory, 2016, 2291-2310). When we should pick between various conceivable acts, our reasons may struggle, and they can contrast in what we can describe as their power, quality, or weight. If I enjoy the taste of mushrooms, I will have an important reason to eat them, but my doctor's advice that eating them is a strong reason for my death gives me a weightier reason not to eat mushrooms because this could be fatal (Jones, 2009).

Normative reason is for acting because it favors someone's action. It derives from the idea that there are norms, principles, and codes that impose actions as right and wrong. For example, when we meet someone for the first time, we shake hands, and in Indian culture, younger people greet their elders with a 'Namaste' and touch their feet (Williams, 1985). The existence of these norms depends on a variety of logical and natural relations, rules, and regulations.

Normal human limitations mean that when people decide how to act, they often have to base their decisions on flawed information or reasoning. Even when agents reason to the best of their ability and form intentions consistent with that reasoning, they sometimes get things wrong. Dominant theories about reasons for action argue that all good, or 'normative', reasons for acting are objective normative reasons. However, objective normative reasons for action are derived from facts about the world that ignore certain facts about human agents (Smith, 2012).

On these accounts of reasons, real human agents can be unable to learn what they have normative reason to do. A common response to this problem is to say that in such situations people act in a praiseworthy way, but their actions are based on false beliefs, and false beliefs cannot be good reasons (Jones, 2009). In this way, when agents reason to the best of their ability and form intentions consistent with that reasoning, agents act appropriately in response to states of the world that are normative reasons for action (Scanlon, 1998).

#### **4. Practical reason is the evidence for the weight of reason**

The phrase "Reason Have Weight with the Evidence of Practical Reason" suggests an exploration of how practical reason provides evidence for the significance of rational considerations in guiding human behavior. It implies that when people consider reasons for action, these reasons exert a certain "weight" or influence on the choices they ultimately make (Scanlon, 1998). Here's an analysis of how practical reason can indeed demonstrate the importance of rationality in decision-making:

Practical reason involves rational deliberation about what actions to take, often guided by ethical or moral principles. For instance, when deciding whether to donate to charity, a person might weigh the potential benefits to those in need against their own financial constraints. This deliberation illustrates how rational considerations can influence behavior by providing reasons that carry weight in the decision-making process (Korsgaard, 1996).

Moreover, practical reason helps individuals align their actions with their long-term goals and values. By considering the broader implications of their choices, people can make decisions that

are consistent with their principles and objectives. For example, someone committed to environmental sustainability might choose to cycle to work instead of driving, even if it's less convenient. The weight of their environmental values influences their practical decision-making (Foot, 1978).

Additionally, practical reason can resolve conflicts between competing reasons. In situations where multiple considerations pull in different directions, practical reason allows individuals to evaluate and prioritize these reasons based on their relative importance. For instance, a person might balance their desire for a luxurious lifestyle with their ethical commitment to avoid exploiting others. The ability to discern which reasons have more weight is a key aspect of practical reasoning (Parfit, 2011). In general, the weight of reasons, supported by practical reason, emphasizes the role of rational deliberation in ethical decision-making. Practical reason provides a framework for understanding how rational considerations can guide behavior, helping individuals make choices that are coherent, principled, and aligned with their values.

Here's an analysis of how practical reason can indeed demonstrate the importance of rationality in decision-making:

**1. Alignment with Goals:** Practical reason involves deliberation about how to achieve one's goals or objectives effectively. By examining how individuals make decisions in pursuit of their goals, we can observe the influence of rational considerations. Choices that are informed by reason tend to be more aligned with long-term objectives and are based on careful assessment of available options and their consequences.<sup>1</sup>

**2. Consistency and Coherence:** Practical reason often leads individuals to make decisions that are consistent with their values, beliefs, and principles. Rational decision-making involves weighing different considerations and choosing the option that best fits one's overall worldview and ethical framework. This coherence in decision-making provides evidence of the weight of reason in guiding human behavior.

**3. Adaptability and Learning:** Rational decision-making involves the ability to adapt to changing circumstances and learn from past experiences. Practical reason enables individuals to evaluate the effectiveness of their choices and adjust their behavior accordingly. This adaptability and learning process demonstrate how rational considerations play a crucial role in guiding human behavior towards more optimal outcomes.

**4. Accountability and Justification:** Practical reason requires individuals to justify their choices based on rational grounds. When people make decisions, they often provide reasons to justify their actions to themselves and others. This accountability demonstrates the importance of rational considerations in guiding behavior, as individuals seek to justify their choices based on logical arguments and evidence.

---

<sup>1</sup> <https://plato.stanford.edu/entries/practical-reason/>

Overall, by examining how practical reason operates in decision-making processes, we can indeed find evidence for the significance of rational considerations in guiding human behavior. The alignment with goals, consistency and coherence, adaptability and learning, and accountability and justification are all aspects that highlight the weight of reason in shaping our actions and choices. According to Dalia Drai delves into the intersection of practical reason, evidence-based decision-making, and the significance of rational considerations in guiding human behavior.

In general, practical reason refers to the capacity of human beings to make judgments and decisions based on rational deliberation and considerations of what actions are most conducive to achieving their goals or fulfilling their values (Kant, 1785). Practical reason is concerned with how individuals ought to act in various situations to achieve desired outcomes. It often operates by considering evidence relevant to a particular situation or problem. This evidence can include empirical data, past experiences, ethical principles, and logical deductions (McDowell, 1996).

The paper explores how individuals gather and evaluate evidence to inform their decision-making processes. A central theme is how individuals weigh different reasons when making choices in their everyday lives. This involves assessing the relative importance or significance of various considerations and determining which course of action is most rational or morally justified based on those considerations (Rawls, 1971).

From practical reason, we assess situations and make choices based on reflective consideration of available options. In daily life, we encounter various situations and options, and practical reason helps us choose the most viable and rationally defensible path to our objectives (Aristotle, 1985). Practical reason not only provides justification for our choices but also motivates us to undertake suitable actions. This consideration is practical in at least two senses:

- (1) In terms of its subject matter, and
- (2) In terms of its consequences

Practical reason is practical in virtue of its intimate relationship with action which constitutes its subject matter. Also, it necessarily considers the expected outcomes of an action for the purpose of viability and evaluation of the same. Practical reason is characterized as both a capacity to respond to the situations and as normative principles for the assessment. The function of practical reason is called practical reasoning which is described as an inferential process through which a justification is offered in terms of reasons drawn from the compelling facts of life. The choice of action or decision is determined with due to consideration to the facts of life.

Through practical reason we will weight to our reason. By instrumental desires we have create a long chain of desires and then weight to our reasonable action. In instrumental desires every action is depending on another desire for valuable or weighable action that we have done. In our daily life we have long chain of instrumental desires. Our every desire is based on some ends and that is for another batter ends but this chains all are start with a number of telic desires. To outline

with one of Aristotle's precedents,<sup>1</sup> on the off chance that I am a doctor treating a patient, my administering end as doctor (leaving aside the issue of willful extermination) is to secure, and I don't deliberate about whether I will secure the patient. I do, notwithstanding, deliberate about methods, state about whether I should provide tablets or just prescribe takes a rest. At that time rate the two conceptions about deliberation are reliable with this entry. On one, the deliberative chain contains sequence of choices prompting an official decision which is, or is at any rate firmly attached to, the principal thing in the order of causation. In this first occurrence, the deliberative chain is decisional: in the event that I choose to recommend for tablets, doing as such turns into a subsidiary end, and I may then ponder about what type of tablets I should give. In the event that I choose on penicillin as methods for treat, I have another auxiliary end and may consider about how I should complete that, state by tablet or injection. In the event that decide on tablets, I may understand that they are in the cabinet to one side. Assume I choose to give a portion of those very tablets; at that point, mindful that I need go after them, I do it. On the second conception predictable with the section, in spite of the fact that I settle on a similar official choice, the deliberative chain is cognitive: instrumental beliefs (or other intellectual components, for example, decisions) express the backup closes; for example, I don't choose to endorse medication, yet do judge recommending it to be ideal and in this way continue to distinguish the best prescription, and after that the best vehicle for giving it. I at last choose to do what is justified by the whole succession: going after the tablets. So, that type of desire is merely instrumental desire, be that as it may, on the off chance that I need after death notoriety for the wellbeing of its own; this telic want would start this specific chain. In that way all the wants that will be rational can verify by practical reason and attain the chain of instrumental desire

From Ancient concept we weight about reason from many methods. Ethical dilemmas abound in the Panchatantra, highlighting the complexities of moral decision-making.<sup>2</sup> Characters often grapple with conflicting values or temptations, forcing them to weigh the consequences of their actions on others. This encourages readers to consider ethical principles such as honesty, integrity, and compassion in their own lives. Vishnu Sharma's famous work on 'Panchatantra' reflected how we can weight our reason in different type's situation. In his fables we can see that how we can use reason in different practice of life. In every part of life, we weight our reason. In 'Panchatantra' one story a rabbit through his weight of reason save his life and the life of other animal lives in jungle from the lion. The story goes like this. Quite a long time ago, there lived a big lion in a forest. Every time he killed several animals to fulfill his hunger. The animals were stressed and they all chose to go to the lion and find out an answer for this issue. The lion said to every animal in the timberland that on the off chance that one of them will come to him as his feast for the day, he would not kill any other animal among them. Every one of the creatures decided to this. At one

---

<sup>1</sup> <https://iep.utm.edu/prac-med/><sup>2</sup> [https://www.researchgate.net/publication/374421004\\_A\\_Study\\_of\\_Value\\_Education\\_in\\_the\\_Panchatantra](https://www.researchgate.net/publication/374421004_A_Study_of_Value_Education_in_the_Panchatantra)

day it was the turn of a rabbit. The rabbit was extremely depressing. As he was setting off to the lion's corner, he went over an old well. He looked in to it. It was deep and risky. He made an arrangement to his psyche and took little longer to go to lion. The lion was extremely irate. The lion approached it the reason behind being late. It acted sensibly and stated, "Sir! I was stopped by another lion, who was claiming to be the king of the jungle and he wants to meet you." The lion thundered "Do you know where he lives?" The rabbit answered "Yes, Sir. If it's not too much trouble accompany me". The rabbit took the lion to the old well. He said "Sir, that lion lives in this well". The lion peeped into the well. He mixed up his appearance for another lion. He thundered and there was a resound. He felt that the other lion was thundering as well and bounced into the well. That was the finish of the lion. The moral of this story is- Intelligence wins over might. It means rabbit from his intelligence save his life and their friend's life also to finish the lion. He can use his intelligence with faculty of reason. Rabbit when reason about to fulfill the desire of hunger lion or save his life. At that time rabbit reason about how he can save his life and create a plan in his mind to finish the lion. The desire of hunger lion is less weighty with comparison of the life of another animal (Chandiramani, 1991).

### **Conclusion**

Overall, the paper aims to contribute to our understanding of how practical reason informs human behavior, how evidence shapes decision-making processes, and why rational considerations are essential in guiding ethical and practical choices in everyday life. Practical reason is a crucial faculty for action in an ethical way of life (Kant, 1785). As discussed, the notion that reason "has weight" rather than being merely "weighed" suggests that reason has intrinsic significance in guiding decisions and actions.

First, value-based theories of reason emphasize that reasons hold weight because they align with moral principles or ethical values (Mill, 1863). According to these theories, the weight of reasons is not a matter of physical measurement but of their moral or practical significance (Aristotle, 1985). From this faculty of mind, we encounter various life problems and weighty situations, underscoring the importance of reason in human life. There are many normative, explanatory, and motivational reasons we weigh, which become central to ethical behavior and societal norms (Rawls, 1971 & Dewey, 1939).

Furthermore, Drai's arguments suggest that practical reason involves assessing the relative importance or significance of different reasons, integrating normative evidence, and reflecting on how these considerations influence decision-making (Drai, 2009). Drai argues that practical reason is integral to resolving conflicts and making choices that align with both personal and societal values. This reflects a broader understanding that while reasons may not have a measurable weight like physical objects, they undeniably shape human actions and decision-making processes (McDowell, 1996).



Philosophers such as Kant, Mill, and Aristotle provide diverse perspectives on how reasons acquire significance and influence ethical reasoning. For instance, Kant's emphasis on moral duties, Mill's focus on utility, and Aristotle's concept of virtue all demonstrate different ways reasons can be weighted in moral deliberation. Additionally, Draï's insights into the role of practical reason highlight how reasons are not merely abstract but play a tangible role in ethical and practical decision-making (Draï, 2009).

Ultimately, the weight of a reason depends on various factors, including moral principles, consequences, virtues, and individual values. Understanding the complexity of reason's weight enriches our understanding of ethics, decision-making, and human behavior.

## References

- Anderson, P. (2022). *The Art of Abductive Reasoning*, Logic Press.
- Aristotle. (350 BCE). *Nicomachean Ethics*. Translated by W. D. Ross. Oxford University Press.
- Arcus, M. E. (1980). Value Reasoning: An Approach to Values Education. *Family Relations*. 29(2), 163-171. <https://doi.org/10.2307/584067>
- Bentham, J. (1789). *An Introduction to the Principles of Morals and Legislation*, T. Payne
- Broome, J. (2002). Reason and Nature: Essays in the Theory of Rationality. In J. L. Millar, *Practical Reasoning* (pp. 85-111). Oxford University Press.
- Broome, J. (2005). *Weighing Goods: Equality, Uncertainty, and Time*. Oxford University Press.
- Broome, J. (2008). Reply to Southwood, Kearns and Star, and Cullity. *Ethics* 119. In *Ethics* (pp. 96-108). University of Chicago.
- Clark, R. (2020). *Unconscious Cognition*, Mind Matters.
- Chang, R. (2020). Do we have normative powers? *Aristotelian Society Supplementary*, 94 (1), 275-300. <http://dx.doi.org/10.1093/arisup/akaa012>
- Davis, L. (2019). *Cognitive Processes in Decision Making*, Cognitive Science Press.
- Dancy, J. (2002). Reasons for Action. In J. Dancy, *Practical Reality*, Oxford University Press.
- Draï, D. (2018). Reasons Have no Weight. *The Philosophical Quarterly*, 60-76.
- Evans, J. (2019). *Justifying Actions Through Reason*, Ethical Press.
- Fogal, D. & Olle, R. (2023). The weight of reasons. *Philosophical Studies*, 180(9), 1-24. <http://dx.doi.org/10.1007/s11098-023-01974-y>
- Foot, P. (1978). *Virtues and Vices*. University of California Press.
- Frege, G. (1892). *Über Sinn und Bedeutung*. In *Zeitschrift für Philosophie und philosophische Kritik*.
- Gorham, G. (Sep., 2002). Descartes on the Innateness of All Ideas. *Canadian Journal of Philosophy*, 32 (3), 355-388. <http://dx.doi.org/10.2307/40232155>
- Gregory, A. (2016). Normative Reasons as Good Bases . *Philosophical Studies* 9, 2291-2310.
- Hempel, C. G. (1965). *Aspects of Scientific Explanation*, The Free Press.
- [https://www.researchgate.net/publication/324271986\\_Reasoning](https://www.researchgate.net/publication/324271986_Reasoning)
- <https://www.mindtools.com/a3ixqae/critical-thinking?from=shared-link>
- <https://doi.org/10.1111/nous.12429>
- <https://plato.stanford.edu/Entries/reasons-just-vs-expl/>

<https://plato.stanford.edu/entries/practical-reason/>

<https://iep.utm.edu/prac-med/>

[https://www.researchgate.net/publication/374421004\\_A\\_Study\\_of\\_Value\\_Education\\_in\\_the\\_Panchatantra](https://www.researchgate.net/publication/374421004_A_Study_of_Value_Education_in_the_Panchatantra)

Jones, B. (2019). *Patterns in Inductive Reasoning*, Learning Publishers.

James, W. (1907). *Pragmatism: A New Name for Some Old Ways of Thinking*, Longmans, Green, and Co.

Johnson, T. (2018). *Logical Inference and Mental Processes*, Academic Press.

Kant, I. (1785). *Groundwork of the Metaphysics of Morals*, Johann Friedrich Hartknoch.

Korsgaard, C. M. (1996). *The Sources of Normativity*. Cambridge University Press.

Maguire, B. (2016). The value based Theory of Reasons. *Journal of Philosophy*, 3, 233-262.

<http://dx.doi.org/10.3998/ergo.12405314.0003.009>

Maguire, E. L. (2017). *Weighing Reasons*. oxford University Press.

Maguire, B. (2016). *The Weight of Reasons: A Study in Value and Decision Making*, Oxford University Press.

Mill, J. S. (1863). *Utilitarianism*, Parker, Son, and Bourn.

Miller, S. (2017). *Complex Problem Solving and Reasoning*, Thoughtful Books.

Rawls, J. (1971). *A Theory of Justice*, Harvard University Press.

Roberts, M. (2018). *Adaptability and Reasoning*. Adaptive Minds Publishing.

Ricco, R. B. & Overton, W. F. (2009). Reasoning. To appear in V. S. Ramachandran (Ed.), *Encyclopedia of human behavior*, 2<sup>nd</sup> Edition.

Sharma, P. V. (1990). Panchatantra . In G. L. Chandiramani. Rupa Publications India .

Kearns, S. & Star, D. (2013). Weighing Reasons. *Journal of Moral Philosophy*, 70-86.

Sinnott-Armstrong, W. (2009). *Moral Skepticism and Moral Judgment*, Oxford University Press.

Singer, P. (1972). *Hunger, Poverty, and the Ethics of Giving*, Random House.

Smith, C. (2020). *Principles of Deductive Reasoning*, Thoughtful Books.

Sartre, J. P. (1943). *Being and Nothingness*, Gallimard.

Taylor, L. (2020). *Financial Planning and Reason*, Money Matters.

Thompson, L. (2020). *The Role of Reasoning in Human Thought*, Cognitive Science Books.

Whiting, D. (2017). Right in some respects: reasons as evidence. *Philosophical Studies*, 175(9), 2191–2208.

<https://link.springer.com/article/10.1007/s11098-017-0954-x>

Wilson, G. (2021). *Principles and Practices of Inductive Reasoning*, Learning Publishers.



Journal of Philosophical Investigations

## Journal of Philosophical Investigations

Print ISSN: 2251-7960 Online ISSN: 2423-4419

Homepage: <https://philosophy.tabrizu.ac.ir>



University of Tabriz

# Kant's Moral Theory Meets Evolutionary Theory

Alireza Mansouri 

Associate Professor, Philosophy of Science Dept., Institute for Humanities and Cultural Studies, Iran. Email: [mansouri@ihcs.ac.ir](mailto:mansouri@ihcs.ac.ir)

### Article Info

#### Article type:

Research Article

#### Article history:

Received 26 June 2024

Received in revised form  
01 July 2024

Accepted 10 July 2024

Published online 07  
August 2024

#### Keywords:

categorical imperative,  
personhood, moral status,  
moral patient,  
evolutionary ethics.

### ABSTRACT

This paper delves into the intersection between Kant's moral theory and evolutionary perspectives on *personhood*. It explores how Kant's emphasis on rationality in moral agency aligns with evolutionary studies on the development of moral behaviors. By examining the transcendental implications of Kant's Categorical Imperative (CI) and the evolutionary origins of moral agency, this study aims to illuminate the link between Kant's conception of *moral agency* and *personhood*. Additionally, it investigates how Kant's call for CI resonates with evolutionary insights on the adaptive nature of social cooperation in human societies. Through this analysis, we seek to deepen our understanding of the cognitive, social dimensions of moral agency and moral status within the framework of Kant's moral theory and evolutionary perspectives on *personhood*.

**Cite this article:** Mansouri, A. (2024). Kant's Moral Theory Meets Evolutionary Theory. *Journal of Philosophical Investigations*, 18(47), 251-264. <https://doi.org/10.22034/JPIUT.2024.62262.3799>



© The Author(s).

<https://doi.org/10.22034/JPIUT.2024.62262.3799>

Publisher: University of Tabriz.

## Introduction

One of the key features of Kant's moral theory is that *persons*, as *rational* agents, have a special value and deserve moral respect (Kant, 1998). To be a *person*, for Kant, is to have the ability to act autonomously, that is, to follow one's own *reason* and not external forces or desires; to have a rational will that is capable of giving universal laws, and to be an end in itself with dignity and respect. Acting autonomously means exercising one's rational will in accordance with the principle of universal law, and being free from domination by one's own inclinations or those of others. Therefore, to be a *person* and to act *autonomously* are inseparable, since only by acting autonomously can one realize one's personhood and dignity, and only by being a person can one have the capacity and obligation to act autonomously (G, 4, 440, 432; Guyer, 2003). Although the concept of *personhood* is crucial for Kant's moral theory, he did not develop a comprehensive theory of it. However, philosophical and empirical approaches, informed by evolutionary biology and psychology, challenge the idea that *personhood* is a *static* or *innate* attribute. Instead, they propose that personhood is a *dynamic* and *relational* notion that arises from the interplay and adaptation of living beings in complex environments (Barresi 1999; 2001; 2016; Moore & Lemmon 2001; Popper & Eccles, 1985, Ch. 4; Thompson, 2006). For instance, Barresi (1999; 2016) contends that personhood is a cognitive and social construct that evolved in humans as a result of their capacity to think of themselves and others as equivalent agents, and to participate in reciprocal and cooperative behaviors. He also maintains that personhood is closely linked to *moral agency*, as it enables humans to apply ethical norms and principles to themselves and others, and to respect the dignity and value of persons. Therefore, it is essential to investigate whether Kant's moral theory can accommodate these evolutionary insights, and what implications they have for his ethical principles and arguments.

According to Kant's *Groundwork of the Metaphysics of Morals* (G), the *Categorical Imperative* (CI) is the supreme principle of morality. Kant defines the CI as an objective, rationally necessary, and unconditional principle that we must always follow regardless of our natural desires or inclinations (G, 4, 415; Johnson and Cureton, 2022). Barresi (2012, 128) briefly suggests that Kant's notion of personhood is based on a human conception of self and others as persons. However, evolutionary studies show that conception has an evolutionary origin in developing reciprocal altruism in hominids. This paper aims to explore this suggestion and examine the link between Kant's moral agency and the evolutionary origins and functions of personhood. It argues that Kant's idea of 'person' as a rational and moral agent needs to be illuminated within a theoretical framework incorporating evolutionary insights into the nature and emergence of personhood.

The structure of the paper is as follows. Section 2 examines the central role of 'personhood' in Kant's formulations of the CI. Section 3 contends that an evolutionary perspective on personhood enriches Kant's moral theory and illuminates its relation to personhood. Section 4 is the conclusion.

## 1. Kant's Moral Theory and Personhood

Kant proposed various formulations of the CI. We contend that the CI, in all its formulations, entails that an agent needs to reach a level of cognitive capacity, namely *personhood*, to be able to make moral judgments. *Personhood* is a concept that allows us to understand both self and others as agents of the same kind, with shared intentional relations and agent-neutral perspectives (Nagel, 1986). We will discover how different formulations presuppose this cognitive ability:

*The Universality Formula.* Kant's Universality Formulation of the CI states ‘

act only in accordance with that maxim through which you can at the same time will that it become a universal law’ (G, 4, 421).

This implies that *the act we are morally obligated to do is motivated by adherence to a principle that could, without inconsistency, be held to apply to any (and all) rational agents*. This formulation underscores the necessity for moral obligations to be based on principles that could apply consistently to all rational agents. However, to ascertain the *universality* of a moral law, one must first possess the cognitive capacity to empathize and recognize oneself and others as *persons* who could be subject to that principle. This transcendental argument highlights the pivotal role of *personhood* in the Universality Formulation, as individuals must attain a level of cognitive ability that enables them to adopt an agent-neutral perspective and consider moral principles universally applicable to all rational beings. In fact, the Universality Formulation underscores the inherent connection between personhood, moral judgment, and the capacity to conceive of moral principles that transcend individual inclinations and desires.

*The Humanity Formula.* Kant's Humanity formula of the CI states (G, 4, 29):

So act that you treat humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means.

This formula implies that one should respect the dignity and worth of every human being, not use them as mere instruments for one's own purposes, but treat them as rational and autonomous agents with their own ends and values. To treat humanity as means is to disregard the autonomy and freedom of human beings, and to treat them as objects or resources that can be manipulated or exploited. Kant wrote elsewhere (G, 4, 428):

Beings the existence of which rests not on our will, but on nature, if they are beings without reason, still have only a relative worth as means, and are therefore called *things*. Whereas rational beings are called *persons* because their nature already marks them out as an end in itself.

This view draws on Kant's conception of *humanity as an end in itself*, and highlights his distinction between rational beings and non-rational beings. It aims to show that the value of an

individual does not depend on its relation to others, but on its own *rational* nature, which makes it a moral agent. It also implies that treating individuals as means violates their rational nature and reduces them to non-rational beings.

The Humanity Formula relies on the idea of the unconditioned good and the role of *reason* in setting *ends* (Korsgaard, 1986). In other words, the concept of humanity refers to the capacity of human beings to set their own *ends* by using *reason*, which sets them apart from animals. However, humanity is not fully rational, as it can be affected by inclinations and passions that are not based on reason. The concept of *personhood* denotes the state of being fully rational and acting from duty, which is the respect for the moral law as the only motive of the will. Therefore, personhood is the fulfillment and perfection of humanity, as it realizes the full potential of rational nature (Korsgaard, 1986). The Humanity Formula suggests that *personhood* involves a cognitive capacity that enables the fulfillment of rationality and humanity. It allows agents to acknowledge the moral law as the only reason for the will and apply it universally to all rational beings. Personhood also enables agents to adopt an agent-neutral perspective, in which they consider what is morally required by the universal law rather than what is dictated by their desires. This perspective transcends individual interests and focuses on the objective demands of morality.

Furthermore, this capacity leads to understanding the principle of reciprocal exchange, in which agents treat others as ends in themselves and expect the same respect in return. Thus, in another way, we come to this conclusion that the Humanity Formula's requirement to treat humanity as an end in itself depends on the cognitive capacity that is related to *personhood*, and this capacity allows agents to recognize universal moral norms, adopt an agent-neutral perspective, and appreciate the reciprocity involved in moral interactions.

The application of the Humanity Formula also implies personhood. Applying the Humanity Formula requires one to possess the ability to identify and appreciate humanity as an end in itself, in their selves and others, and to behave accordingly. This means that one must respect the rational nature of individuals, which is the source of their intrinsic worth or dignity. It also means that one must acknowledge the autonomy and freedom of individuals, which is their ability to act according to their own reason, without being determined by external factors, influences, and desires, *in any given situation* (Kant, 1998; Wood, 2011, 58-82). This implies that the moral agent must have the capability to conceive *abstract person*. Hence, the Humanity Formula presupposes that the agent has the cognitive capacity of *personhood*.

In sum, the Humanity Formula connects personhood, as the ability to adopt an agent-neutral perspective, with the moral value of rational beings as ends in themselves, and by doing so, it supports the idea that true moral agency depends on a level of cognitive development. Furthermore, this perspective highlights the importance of acknowledging and respecting the personhood of individuals in ethical decision-making. It also stresses the transcendental condition for moral

agency as attaining a stage of personhood where one can recognize the inherent worth and dignity of all rational beings as *persons*.

Furthermore, personhood manifests in Kant's concept of 'respect for person'. the idea that *respect for persons* comprises the core of morality has long been associated with Kant and his ethics in the *Groundwork of the Metaphysics of Morals*. As a received view, the Humanity Formula is often glossed as enjoining us to respect persons as such. However, while Kant does remark that, inasmuch as 'their nature already marks them out as an end in itself' persons are *objects of respect*, he does not explicitly characterize the Humanity Formula as a principle enjoining respect for *persons*. Kant even repeatedly tells us in the *Groundwork* that the only proper *object of respect* is the *moral law* (G, 4, 400-1, 403, 405, 436).<sup>1</sup> There have been attempts to give an account of *respect* that could reconcile the scattering remarks of Kant (Klimchuk, 2004). But I think if we take 'person' as an *abstract concept*, independent of individuals, we will find that *respect for the law* and *respect for an abstract person* are two sides of the same coin as they depend on each other, being interdependent concepts. Respecting a moral law in relation to an individual means acknowledging their freedom, and respecting an abstract person means respecting their humanity, that is, acknowledging their freedom and autonomy. In addition, acting according to rules requires a conception of *person* that has the ability to follow rules. In other words, she has the capacity to set and pursue ends. "A being who has *humanity* can also act with the help of rules that have a wider scope, indeed ones that apply to all persons.", namely, act according to universal laws; and according to the Humanity Formula, in moral behavior, the end is 'humanity' itself (Kerstein 2013, 17; 2023). In sum, it seems in Kant's moral theory 'respect for persons' presupposes an abstract sense of 'person' which demands the conception of *personhood*.

*The Autonomy Formula.* As Kant writes, it is 'the idea of the will of every rational being as a *will giving universal law*' (G, 4, 432). Although Kant does not state this as an imperative, as he does in the other formulations, it is easy enough to put it in that form: Act so that you could be a *legislator of universal laws* through your maxims. In this formula, Kant sees the individual as a *legislator* of universal laws. A 'rational will' merely bounded by universal laws could act according to natural and non-moral motives, such as self-interest. However, in order to be a legislator of universal laws, such contingent motives must be set aside (Johnson & Cureton 2022). Because for Kant, moral action should be in accordance with 'duty', and we do our moral duty when we act

---

<sup>1</sup> One may question the possibility of acting purely out of respect for the moral law. For instance, if we claim that we keep the promise only for *desire-independent reason* – for the sake of the moral law – someone may object that there are various prudential reasons that motivate us to keep the promise. However, following Searle (2005, 21), this objection can be countered by distinguishing between thought experiment and empirical hypothesis. It may be the case that there are various prudential reasons involved in fulfilling my promise, but the crux of the matter is that fulfilling the promise as a moral act requires imagining a situation in which all those prudential reasons are absent and it is only the call of duty with the intention of respecting the moral law that guides my action. Irrespective of any empirical hypothesis, nothing precludes us from conceiving such a thought experiment.

according to a *principle* recognized by reason, rather than the *desire* for achieving a *consequence* or emotional feeling.

Human beings possess *reason* regardless of any specific level of cognitive ability. However, our *reason* (or our intellectual faculty) can only act according to *principles* only if it has the cognitive capacity to differentiate between actions based on *principles* and actions based on *desires*. This cognitive capacity is a faculty that allows moral agents to conceive of themselves and others as *persons*, and the *Autonomy Formula* assumes that humans have this faculty. Hence, this formulation also implies that legislation is impossible without the ability to adopt a first- and third-person perspective and unite them (personhood), which enables acting according to laws rather than desires.

*The Kingdom of Ends Formula.* This is the fourth formulation of CI. Though ‘*Reich der Zwecke*’ is usually translated as ‘Kingdom of Ends’, the German word *Reich* is perhaps more appropriately translated as ‘realm’. Kant introduced this concept in his *Groundwork of the Metaphysics of Morals* (439), envisioning a world in which all human beings are treated as ends, not merely as a means to an end for others. He used this concept to mean ‘a systematic union of various rational beings through common laws’ (G, 4, 433). These common laws, established by the CI, are the gauge used to evaluate the worthiness of an individual’s actions. This systematic whole is the Kingdom of Ends, which contributes to universalizing ethical theory for society.

When all the kingdom’s individuals live by the CI—particularly Kant’s second formulation—each one will treat all of his fellowmen as ends in their selves, instead of means to achieving one’s own selfish goals. This means that they must judge themselves and their actions from the perspective of a *person*. Therefore, as implied by the second formulation, each individual in this kingdom needs a cognitive capacity for the conception of *personhood*, which goes beyond individuals’ desires.

*Equivalency claim.* Another reason for supporting the idea that ‘personhood’ plays a central role in different formulations of CI is Kant’s *equivalency claim*. Unfortunately, Kant does not say in what sense different formulations of CI are equivalent. All that he says is that ‘the above three ways of representing the principle of morality are at bottom only so many formulae of the very same law, and any one of them of itself unites the other two in it’ (G, 4, 436), and that the differences between them are ‘subjectively rather than objectively practical’ in the sense that each of them aims ‘to bring an idea of reason closer to intuition (by a certain analogy), and thereby closer to feeling’ (G, 4, 436). He also says that one formula ‘follows’ from another (G, 4, 431) and that the concept that is foundational to one formula ‘leads to’ a concept that depends on another formula (G, 4, 433).

Kant’s remarks on the *equivalency claim* can be understood in different ways. A simple interpretation might be that each formula generates the same and only the same duties, if followed or applied (Allison, 2011; O’Neill, 1975; Engstrom, 2009; Sensen, 2011). However, Allen Wood



rejects this interpretation. He argues that Kant's formulas do not *prescribe* actions. They assign a distinct function (or set of functions) to each formula, and in that fairly obvious sense, they are not *equivalent*. Wood points out that for Kant these are 'so many formulas of the very same law' But 'he never says: that they are *equivalent*' (Wood, 2017, 74-76).

We can interpret Kant's project as aiming to provide a 'metaphysics of morals' that explores the basis of moral obligation in the nature of a rational being. In doing so, he confronts a *transcendental question*: What constitutes *moral agency*? Kant's CI, by transcendental reasoning, seeks to provide the ontological and metaphysical condition for 'being a moral agent'. This reading might become more plausible, especially if we consider that Kant's general transcendental approach implies deriving from these formulas some kind of *transcendental condition* for the applicability of the CI. Assuming CI formulas as different versions of the same law entails that understanding and applying such a law requires the ability to conceive oneself and others as *persons*. In other words, despite their differences, every formulation of the CI could reveal a transcendental condition under which an agent could be a *moral agent*. This condition could be the cognitive capacity of *personhood* for the applicability of the CI. Kerstein (2013) and Wood (1998) have offered similar interpretations of this idea. Wood argues that the Humanity Formulation of the CI need not lead Kant to conclude that only *actually rational* beings are ends in themselves. This consequence only arises because Kant also accepts what Wood dubs the 'personification principle,' which states that 'rational nature is respected only by respecting humanity *in someone's person*, hence that every duty must be understood as a duty toward a *person* or persons' (Wood, 1998, 198). Therefore, Wood's view seems to support our suggestion regarding the *centrality of personhood* in different formulations of the CI.

Furthermore, according to Wood, a plausible interpretation of Kant's view of *respect* for the rational nature requires that children and people with mental disorders should not be treated as mere things or mere instruments; they should be regarded as 'persons' in the *extended sense*. Persons in the *extended sense* do not have precisely the same *moral status* as persons in the strict sense. But they do not have a *lesser* status. Their lack of rational capacities to autonomously direct their own lives implies that they do not have the same right to self-determination that persons in the strict sense possess. However, persons in the extended sense have the same right to life as persons in the strict sense, and we have the same duties to respect their interests and regard them as ends in themselves (Wood, 2008, 97-100). However, Wood does not clearly explain this extended sense of person. For him, this extended sense of person seems to require a 'right relation' towards rational nature, so part of respecting rational nature is respecting its *potential* within a young child. Thus, although young children are not, strictly speaking, Kantian *persons*, their potential rationality bears a 'right relation' to actual rationality, so they must be treated as an end in themselves (Wood, 1998, 197, 198).

We concur with the part of Wood's approach that emphasizes the *extended sense of person*, yet suggesting a *theory* of personhood might be preferable to an explanation based on *potential rationality*. This theory should account for the emergence of personhood exclusively among humans as rational beings. The following section will explore how an evolutionary view of personhood can shed light on Kant's moral theory.

## 2. Personhood: An Evolutionary Approach

As mentioned above, *personhood* has a pivotal role in Kant's moral theory and his CI. In his view, personhood involves the ability to act autonomously, guided by one's own reason rather than external influences. Autonomy allows individuals to make moral choices based on rational principles. In order to elucidate Kant's conception of 'personhood', we need to offer an explanation of it. Nevertheless, this requirement does not seek to supply some descriptions for the idea of personhood. Rather than concepts, which may have a lot of suggestive power, *theories* are what matter. what should interest us are theories, truth, problems, and arguments. Many philosophers and scientists erroneously assume that concepts and conceptual systems (and problems of their meaning, or the meaning of words) have a similar importance to theories and theoretical systems (and problems of their truth, or the truth of statements). While concepts are partly means of formulating theories and partly means of summing up theories. In any case, their significance is mainly instrumental'; and they can always be substituted by other concepts (Popper, 1979, 123-4). Consequently, concepts such as 'personhood' derive meaning and importance from theories and serve them instrumentally. Therefore, the appropriate way to formulate the above question is to seek a satisfactory *theory of personhood* and its relation to morality.

The possibility of providing an evolutionary explanation of human morality has attracted much attention in recent decades (Binmore 2005; Boehm 2012; de Waal 2006; Greene 2013; Haidt 2012; Joyce 2006; Kitcher 2011; Krebs 2011; Nichols 2004). Following this line of thought, Barresi (2012; 2016) proposed an *evolutionary account of personhood* and its connection to morality, which integrates first- and third- person perspectives. Based on his account, we offer an understanding of personhood from an evolutionary perspective, which may illuminate Kant's moral view and the CI.

Building on this evolutionary perspective, we explore the idea that a significant evolution occurred from emotion-based pro-social motives to explicitly moral motivations in humans. Kitcher (2011) argues that altruism based on sympathetic responses found in chimpanzees is insufficient to sustain a large society. Collective intentionality and normative rules, which transcend emotion-based motives, are necessary for achieving a high level of collaboration in large societies. Similarly, Barresi (2016) claims that the normative foundation of human moral life depends on the cognitive capacity to conceive of their selves and others as *persons*. The essential requirement for moral norms to be universal and independent of one's emotional tendencies is to

conceive both self and others equally as members of a class of agents whose duty is to accede to those demands. For instance, when we say one should not cause pain to another human being, it implies that pain should be avoided regardless of whose pain it is. In doing so, a concept of 'agent' that bridges the gap between the self and other is needed. That concept is 'person'. Nagel (1970) aptly expresses this notion of person when he depicts the situation of one individual standing on the foot of another. 'Recognition of the other person's reality, and the possibility of putting yourself in his place, is essential. You see the present situation as a specimen of a more general scheme, in which the characters can be exchanged' (82). Nagel considers the concept of the 'agent-neutral' as a pre-condition for moral experience (Parfit 1984; Nagel 1986). Because he holds that moral experience entails impersonal motives and reasons, unlike emotional experience, which is 'agent-relative' and typical in other animals. As Barresi (2016, 103) puts it:

In imagining the exchange, you would expect the other person to release his or her foot, not because it would reduce 'your' pain, but because it would reduce 'someone's' pain, some 'person's' pain. This is a unique aspect of the human moral order not found in other animals. Humans can conceive of themselves as just another person and that all persons should be treated equally with respect to moral norms. ... Without a rich capacity for perspective taking that makes possible full imagination of the reversal of positions, mere aversion to pain in another would vary with one's personal relationship to the other, and this would make it a self-interested motive, not a motive based on a conception of self and other equally as persons whose pain ought to be avoided.

We experience the personhood of self and others in different ways. Our experience of our own personhood is a first-person experience of our self, while our experience of each other's personhood is from a third-person perspective (Barresi, 2016, 107).

During the evolutionary process and pressure, humans, unlike chimpanzees and other animals, have been able to unite the two experiences through the need for a high level of *cooperative activity*. Barresi suggests that the human organism is the primary source of unity for the ontological, experiential, and social dimensions of person, and that the first-person experience of self and the social relations with others are essential aspects of personhood and selfhood (2012, 124). He points to various experiments to show that in humans, the first-person and the third-person are combined in a single common format that applies uniformly to self and others and could engage in normative guidance based on agent-neutral situational rules generated within one's group, not just on personal relations (Barresi, 2016, 104). Humans' conception of the self and others as *persons* becomes available in the second year of life and goes through several development stages crucial to our way

of life as cooperative organisms. This development leads to passing through emotional and social motivations and reaching the stage of personhood (Barresi, 2016, 104).

Raising infants in humans are among the needs that have required collective cooperation. In the case of chimpanzees, this breeding takes less time and is only with the mother's help. However, in humans, parents and other group members are also involved in caring for the baby (Chapais, 2008; Hrdy, 2009). Therefore, the difference between humans and chimpanzees is not in degree but in quality. It requires a kind of equality- and equity-demand that is seen only in humans. The evolutionary scenario explains this by the cognitive capacity of personhood which is specific to humans (Barresi, 2016; Binmore, 2005; Sterelny, 2012).

A related phenomenon is the contrast between adults and children in their adherence to promises. Research suggests that 2- and 3-year-olds have developed a sense of self and others as persons, and can reason about them in agent-neutral ways. Consequently, these children are very concerned about following rules and expect others to do the same. Moreover, they can differentiate between moral and conventional norms, regarding the former as universal and the latter as contingent and group-specific (Nichols, 2004; Rakoczy & Schmidt, 2013; Tomasello & Vaish, 2013). However, their conception of *person* and *self* is limited to here and now, or extended only with respect to well-known routines. Whereas for participating in normative and moral activities, it is necessary that one be conscious, at different times, of the numerical identity of its self. Two more significant advances are required for children and adolescents to achieve an adult concept of 'person' (Barresi 1999; Thompson, 2006). These two advances are the concept of a temporally extended person and self, and a life-course narrative identity. Both of these concepts are essential for recognizing and performing adult activities, like keeping promises, in an agent-neutral manner. A major change occurs during the fourth and fifth years of growth. At this time, self-reflection enters fully into the temporal domain, and the child can perceive personal identity over time (Barresi 2001; 2016, 112; Moore & Lemmon, 2001). At this stage, the child can distinguish between her past and future representations of reality as distinct from her present representations. At this point, they begin to appreciate their selves as well as others as selves extended in time. Before this time, their experiences are not linked to an autobiographical stream. But now, retrospective memory and future anticipation have this structure (Barresi, 2016, 111-112). Thus, only in adolescence do human beings acquire the ability to differentiate between temporary laws and moral norms, and to extend their perception of 'person in the present' to individuals' past and future situations. This ability is essential for keeping promises, as the agent must comprehend an abstract concept of their self beyond the present time. This ability results from the cognitive capacity of personhood, which is unique to human beings and develops further during the developmental stages. Personhood emerges from the interaction of the self with other selves and objects in the environment (Popper, 1985, Ch. 4). Therefore, personhood in humans is not static

but dynamic and evolving. Wood's interpretation of Kant supports this view, as he refers to this development (Wood, 2008, 99):

Of course, it is one thing to say that parents should be thought of as bringing a person into being, and even that they have duties of care to their offspring from conception. It is quite different to say that the offspring is a *person* from conception onward. The first two things Kant does appear to say; the third is something he never quite says.

At a certain stage of development, humans can conceive themselves and others as *persons* with consistent individual personalities as well as life-course identities in agent-neutral ways (McAdams, 1990; Barresi, 2016, 112). Popper emphasizes this point when he says: for Kant 'a person is something that is conscious, at different times, of the numerical identity of its self' (translated and quoted by Popper (1985, 115).

Barresi argues that our conception of person and self evolves as we grow and gain more capacity for more comprehensive agent-neutral forms of representation. The earlier concepts of person and self are more restricted and agent-relative than the higher levels. Only an advanced conception of the person, which only humans have achieved, enables agent-neutral cooperative and moral activity. Without the adaptive need for more intense forms of cooperation than chimpanzee life requires, our ancestors may not have developed the kind of agent-neutral thinking that treats self and others equally as persons, which is a necessary conceptual capacity for human moral life (Barresi, 2016, 113).

## Conclusion

Evolutionary studies may shed some light on the role of *personhood* in Kant's moral theory. They also seem to support Kant's distinction between moral and emotional actions. Kant maintains that actions based solely on emotional motives are *not* moral; they must stem from a sense of 'duty' and acceptance of a universal moral rule or principle. Two conditions are necessary to accept and act on normative rules as a duty: first, the rules must be the same for everyone. Second, each person must regard themselves and others as agents whose duty is to follow those rules. Hence, performing a duty requires the ability to view a situation from first- and third-person perspectives and unite them, which entails reaching the advanced level of personhood. Only at this stage can the agent see the rules as universal principles; disregard their feelings and preferences, and follow the law and the rule as a 'duty'. Laws may coincide with individual or collective desires, but for collective cooperation to take place, their application must be independent of desires and feelings. Therefore, the CI could be seen as an agent-based prescription.

This conclusion implies that *personhood*, rather than *reason* and *rationality*, might be the focus of Kant's moral theory. This might make his moral theory more defensible and avoid the criticism

of the sort of *moral patient*. Taking personhood seriously does not necessarily force us to deny the moral status of non-rational beings, because the evolutionary scenario suggests that there is no necessary symmetry of personhood between moral agents and moral patients. Moral duty can only be attributed to those beings who have attained the cognitive capacity of personhood. Thus, adults may have a moral duty towards children or infants, who have moral status. However, an infant cannot be a moral agent, nor have any moral duty, because they have not reached the stage of personhood. Therefore, the infant has no duty or responsibility. Moral patients are beings towards whom moral agents may have moral responsibilities but do not have to be moral agents themselves. Only moral agents can act as the bearers of moral obligations towards others. While moral patients can be the recipients of the moral obligations of others, they do not need moral agency capacity because they may not have achieved an advanced level of personhood. In this line of thought, even non-human animals that lack the cognitive capacity to conceive personhood for moral agency, could be classified as moral patients.

## References

- Allison, H. (2011). *Kant's Groundwork for the Metaphysics of Morals: A Commentary*. Oxford University Press.
- Barresi, J. (1999). On becoming a person. *Philosophical Psychology*, 12 (1), 79–98  
<http://dx.doi.org/10.1080/095150899105945>
- Barresi, J. (2001). Extending self-consciousness into the future. In C. Moore, & K. Lemmon, *The Self in Time: Developmental Perspectives*, pp. 141–161. Erlbaum.
- Barresi, J. (2012). On seeing our selves and others as persons. *New Ideas in Psychology*, 30(1), 120-130.  
<http://dx.doi.org/10.1016/j.newideapsych.2009.11.003>
- Barresi, J. (2016). Personhood and humanhood: An evolutionary scenario. In J. Kiverstein, *The Routledge Handbook of Philosophy of the Social Mind* (pp. 102-114). Routledge.
- Binmore, K. (2005). *Natural Justice*. Oxford University Press.
- Boehm, C. (2012). *Moral Origins: The Evolution of Virtue, Altruism, and Shame*. Basic Books.
- Chapais, B. (2008). *Primeval Kinship: How Pair-bonding gave Birth to Human Society*. Harvard University Press.
- de Waal, F. M. (2006). *Primates and Philosophers: How Morality Evolved*. Princeton University.
- Engstrom, S. (2009). *The Form of Practical Knowledge*. Harvard University Press.
- Guyer, P. (2003). Kant on the theory and practice of autonomy. *Social philosophy and policy*, 20(2), 70-98.  
<http://dx.doi.org/10.1017/S026505250320203X>
- Haidt, J. (2012). *The Righteous Mind: Why Good People Are Divided by Politics and Religion*. Pantheon.
- Hrdy, S. (2009). *Mothers and Others*. Harvard University Press.
- Johnson, R., & Cureton, A. (2022). *Kant's Moral Philosophy*. Editor by E. N. Zalta, Retrieved from The Stanford Encyclopedia of Philosophy: <https://plato.stanford.edu/archives/spr2022/entries/kant-moral/>
- Joyce, R. (2006). *The Evolution of Morality*. MIT Press.

- Kain, P. (2009). Kant's Defense of Human Moral Status. *Journal of the History of Philosophy*, 47(1), 59–101. <http://dx.doi.org/10.1353/hph.0.0083>
- Kant, I. (1998). *Groundwork of the Metaphysics of Morals*. translated by M. J. Gregor, Cambridge University Press.
- Kerstein, S. J. (2013). *How to Treat Persons*. Oxford University Press.
- Kerstein, S. J. (2023). *Treating Persons as Means*. Editor by E. Zalta, Editor, Retrieved from The Stanford Encyclopedia of Philosophy: <https://plato.stanford.edu/archives/sum2019/entries/persons-means/>
- Kitcher, P. (2011). *The Ethical Project*. Harvard University Press.
- Klimchuk, D. (2004). Three Accounts of Respect for Persons. *Kantian Review*, 8, 38-61. <http://dx.doi.org/10.1017/s1369415400001862>
- Korsgaard, C. (1996). *Creating the Kingdom of Ends*. Cambridge University Press.
- Korsgaard, C. (1986). Kant's formula of humanity. *Kant Studien*, 77(1-4), 183-202. <https://doi.org/10.1515/kant.1986.77.1-4.183>
- Krebs, D. L. (2011). *The Origins of Morality: An Evolutionary Account*. Oxford University Press.
- McAdams, D. P. (1990). Unity and purpose in human lives: The emergence of identity as a life story. In A. I. Rubin (Ed.), *Studying persons and lives* (pp. 148-200). Springer.
- Moore, C., & Lemmon, K. (2001). *The Self in Time: Developmental Perspectives*. Erlbaum.
- Nagel, T. (1970). *The Possibility of Altruism*. Oxford University Press.
- Nagel, T. (1986). *The View from Nowhere*. Oxford University Press.
- Nichols, S. (2004). *Sentimental Rules*. Oxford University Press.
- O'Neill, O. (1975). *Acting on Principle*. Columbia University Press.
- Parfit, D. (1984). *Reasons and Persons*. Oxford University Press.
- Popper, K. R. (1985). *Self and Its Brain*. Springer International.
- Popper, K. R. (1979). *Objective Knowledge: An Evolutionary Approach* (Revised edition ed.). Oxford University Press.
- Rakoczy, H. & Schmidt, M. (2013). The early ontogeny of social norms. *Child Development Perspectives*, 7, 17–21. <http://dx.doi.org/10.1111/cdep.12010>
- Searle, J. R. (2005). What is an institution? *Journal of institutional economics*, 1(1), 1-22. <http://dx.doi.org/10.1017/S1744137405000020>
- Sensen, O. (2011). Kant on Duties Toward Others From Respect (TL 37–44), ” in Andreas Trampota, and Jens Timmermann (eds.), Kant's “Tugendlehre”: A Commentary, Berlin: Walter de Gruyter, 343–364. In A. Trampota, O. Sensen, & J. Timmermann, *Kant's “Tugendlehre”: A Commentary* (pp. 343-364). Walter de Gruyter.
- Sterelny, K. (2012). *The Evolved Apprentice: How Evolution Made Humans Unique*. MIT Press.
- Thompson, R. (2006). The Development of the Person: Social Understanding, Relationships, Consciencescience, Self. In N. Eisenberg, *Handbook of Child Psychology: Social, Emotional, and Personality Development*. John Wiley and Sons.

- Tomasello, M. & Vaish, A. (2013). Origins of human cooperation and morality. *Annual Review of Psychology*, 64, 231–255.
- Wood, A. (1998). Kant on Duties Regarding Nonrational Nature. *Supplementary Volumes Proceedings Aristotelian Society*, 72 (1), 189-210.
- Wood, A. (2008). *Kantian Ethics*. Cambridge University Press.
- Wood, A. (2011) Humanity as End in Itself, in Parfit, D. (ed.) *On What Matters: Vol. 2*, Oxford University Press.
- Wood, A. (2017). *Formulas of the moral law*. Cambridge University Press.



## Shame and ‘Shame Instinct’ in Kant’s Pre-Critical Texts RH\*

Ana Cristina Falcato 

Research Fellow in Philosophy at IFILNOVA, The New University of Lisbon, Spain. Email: [anafalcato@fcsh.unl.pt](mailto:anafalcato@fcsh.unl.pt)

### Article Info

**Article type:**

Research Article

**Article history:**

Received 07 May 2024

Received in revised form  
24 June 2024

Accepted 10 July 2024

Published online 07  
August 2024

**Keywords:**

Kant, shame, state of  
nature, shame instinct,  
Bernard Williams.

### ABSTRACT

This paper corrects a historical injustice that has been perpetrated against Kant for some time now. Mostly on good grounds, Kantian ethics has been accused of neglecting the role played by the emotions in moral deliberation and in morally informed action. However, the contemporary moral philosophers who have put forth such a claim tend to bypass textual sources, on the one hand, and to downplay the role played by the anthropological writings on Kant’s practical philosophy as a whole, on the other. Relying on highly relevant pre-critical texts in which Kant sketches future argumentative patterns and discusses the role of a negative emotion like shame on the improvement of the human species, I address a mistaken conclusion about Kantian ethics as a whole that is common in contemporary Anglo-American philosophy. I also raise paradoxical conclusions that follow from Kant’s argument, once its implicit premises have been brought to light. I conclude that Kant did indeed think seriously about a so-called ‘shame-instinct’, however much his central ideas diverge from contemporary readings of the emotion, and fall short of fulfilling the ultimate target one can assume his insights would be drawing at.

**Cite this article:** Falcato, A. C. (2024). Shame and ‘Shame Instinct’ in Kant’s Pre-Critical Texts RH. *Journal of Philosophical Investigations*, 18(47), 265-280. <https://doi.org/10.22034/JPIUT.2024.62394.3812>



© The Author(s).

<https://doi.org/10.22034/JPIUT.2024.62394.3812>

Publisher: University of Tabriz.

\* This text is a reprint from my 2021 published essay, ‘Shame and ‘shame instinct’ in Kant’s Pre-Critical texts’. The text was originally published at *Cosmos and History: The Journal of Natural and Social Philosophy* (Cosmos and History: The Journal of Natural and Social Philosophy, vol. 17, no. 1, 2021), and is now reprinted with the generous permission of the Editors – especially Professor Arran Gare, to whom I am grateful.

## Introduction

### On and Beyond Contemporary Approaches to Kant's Thought

As pretty much every reader of Kant's practical philosophy will ascertain, there are two historical figures that have strongly impacted on the development of the complicated, strict and highly demanding account of morality our philosopher ends up delivering, first in the *Groundwork* (1785), and then, three years later, in the *Critique of Practical Reason* (1788): Jean Jacques-Rousseau and David Hume. As targets, sources of inspiration or, more often, both, these two giants of the European Enlightenment branded Kant's philosophy in ways not always easy to reconstruct. It is also famously the second-best anecdote about the modesty and iconoclastic spirit reigning in Kant's life as well as in his dwelling, the fact that in it – so the story goes –, only a small portrait of Jean-Jacques Rousseau was to be seen hanging on a wall.

In spite of its crucial debt towards Rousseau's moral views, Kantian ethics has often been accused of neglecting, or – even worse – of *condemning* the role of emotions in moral action. There are some remarkable exceptions to this prejudice in both Kantian and anti-Kantian scholarship produced nowadays, such as the admirable work carried out by Krista K. Thomason and Nancy Sherman in the US (Sherman, 2014; Thomason, 2013, 2015), by Inge Römer in Germany (Römer, 2014) or by Núria Sanchez-Madrid in Spain (Madrid, 2018, 2021). But well-known moral philosophers working mainly in the analytic school in the second half of the last century were extremely keen on systematizing traditional moral outlooks (such as Kant's), without a close eye on the evolution of the textual sources documenting those views. This was the case of Bernard Williams, who's inspiring thematization of the practical role of emotions motivates the present discussion.

In an often-quoted passage from *Shame and Necessity* (Williams 1993, 77), Bernard Williams accuses Kant – or at the very least a hypothetical Kantian philosopher – of utterly dismissing the role of shame in human relations in general, and especially within a framework for moral decision-making which stresses the universality of such a procedure, as against individual character traits or personal choice. The fundamental passage from *Shame and Necessity* on (but also against) which I will anchor my own reflections in this essay reads as follows:

In the scheme of Kantian oppositions, shame is on the bad side of all the lines. This is well brought out in its notorious association with the notion of losing or saving face. 'Face' stands for appearance against reality and the outer versus the inner, so its values are superficial; I lose or save it only in the eyes of others, so the values are heteronomous; it is simply my face to save or lose, so they are egoistic (Williams 1993, 77).

As structurally accurate as this brief description of the experience of shame may appear, there is not, in the whole of *Shame and Necessity*, a shred of textual evidence that this is Kant's genuine

view on the matter. Williams never quotes Kantian sources; he does not even allude to them. However, an accurate examination of some crucial texts in the Kantian *corpus*, on the one hand, and a deep analysis of Williams' complex reasoning, on the other, disclose a much more nuanced treatment of the topic – by both philosophers.

I shall thus first explore Kant's discussion of the role of shame in the evolution of the relationship between the sexes – in particular, the *historic* and *genetic* development of a complex line of reasoning that Kant seems to have first presented in his courses on ethics, dating from the beginning of the 1760s (Kant 1997, 48; 27-52; 22). When faced with certain perplexities that we can assume were roused in him by the reading of Rousseau's *Discourse on Inequality*, Kant confronts his students with a metaphysical hypothesis that accounts for a way out of the state of nature and which involves exploring the most primitive dynamics of the shame instinct. I present this hypothesis in Section 1.

Section 2 furthers the analysis of this problem in Kant's early texts with a comparative approach to the structural expansion of this model, applied to a justification of the ultimate purpose of the relationship between the sexes. By displaying a multi-perspectival description of how the intricate dynamics of this primitive shame instinct supposedly works in relations between men and women, I end Section 2 with a note on the paradoxical conclusions to which it gives rise when compared with contemporary views on shame, understood as a moral emotion. The essay's ultimate point is to show how, properly understood, the feeling of shame represents for Kant a self-sufficient drive for preservation that is independent of each agent's will and that strives for the strengthening of the human species as a whole.

### 1. Continuities and Breaks in Kant's Anthropological Texts

Reading Kant's pre-Critical texts today, it is hard to avoid the impression that the very evolution of his thought anticipates concepts and characteristic patterns of reasoning that only the mature works would bring to fruition.

One such reasoning model, which displays a recognizable teleological structure while describing the sheer appearance of phenomena, can be detected in the second section of *Observations on the Feeling of the Beautiful and Sublime* (1764) and in some lecture notes on moral philosophy, world history or physical and pragmatic anthropology, also dating originally from the 1760s. This model traces a deep connection between one crucial end of nature – the propagation of the human species – and a sense of shame. However, Kant falls short of providing an argument for this link. Accomplishing this will be the main purpose of the present essay.

To frame my analysis of the connection Kant draws between the shame instinct and human propagation, I begin by looking at a well-known set of lecture notes taken by Johann Gottfried Herder early on in Kant's career. Herder attended Kant's lectures at the Albertina University in Königsberg between 1762 and 1764. Although in later years he would strongly dispute Kant's most

important views on history and the role of rationality as the ultimate source of normativity, in his student days Herder considered himself a disciple of Kant.

Reading the many surviving notes from Kant's early courses, one can at least infer that Herder attended the lectures on Ethics – or *Moralphilosophie*, as they were commonly announced and scheduled in Königsberg. It was the German scholar Paul Menzer who, in 1924, (Menzer 1924) on the occasion of the bicentennial of Kant's birth, first compiled the notes from those courses, and despite certain discrepancies between the Menzer edition of *Eine Vorlesung Kants über Ethik* and the collection of remarks attributed to Herder, some of the topics discussed in both texts clearly overlap. The range of problems discussed in what is now commonly referred to as the 'Herder Lecture Notes' (Kant 1997, 1-37) is impressive, ranging from the difference between physically good actions and morally good ones, the nature of moral motivation, atheism in *sensu privationis* and *sensu contradictoriae*, and the natural evolution of the sexual impulse (Menzer 1924, § 27, 48 ss).

It is precisely when speculating on the evolution of the sexual drive, from the state of nature to its present form in civilized society, that Kant refers to a basic shame instinct (a 'true shame instinct', as he calls it), which not only performs a bridging role between the two phases of humanity's development but even justifies the attainment of the latter and its perfectible teleology. Here is what Herder reports Kant as having said in his public lectures:

The sexual impulse would not have developed so early, but once the powers of the body had matured, for it would not have been accelerated by instruction. The impulse satisfied itself merely by immediate pleasure, and there would probably not have been a permanent bond. But since, no doubt, the man would have felt that the impulse would recur, he would allow the woman to follow him into the forest; she became his companion, and both would have cared for the children. He would have had to help her while she was suckling them, and thus arouse monogamy, since there are as many women as men. The impulse would not have been so rampant then, since the fantasied pleasures of the civilized were lacking. Moreover, this impulse is covered with the veil of shame, which is also found among the majority of savages, and is quite unlike any other form of shame and restrains the impulse.<sup>1</sup> There is much truth in the objections of the cynic: we should be ashamed only of what is dishonourable; but for all that, there is a genuine shame instinct, which has indeed no rational cause, and is strange, but whose aims are, 1) to restrain the untamed sexual impulse; and 2) to maintain the attraction of it by secrecy. The male sex, which has more principles, possesses this shame in a lesser degree; for want of principles, the woman has a

---

<sup>1</sup> My italics.

great deal of it, and it dominates her; and where this shame has already been uprooted in women, all virtue and respectability have lost their authority, and they go further in shamelessness than the most dissolute of men. Such shame, moreover, has an analogon with an act that is intrinsically dishonourable, and this has produced the stupid shame of monkishness. It is not, however, in itself the mark of an unpermitted act, but the veil of an honourable one, which propagates mankind. (Herder 1997, 22).

The train of Kant's reasoning is not altogether clear, and some reconstruction is needed if we are to gain a firm grasp on the inner logic of his position and a sense of its paradoxical outcomes. One point that Kant is certainly trying to put across concerns the establishment of monogamy in civilized societies, by sketching out a sort of metaphysical story about the first encounters of members of the two sexes, linked with the persistence of the sexual impulse and a supposedly equal number of members of each sex. But this is not the whole purpose of this speculative account of the state of nature and its evolution into a civil state; it is only a stepping stone. In Kant's complex reflection on the ultimate purpose of the association between men and women, the establishment of monogamy is but a primary condition for the propagation of the species and, one might safely infer from many other texts in the Kantian corpus besides the *Vorlesungen*, its perfectibility throughout the generations.<sup>1</sup> Monogamy, Kant believes, is the first condition for a steady increase in the survival rate of the offspring of natural man and woman and, *a fortiori*, a condition for the improvement of their descendants.

As it turns out, the lengthy remark recorded by Herder on the connection between shame and the maturation of the sexual impulse in the history of humankind first introduces a reasoning pattern, which Kant will later<sup>2</sup> identify as the enabling principle for the satisfaction of the ultimate purpose of the encounter between members of the two sexes. *Qua* form of restraint, Kant will argue, shame is said to creep along intersexual relations as a means of disciplining the sexual drive, taming sensual excesses, and ensuring that the human species moves forward.

When setting out to identify the most immediate and apparent features of the forms of relationship that men and women establish and maintain with each other, Kant opts for a twofold kind of description: on the one hand, he points out how behavioural norms come about, develop and become conventions among members of each sex; on the other, he attempts to figure out what their ultimate meaning and purpose might be.

<sup>1</sup> See especially his *Observations on the Feeling of the Beautiful and the Sublime*, Third Section, 2: 236: 'This whole enchantment is at bottom spread over the sexual drive. Nature pursues its great aim, and all refinements that are associated with it, however remote from it they seem to be, are only veils, and in the end derive their charm from the very same source. [...] If this taste is not exactly fine, still it is not on that account to be despised. For the greatest part of humanity follows by its means the greatest order of nature in a very simple and certain manner. By this means most marriages are brought about, and indeed among the most industrious part of humanity [...]' (Kant 1997: 46).

<sup>2</sup> See Kant 2007: 18-63.

In Section Three of *Observations on the Feeling of the Beautiful and the Sublime* ('Of the Distinction of the Beautiful and Sublime in the Interrelations of the Two Sexes'), published at the end of the timeframe during which Herder might possibly have taken down his course notes, Kant writes the following on the 'sense of shame':

The sense of shame is a secrecy of nature aimed at setting bounds to a most intractable inclination, and which, in so far as it has the call of nature on its side, always seems compatible with good, moral qualities, even if it is excessive. It is accordingly most necessary as a supplement to principles, for there is no case in which inclination so readily becomes a sophist cooking up complaisant principles as here. At the same time, it also serves to draw a secretive curtain before even the most appropriate and necessary ends of nature, so that too familiar an acquaintance with them will not occasion disgust or at least indifference with respect to the final aims of a drive on to which the finest and liveliest inclinations are grafted. (Kant, 2011, 41)

This excerpt presents us with an extremely complex reasoning, which only a careful, step-by-step analysis can properly unravel. The attentive reader, however, is left with no doubt that here, too, Kant is tackling a massive theoretical issue, which relates the natural-historic taming of the sexual drive to the propagation of the species. The most important point to be elicited from this strange passage is this: the shame instinct works as a natural bridge between the two purposes ultimately ascribed to the satisfaction of the sexual impulse that brings together human beings of the two sexes. If there is a sexual force attracting men and women to each other, there is also a further drive which takes sex beyond sex, and members of each sex beyond the peculiarities of their circumstances, and that is the very 'drive of the species', which is set to continue and improve itself.

Needless to say, such a supra-individual force of perfectibility, if at all meaningful, falls beyond the reach of empirical ascertainability, and its explanatory role cannot but be inferred from other, observable phenomena – in the present case, from the behavioural pattern which sets itself between the members of each sex. Kant seems to have been extraordinarily impressed by the fact that one purpose of this pattern of attraction is immediate and inconsequential, whilst the other is noble and world-historical, and by the fact that it is shame that mediates between the two.<sup>1</sup> Carefully

---

<sup>1</sup> Kant picks up the basic thread of his reasoning from Rousseau's *Discourse on Inequality*. As usual, Rousseau's influence on Kant's moral thinking cannot be dismissed, but their respective ideas on this point are almost at loggerheads. Rousseau wants to claim that, since in the state of nature the cultural elements of the fundamental human drives – including the sexual drive – are absent, the impulse is quickly satisfied and dies out. In the long run, it is hard to see how, given this framework, the species could have evolved. Kant, however, will insist not only that the spontaneous meetings occur more and more often, but that somehow the performance of the species overcomes the limitations of individuals, and that, through *shame*, the immediate satisfaction of the drive is both restricted and rationally linked with the continuation of humankind beyond the natural state (See Rousseau [2002] and Kant [2011]).

examined, the supposed sensible features characterizing – or even determining – a feeling or impression of shame, particularly in connection with the sexual drive, are overruled by their structural design of taming a more primitive sensual experience. Although labelled 'an instinct', Kant's peculiar account of shame turns the emotion into a form of self-control in the face of a most overwhelming expression of a natural drive, and thus into a pre-condition to virtue. However much one wants to insist on the sensible nature of this form of self-restraint, its functional role is to curb a much more undisciplined human instinct – one we share with non-human animals –, and one can thus witness in this peculiar genetic interpretation of the sexual impulse a sensible feeling enabling moral behaviour<sup>1</sup>.

It is precisely this added twist in Kant's genetic story about the maturation of the sexual drive, however, which will begin to expose his reasoning to paradoxes. I now turn to these paradoxes and to possible ways of resolving them.

## 2. Irreconcilable Viewpoints

As I hope my explanation of the evolution of Kant's thoughts about the disposition of men and women in the state of nature has made clear, it is when Kant goes from speaking about the frequency of sexual intercourse between members of each sex to suggest an explanatory model for the way out of the natural disposition that problems emerge with his hermeneutical hypothesis. I shall recur to the reflective strategy of testing an interpretive proposal from different standpoints to illustrate exactly how the model flounders.

Reconstructing the two long and convoluted remarks quoted in the previous section, one shall stress first that it is the sexual impulse itself that Kant says is covered with the veil of shame. When one interprets the first passage, from the Herder Lecture Notes, with a view to unfold its rationale, what is there described as 'a sense of shame' appears to work thus: by restraining the naturally ungovernable expression of the sexual drive, the so-called 'veil of shame' precisely allows it to attain its ultimate goal, which is not immediate satisfaction but the propagation of mankind [*die Menschenpflanzung*]. However, one should bear in mind throughout that Kant believes this to be a kind of phylogenetic inheritance from our forebears, and thus this peculiar type of experience should map out both the way the sexual drive is felt by the uncivilized man and the way we

---

<sup>1</sup> It is crucial to bear in mind, though, that such form of strengthening or improving of the species as a whole has no relation whatever with what Kant usually means by moral improvement or perfectibility and which presupposes a detachment from sensible feeling and a strict grounding of moral action on strict rational principle. My reasoning here does not attempt to make of Kant an empiricist (or even a sensualist) about the ultimate goal of human evolution or its subjective take-off in relation to sexual attraction and a reproductive pattern this puts in place across generations – starting with a primeval pull to get man out of the state of nature. All I am emphasizing is the intelligible design of this primitive impulse drawing members of the two sexes to each other. In that regard, such thing as the 'moral development of the human species' is akin, first, to species survival and, in the second place and ideally, to the betterment of its members throughout time.

experience the drive in a civilized state<sup>1</sup>. Once this pattern is analysed in detail, however, structural problems immediately stand out.

For one can reason as follows. From an internal, strictly subjective point of view, how can we so much as live through the curbing of the sexual impulse – mostly experienced as one peculiar kind of social restraint –, as ‘disgust’, ‘indifference’, ‘moving away’ (and thus, still as a point of contact or articulation) with the final design of that drive, which is entirely rational and concerns the sustained propagation of the species? One is usually capable of both thinking about the continuation of the human species as one of the ends fulfilled by reproduction and of experiencing the restriction of the sexual impulse as both a natural and a cultural necessity. But it hardly encompasses our experience of a limited, organized, socially framed sexual expression that it be a function of a higher purpose which, on top of everything, and being by nature *expansive*, precisely contradicts a restrained form of activity.

However, an attempt at describing this very situation in purely external terms does not fare much better. When reconstructing the fundamental moments in Kant’s argument, we are able to understand at least this much: the original force of the sexual drive must be curbed, so that its ultimate purpose can be accomplished. Kant alleges that an effective sense of shame is what performs this dual role of i) keeping the impulse within bounds and ii) ‘maintaining its attraction by secrecy’. However, and conceiving of the experience in its functional, dual-aspect role – as an inner feeling with a given external configuration pointing at a teleological design which goes beyond experience itself – one further question begins to take shape: how is *shame* supposed to do this? How can a deeply first-personal negative emotion like shame perform this ambitious role of bringing together the natural and the rational dimensions of a primitive disposition in us?

If one looks at available theories of shame (let alone first-hand experiences), ranging from first-personal approaches to the emotion as described by twentieth-century phenomenology, (Sartre (2003) and Max Scheler (1913).) contractualist discussions of its impact on an individual’s self-esteem, such as the Rawlsian model presented in Part III of *A Theory of Justice*,<sup>2</sup> or to the fast-

<sup>1</sup> To confirm the uniformity of this notion in Kant’s mind, both when speculating on the origins of basic drives in ‘Natural Man’ and when settling on a view of the same issue regarding human beings living in a state of civilization, we need but look at the similar note recorded in *Observations on the feeling of the beautiful and sublime*, published in 1764 (at the end of Herder’s stay in Königsberg), which accounts for the general features of the relation between members of the two sexes strictly from a phenomenal point of view.

<sup>2</sup> See Rawls 1999: 388-389, § 67: ‘We may characterize shame as the feeling that someone has when he experiences an injury to his self-respect or suffers a blow to his self-esteem. [...] Shame is the emotion evoked by shocks to our self-respect, a special kind of good. Now both regret and shame are self-regarding, but shame implies an especially intimate connection with persons and with those upon whom we depend to confirm the sense of our own worth’. Rawls thinks about shame at a very peculiar conceptual juncture. He claims that human virtues are *excellences*, i.e. personal attributes that increase interpersonal common good, and he furthermore takes self-respect to be a basic kind of good, essential for the flourishing of any individual agent. Shame, to the extent that it constitutes a blow to self-respect – the fundamental feature of personal integrity – threatens the possibility of social cohesion.



growing specialized literature from moral psychology and the empirical sciences broadly construed, one sees that shame is uniformly characterized as a deeply personal emotion that damages the core of the self.<sup>1</sup> Even Bernard Williams' extremely nuanced analysis of this emotional pattern, whose main purpose is to disassociate the experience of shame from a purported Kantian critique of moral superficiality, matches well this intuitive description.

Perhaps the most acute criticism directed by Williams at contemporary advocates of Kantianism is that modern readings of Kant's practical philosophy in which the structural *moral significance of shame* is rejected (in which shame, often stemming from involuntary action, is lumped together with a recognizable cluster of subjective responses that differ fundamentally from duty or action springing from sheer respect for the moral law) function unselfconsciously against the distorting background of the Kantian moral paradigm, which is taken as *given*.

Williams's critique of the moral context that privileges the absolute autonomy of the agent's will has two fundamental stages. Within the Kantian deontological framework – so his argument goes –, the autonomy of the moral agent is defined in terms of her immunity to the judgements of others in practical decision-making, i.e. her ability to disregard such influences and instead act on principle. Autonomous freedom of choice is set against an adaptive model following which each agent determines her actions with a view to meeting external expectations. Because the experience of shame is mainly connected with the disadvantageous exposure of oneself to the gaze of another, the values at stake are superficial, and superficial in a double sense. To the extent that the basic experience is related to the physical, to appearance, it concerns the appraisal of external features. For precisely this reason, the moral relevance of shame can be criticized: shame does not concern the intrinsic value of the person and indeed replaces true value with what lies on the immediate surface—the first impression. If this were so, Williams continues (imagining the Kantian line of argument he wants to deconstruct), the experience of shame would indeed be linked to heteronomy and lack true moral value, since the self-directed adverse judgement involved in shame brings with it awareness of both the critical gaze of the other and how I appear in the light of that gaze, but nothing more. On the other hand, and because the impression of disadvantage is merely subjective, the experience might be said to lack the universality that defines the Kantian normative model, and might therefore be dismissed as *egoistic*. In fact, the two defining features tend to complement each other. An action is classified as egoistic if it is structurally motivated by the individual's desire for success and acclaim, and, it can be argued, such a practical pattern makes sense only given an environment in which the agent believes she will be seen and possibly receive acclaim. To be sure: agents can stipulate for themselves ends they would like to achieve, independently of third-party assessments or prescriptions. But even internalized expectations of personal fulfilment are learned and structured as patterns of self-assessment in the context of appearing in a world shared with

---

<sup>1</sup> See Taylor (1985), Sartre (2003), and Thomason (2013, 2015).

others. What becomes in some sense inner—a part of my personal moral makeup, the backdrop against which I assess myself and my progress in the world—has its psychological roots at the level of my appearance among others.

Now, if one goes back to Kant's direct pronouncements on the topic of shame – which Bernard Williams never mentions –, it is extremely difficult to relate the first, naturalized function, highlighted by the quotations above, and which Kant (alongside many anthropologists, phenomenologists and moral philosophers) ascribes to the emotion with the supposedly noble, teleological aim, which that first, intuitive role should pinpoint and which concern the betterment of the human species. For unless one can prove either that Kant is using the notion here with a double connotation – objective and subjective, let us say, or external and internal – or that he is dealing with a *sui generis* notion of 'shame', it is going to be hard to make a case for this interpretative model once its rationale is fully understood.

It might be useful to insist, once more, that this complex scenario admits of three separate descriptions. There is, first of all, a reflective exercise one must undertake in order to grasp the inner logic of Kant's association of an ungoverned sexual drive, which is dominated by shame, with the fulfilment of its ultimate end – an accomplishment that is only reached to the extent that shame curbs the instinct, thus enabling propagation. In a way, this effort objectifies the content of Kant's ideas about the problem in the sense of making it fully explicit. A second description focuses on the purely subjective experience of this supposed association between the two aims of the impulse (a person feels ashamed of his or her sexual appetite and tames it, immediately unaware that she is thus fulfilling its ultimate evolutionary purpose). And finally, this inner experience or set of related experiences matches a kind of behaviour in the individual arguably targeted by this form of shame, which we can also describe, but *hardly as shame behaviour*, as long as one respects the two main features suggested by the first description, i.e., restriction of the drive plus a rational justification of its ultimate design<sup>1</sup>. The latter, being empirically unverifiable, demands the stretching of our rational powers, which enables us to make sense of a natural phenomenon.

Now, having unpacked the main elements structuring Kant's complex reasoning, I would like to specify where my hermeneutic difficulties ultimately lie, when one brings together the external and the internal features of the phenomenon under analysis.

---

<sup>1</sup> In the *Metaphysics of Morals* Kant will again discuss a two-pronged approach to the morally flawed action and underline the duty to expose its shameful aspect as against its harmfulness to the agent him- or herself. The brief assessment of the shameful aspect of the morally bad action (to the effect that its shameful aspect has precedence over its badness whenever the education of the moral sense is at stake – 1996a: 595; MS 6: 483) contrasts with the notes on the shame-instinct I have been exposing thus far in two major ways. At the level of emotional self-discipline, one has, in the latter example, a form of third-personal control and policing over the morally bad behavior as against a first personal, natural mechanism of instinct curbing and restraint. On the other hand, it is the external, morally defective aspect of the shameful act, which is exposed and expunged with a view to educate the moral sense in young people as against the inner control over the excesses of sexual inclination.

It is certainly conceivable that the kind of experience Kant describes as (i) restricting an untamed sexual impulse and (ii) maintaining the attraction to the impulse via secrecy, is observable as shame behaviour. As a matter of fact, the two features might well encompass, in broad terms, a certain awkwardness with which human beings live out their sexuality in different social settings. At the same time, it is at least plausible that this can be seen as a typical form of self-restraint characteristic of shame feelings. What is in any case counter-intuitive is adding to this form of shyness the functional role that Kant claims it is meant to perform – the higher goal of propagating the species – while retaining the necessary features that define experiences of shame.

Should Kant's own thought experiment seem too grand, or my exposition thus far too abstract, think of the following, quite common, first-personal reflection. Kant's ingenious line of reasoning concerning the fulfilment of the 'great design of nature' would seem to find support in the calculated planning involved in crafting policies, whether in the home or at the state level, concerning future generations. Individuals, couples, states and corporations make and participate in birth control programs, design laws and social plans to increase (or decrease) birth-rates, rely on fertility cycles, even shape entire aspects of work and social life in accordance with the limited potential modern life allows for reproduction – and this may imply the effective channelling of sexual energy to the right moments.

But it is hardly imaginable that any sort of prudential measures springing from thoughts akin to the one Kant expounds in both texts encompasses anything even remotely associated with a shame instinct. There may be self-imposed restrictions, and typical, primitive impressions of shame in relation to the sexual act may equally occur, but the very plan of leaving progeny will remain unaffected by both.

There may be objections to this brief account of how the shame instinct operates in Kant's framework, and I alluded to some of these above. One might think, for instance, that Kant is here dealing with a rather technical – or circumstantial – notion of 'shame' (more precisely, a 'shame instinct'), and that applying it to common forms that the emotion assumes in daily life is simply out of place. At least the first half of this objection is accurate, and it is indeed pointed out by Kant himself.<sup>1</sup> It does not invalidate my reconstruction of the argument, however, and this for two reasons. On the one hand, the objection is unsatisfactory because although the concept Kant is treating is *sui generis* and somehow *ad hoc*, its components are those I have stressed above, and in at least some of its applications, as my above example shows, their conditions for truth are mapped out by very primitive shame experiences with which we are all familiar. On the other hand, there

---

<sup>1</sup> If one is to trust Herder's notes, it seems likely that Kant was aware of the peculiarities surrounding the experience of shame in relation to sexual intercourse. See esp. Kant 1997: 27:99: '[...] there is a genuine shame-instinct, which has indeed no rational cause, and is strange...'.

are other contexts – for instance, when relating shame to the privacy of prayer<sup>1</sup> – where Kant deploys an idea of shame that precisely maps the definition laid out for us by moral psychologists and phenomenologists. Therefore, even in purely conceptual, non-empirical terms, Kant certainly had the notion that modern moral theories<sup>2</sup> commonly endorse.

One could further object that what Kant in fact proposes is a functional account of how a process which is based on experience becomes ultimately intelligible only once reason is stretched beyond experience in order to make full sense of it, and thus that any appeal to an empirical order which justifies it is question-begging and any explanation of its operative dynamics from distinct viewpoints utterly off the mark. I still find fault with this objection, however, and again for two reasons. For even if what Kant is in fact providing is a kind of a-perspectival explanation of a very complex, natural and historical phenomenon which only a fiction of reason enables us to fully grasp, nothing prevents the modern interpreter of these lecture notes from shaping their most abstract ideas from the point of view of individual agents. In fact, when one frames Kant's *modular experience of the shame instinct* in terms of singular perspectives, common features of our personal and interpersonal acquaintance with it immediately stand out. The structural articulation of its parts by no means functions *only* as a rational fiction.

It seems indeed highly plausible that, if anything in the structure of Kant's reasoning, as presented both in the *Lectures on Ethics* and later in the *Observations*, works to block full intelligibility of its composing elements when brought together under a unified standpoint, it is the supra-individual justification of partial features of individual experience. Thus, if Kant's rational account of a chain of natural phenomena fails, this is due to the unwarranted confluence of its explanatory layers rather than the presumed partiality of descriptive points of view. As stressed above, his full explanatory model is altogether general and supra individual, but parts of it only make sense once framed from personal standpoints.

Finally, and precisely on account of the reconstructive effort taken to explain how exactly it is that Kant tries to relate two dimensions of a single phenomenon with the help of a third notion – a phenomenon which corresponds to one of the most primitive forms of human experience – an objection (or perhaps merely a query) to the effect that he might be using the notion with distinct subjective or objective/internal or external interpretations is blocked. Kant has a persistent but undeveloped idea of how the shame instinct operates in order to accomplish all that it is designed to accomplish. Once the hidden pieces of this puzzle are brought into the open, problems emerge:

---

<sup>1</sup> See *Moral Philosophy Collins*: 'The more upright a man is, the more readily he is ashamed if surprised in an act of devotion. A hypocrite will not be ashamed, but on the contrary, will let himself be seen. For a man is ashamed if another thinks ill of him, even though he has committed no fault' (Kant 1997: 27: 337, 119).

<sup>2</sup> The more robust contemporary accounts of shame as a moral emotion – aside from Williams's –, undertaken from the standpoint of its intersubjective and subjective impact upon us are, respectively, those provided by John Rawls (1999) and Gabriele Taylor (1985).

not only is it extremely difficult for the individual to experience restraints on the sexual impulse as a means of expanding the human race, but it is also counterintuitive to associate this behavioural pattern with the emotion of shame, intuitively understood and phenomenologically described.

No matter how complex and ultimately paradoxical Kant's interpretation of the shame instinct turns out to be, however, there is certainly a highly structured explanatory framework to be found behind these undeveloped associations. Following their lead gives us a genuinely *Kantian* view of the emotion of shame – a view which is true to the letter and not just the spirit of what acting morally 'in the scheme of Kantian oppositions' amounts to. If he had taken a more congenial approach, this aspect of Kant's thought, put forth in some crucial writings, could also have been encompassed by Bernard Williams in *Shame and Necessity*.

## Conclusion

In the present essay, I have tried to put an end to two 'hermeneutic myths' that are deeply ingrained both in popular accounts of Kantianism and in certain trends in Kantian scholarship itself. One of these concerns the supposed lack of relevance of the emotions to Kant's practical philosophy, broadly construed. The second myth concerns an often-assumed theoretical and formal *hiatus* which is taken to separate the pre-Critical from the Critical texts.

Bernard Williams – who never presented himself as an advocate of Kantian doctrines – is probably most responsible for spreading the first prejudice against Kantian ethics in twentieth-century Anglo-American philosophy. Always keen to take up philosophical challenges head-on and to restrict his discussions of Kantianism as a moral system to the better-known treatises, i.e. the *Groundwork* and the *Critique of Practical Reason*, Williams omits from his interpretive model a wide array of reflections on anthropology, history, religion, pedagogy, and the natural sciences, in which Kant never loses sight of moral considerations. As it happens, the longest extant textual discussion of the role of shame in human history can be found in notes from the courses that Kant gave on the subject of ethics for more than three decades.

In the first half of the 1760s, Kant clearly based part of his teaching on the most important pieces Rousseau had written the previous decade – even picking up the same anthropological models and specific examples – while putting forward solutions to the enigmas at the root of morality that were almost the opposite of Rousseau's.

The astute reader of the essays and notes from this period will identify the emergence of the more personal formal features which define the Kantian mode of explaining both natural and moral phenomena – a rational and finalistic mind-set – even when he covers the slippery ground of attempting to justify the development of a basic affective disposition in 'Natural Man'. It is of course at this problematic juncture that the idea of a primitive instinct of shame is first discussed, and we have seen that Kant was somewhat obsessed with discovering its ultimate purpose in the

history of humankind. One is well justified to infer that, if the problems are Rousseauian, the solutions bear a distinctly Kantian mark.

At this point in the evolution of his philosophical system, which would be fully consolidated only within two decades, Kant deployed reason as a ‘thinking disposition’ more than a separate faculty of the mind. In other words, in the 1760s Kant seemed to conceive of human reason as an indispensable intellectual tool for explaining reality rather than a *thing* with supra-historical powers and attributes.

As such, not only could no natural or metaphysically hypostasized phenomenon remain unexplained – as in the case of the supposed sexual dynamics of the race in the natural state – but Kant wanted to discover, *rationally*, where it led. And to that end, an emotion like shame – with its intrinsic double bind, implying exposure and spontaneous recoil – is instrumental. Thus, the role played by shame in Kant’s reasoning becomes clear once its most primitive features are rationally displaced. Once this process of explanatory displacement is carefully thought through, two things become apparent. One is that, even in his early lectures, Kant begins to sense that there is something more substantial behind animals’ (including human animals) sexual appetite – something like a trans-individual force fighting for the species and oblivious to our will.<sup>1</sup> One logical counterpart of this supra-personal instinct in us, however, is that individuals can just as well be oblivious to the true motives of their sexual behaviour – even when claiming that they are planning to have *their* children.

On all accounts, the idea matters morally to Kant, as does shame, as the enabling emotion underlying it.<sup>2</sup>

## References

- Aramayo, R. P. (2002). *La Cara Oculta del Formalismo Ético, Introduction to Lecciones de Ética*. Barcelona: Editorial Crítica.
- Colegrave, N. (2002). Sex releases the speed limit on evolution. In *Nature*. Vol 420, 664-666. <https://doi.org/10.1038/nature01191>
- Darwin, C. (1859). *On the Origin of Species by means of Natural Selection*. John Murray.
- Deigh, J. (1993). Shame and Self-Esteem: a Critique. *Ethics*, 93(2), 225-245.
- Deigh, J. (2006). Shame and Punishment in Kant’s Doctrine of Right. *Journal of Ethics*, 10(4), 383-418.

---

<sup>1</sup> This feature of Kant’s thought has been widely acknowledged by evolutionary theorists, from Darwin to the present day. See, e.g., Darwin (1859), Maynard-Smith (1978) and Colegrave (2002). For more than twenty years now, the biologist Nick Colegrave has been arguing that sex can speed the rhythm of evolution, helping new organisms to adapt to changing environmental conditions. Although he insists the evolutionary advantage varies with population-size and the structure of the mating organisms, this is one crucial point where some insights of Kantian Anthropology meet modern Theories of Evolution.

<sup>2</sup> Research leading to the writing of this paper is supported by FCT under the contractual programme in accordance with articles 4, 5, and 6 of the Law Decree nos. 57/2016, of August 29, altered by Law no. 57/2017, July 19.

- Falcato, A. (2020). Shame and ideas of oneself: genetic approach and ethical readings. In *Filosofia Unisinos*, 21(2). DOI: <https://doi.org/10.4013/fsu.2020.212.08>.
- Kant, I. (1996a). The Metaphysics of Morals, In *Practical Philosophy*. translated by M. Gregor, Edited by M. Gregor, Cambridge University Press.
- Kant, I. (1997). *Lectures on Ethics*. Translated by P. Heath & J. B. Schneewind. Cambridge University Press.
- Kant, I. (1764). *Beobachtungen über das Gefühl des Schönen und Erhabenen*. AA. II.
- Kant, I. (1759). *Vorlesungen über Moralphilosophie*. AA. XVII.
- Lehmann, G. (1967). Kant's Entwicklung im Spiegel der Vorlesungen'. In *Studien zu Kants philosophischer Entwicklung*. Hrsg. Heinz Heimsoeth, Dieter Henrich und Giorgio Tonelli. Olms Verlag.
- Lehmann, G. (2021). *Einleitung zur Kants Vorlesungen über Moralphilosophie*. AA 4.
- Madrid, N. S. & Failla, M. (2021) *Kant on Emotions*. Walter de Gruyter (Forthcoming).
- Madrid, N. (2018). Controlling Mental Disorder: Kant's Account of Mental Illness in the Anthropology Writings'. In *Knowledge, Morals and Practice in Kant's Anthropology*. Edited by G. Lorini & R. Loudon, Palgrave Macmillan.
- Maynard-Smith, J. (1978). *The Evolution of Sex*. Cambridge University Press.
- Menzer, P. (ed.) (1924). *Eine Vorlesung Kants über Ethik*. Pan Verlag Rolf Heise.
- Rawls, J. A. (1999). *Theory of Justice*. Harvard University Press.
- Rinofner-Kreidl, S. (2014). Neid. Zur moralischen Relevanz einer 'Outlaw Emotion', In Römer, I. *Affektivität und Ethik bei Kant und in der Phänomenologie*, De Gruyter.
- Römer, I. (2014). Reue und Freiheit. Ansatz zu einer Phänomenologie der Freiheit im Ausgang von Kants Begriff der Reue. In *Affektivität und Ethik bei Kant und in der Phänomenologie*. Edited by R. Inge. De Gruyter.
- Römer, I. (Eds.) (2014). *Affektivität und Ethik bei Kant und in der Phänomenologie*. Walter De Gruyter. <https://doi.org/10.1515/9783110345162>.
- Rousseau, J. (2002). *The Social Contract and the First and Second Discourses*. Edited by S. Dunn, Yale University Press.
- Sartre, J. P. (2003). *Being and Nothingness: An Essay in Phenomenological Ontology*. Translated by H. E. Barnes. Routledge.
- Scheler, M. (1913). *Über Scham und Schamgefühl*, in *Schriften aus dem Nachlass, Zur Ethik und Erkenntnislehre*. vol. 1 Francke Verlag.
- Sherman, N. (2014). The Place of Emotions in Kantian Morality. In *Kant on Emotion and Value* pp. 11-3, Edited by A. Cohen, Palgrave Macmillan.
- Sussman, D. (2008). Shame and Punishment in Kant's Doctrine of Right'. *Philosophical Quarterly*, 58(231), 1-19. <http://dx.doi.org/10.1111/j.1467-9213.2007.530.x>
- Taylor, G. (1985). *Shame, Pride and Guilt*. Clarendon Press.
- Thomason, K. K. (2013). Shame and Contempt in Kant's Moral Theory. *Kantian Review*, 18(2), 221-240.
- Thomason, K. K. (2015). Shame, Violence, and Morality in *Philosophy and Phenomenological Research*., 1, July 2015.
- Vaihinger, H. (1911). *Die Philosophie des Als Ob. System der theoretischen, praktischen und religiösen Fiktionen der Menschheit auf grund eines idealistischen Positivismus*. Reuther und Reichard.

- Williams, B. (1973). Morality and the Emotions. In *Problems of the Self*, pp. 207-229, Cambridge University Press.
- Williams, B. (1993). *Shame and Necessity*, University of California Press.
- Williams, B. (2011). *Ethics and the Limits of Philosophy*, Routledge Classics.
- Williams, H. (2003). *Kant's critique of Hobbes: Sovereignty and Cosmopolitanism*, University of Wales Press.



## Maturity, Freedom of Thought, and Emancipation — On Kant’s *What Is Enlightenment?*

**Dennis Schulting**

Independent Scholar, University of Warwick, United Kingdom. Email: [ds196901@gmail.com](mailto:ds196901@gmail.com)

| Article Info                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | ABSTRACT                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p><b>Article type:</b><br/>Research Article</p><br><p><b>Article history:</b><br/> Received 20 June 2024<br/> Received in revised form 28 June 2024<br/> Accepted 28 June 2024<br/> Published online 07 August 2024</p><br><p><b>Keywords:</b><br/> maturity, enlightenment,<br/> freedom of thought,<br/> emancipation</p>                                                                                                                                                                         | <p>In this essay, I want to address two main aspects of the arguably central topic of Kant’s treatise on Enlightenment, namely maturity: these concern the notion of the freedom of thought (Section I) and the idea of emancipation that is conveyed by maturity, the fact that it involves a process of growing up to become a citizen (Section II). Freedom of thought denotes the idea of self-agency which all human beings possess in principle whereas emancipation points to the fact that maturity is something that is not a matter of course, neither in individual cases nor for society at large. Maturity is not a natural disposition that one either possesses or not but must be developed. This means that not all human beings are de facto mature always, everywhere and all of the time. That this situation is not something that is in principle unalterable can be demonstrated by way of assessing Kant’s apparent views on women’s presumable minority ‘according to sex’.</p> |
| <p><b>Cite this article:</b> Schulting, D. (2024). Maturity, Freedom of Thought and Emancipation— On Kant’s <i>What Is Enlightenment?</i> <i>Journal of Philosophical Investigations</i>, 18(47), 281-302.<br/> <a href="https://doi.org/10.22034/jpiut.2024.18255">https://doi.org/10.22034/jpiut.2024.18255</a></p>                                                                                                                                                                                |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| <div style="display: flex; justify-content: space-between; align-items: center;"> <div style="display: flex; align-items: center;">  <div style="margin-left: 10px;"> © The Author(s).<br/> <a href="https://doi.org/10.22034/jpiut.2024.18255">https://doi.org/10.22034/jpiut.2024.18255</a> </div> </div> <div style="text-align: right;"> Publisher: University of Tabriz. </div> </div> |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |

*To the courageous and self-thinking women of Iran*

## Introduction

Kant's minor essays are often seen as less important than the major works of the critical period, in particular the three *Critiques*, the *Groundwork* and the *Religion*. This is unfair though. The series of essays that Kant published, in the 1780s and 1790s, in the *Berlinische Monatsschrift*, a well-known monthly in Berlin Enlightenment circles, and once in the *Teutsche Merkur*, represent *in nuce*, and often in a stylistically superior writing style, some of the core ideas of Kant's mature thought. At the same time, in those essays various themes are broached which are not or only tangentially dealt with in the major writings. One theme that implicitly runs through the so-called Enlightenment essay of 1784, *An Answer to the Question: What Is Enlightenment?* (henceforth *Aufklärung*),<sup>1</sup> arguably Kant's most popular and influential work, is the notion of self-determination (*Selbstbestimmung*) or self-legislation (*Selbstgesetzgebung*), a theme that is central to his thought in general. This hangs together intimately with Kant's views on the centrality of the thinking subject and reason in general, that is, thinking itself. Thinking and the rational capacity that each human being possesses, as the measure with which we judge or evaluate anything, is centrally thematized in *Aufklärung* in the context of a question that was posed in an earlier installment of the *Berlinische Monatsschrift*, namely the question: What actually is enlightenment?<sup>2</sup>

Kant's answer to that question diverges from the then as now usual definition or view of Enlightenment in terms of an intellectual movement or historical stage in Western society—a view that Moses Mendelssohn, too, espoused in his own answer to the question, published shortly before Kant's in the same journal.<sup>3</sup> According to Kant, enlightenment is however not about theoretical knowledge or its dissemination or transference, or the increase in rational knowledge as such, which should be seen as a counterbalance to other non-rational views or ideas of life and society, religious or otherwise. Nor is it just about scientific progress, or merely about cultural or societal development. On the contrary, enlightenment primarily concerns *a way of thinking*. For Kant, enlightenment is also always critique, a capacity for critical thinking. Enlightenment is therefore not just *theory*, or the possession of knowledge (in Kant's text in plural: *Erkenntnisse*), but also a certain perspective on or approach to knowledge, how it can be acquired and applied methodically. The practical dimension is in fact primary. This is evidenced by various passages in *Aufklärung*.

<sup>1</sup> Kant's essay was published in *Berlinische Monatsschrift* 4 (Dez. 1784) under the title "Beantwortung der Frage: Was ist Aufklärung?" (Kant 1994).

<sup>2</sup> For a fine account of the historical background of Kant's essay, see Deligiorgi (2005).

<sup>3</sup> Mendelssohn's essay, "Über die Frage: was heißt aufklären?" was published in *Berlinische Monatsschrift* 4 (Sept. 1784), 193–200 (see Mendelssohn 1994).

In this essay, I want to address two main aspects of the arguably central topic of Kant's treatise on enlightenment, namely maturity: these concern the notion of the freedom of thought (Section I) and the idea of emancipation that is conveyed by maturity, the fact that it involves a process of growing up to become a citizen (Section II). Freedom of thought denotes the idea of self-agency which all human beings possess in principle whereas emancipation points to the fact that maturity is something that is not a matter of course, neither in individual cases nor for society at large. Maturity is not a natural disposition that one either possesses or not but must be developed. This means that not all human beings are *de facto* mature always, everywhere and all of the time. That this situation is not something that is *in principle* unalterable can be demonstrated by way of assessing Kant's apparent views on women's presumable minority "according to sex."

### I. Maturity and Freedom of Thought

Let us first look at how Kant defines "enlightenment." "Enlightenment is the human being's emergence from his self-incurred minority," Kant writes. "Minority is [the] inability to make use of one's own understanding without direction from another" (*Aufklärung*, AA 8, 35 [Kant 1999, 17]). The term "minority" or "immaturity" is first used by Kant in his early anthropology lectures. In the Friedländer text from the 1775–76 semester, it is said for example: "The understanding can either be used under the guidance of another, or also without the guidance of another. The first is immature, the other is mature" (AA 25, 541 [Kant 2012, 103]). In the same year in which Kant published *Aufklärung*, in 1784, he again lectures in anthropology; in the Mrongovius notes of the anthropology lectures from the 1784–85 semester, one finds this same-sounding passage: "Immaturity is the inability to use one's own understanding without guidance from another" (AA 25, 1298 [Kant 2012, 412]).

There are a number of core features of Kant's method of analyzing that are, here at the outset of the *Aufklärung* essay, immediately apparent. There is talk of an activity, an exiting (*Ausgang*) from a state for which one is to blame oneself; that is to say, that state is not a state of nature in which one simply and blamelessly finds oneself, without our being able to do anything about it —the English translation of *Ausgang* as "emergence" misses this active aspect, which makes the aspect of culpability more difficult to understand. Kant speaks of an "inability" for which one is responsible oneself, an inability that is "self-incurred" (*selbstverschuldet*). After all, every human being possesses by nature a rational capacity, an intellect, which he is able to make independent use of. One has an epistemic duty to use one's capacity for understanding, a capacity that is innate to every adult human being. To fail using the capacity of our intellect implies an incapacitating of that capacity. It is this failure to conform to an epistemic duty that signifies immaturity.

This epistemic duty is related to an innate *right*, the only innate right that a human being has according to Kant, namely his freedom (AA 6, 237; AA 8, 291–3). Human freedom is directly related to the use a human being makes of his or her rational capacity, the capacity for being

mind. This freedom concerns the will to think for oneself, independently from others. As Kant says, the cause of minority lies not “in lack of understanding, but in lack of resolution (*Entschließung*)” and “courage to use it without direction from another” (*Aufklärung*, AA 8, 35 [Kant 1999, 17]). Enlightenment is thus, put succinctly, having the courage “to use your *own* understanding,” which is underlined by the quotation from Horace: “sapere aude,” dare to know.<sup>1</sup> Notice also that Kant uses exclamation marks: “Sapere aude! Habe Mut dich deines *eigenen* Verstandes zu bedienen!” This means that, first and foremost, the motto of enlightenment rests on a command, an exhortation or imperative to do something, to act from an epistemic duty, rather than a mere description of a kind or body of knowledge. However, this command is not an external command, imposed on us by others. Rather, it is a command that is intrinsically connected to the innate ability to make use of one's own capacity for understanding.

What is at issue here is authority and autonomy. Every human being possesses the freedom to break free from the authority of another—be that authority a person, institution, or tradition, religion, or culture. Each person can make the decision to think for themselves. Of course, that does not mean that the thinking “I” solely relies or can rely on himself as a *source* of knowledge! What matters here is where the ultimate formal authority on which a particular knowledge claim rests lies. That is to say, it concerns the possibility for some knowing subject of “holding-something-to-be-true,” a *fürwahrhalten*—what is not at issue is the fact of *knowing* it to be true or for something to *be* true (see *Critique of Pure Reason*, A820ff./B848ff.; cf. AA 8, 140–1). This subjective authority carries with it a certain epistemic responsibility: one is enjoined to provide reasons for one's holding something to be true. For Kant, that epistemic responsibility resides entirely and solely with each individual human being who makes use of his or her own mind. This constitutes the freedom of a thinking subject.

The connection between freedom and thought is not only mentioned in *Aufklärung* but also in an essay that Kant published some two years later, also in the *Berlinische Monatsschrift*, and which played a role in the so-called *Spinozastreit*, namely, the essay *What Does It Mean to Orient Oneself in Thinking?* (hereafter *Orientieren*). There, Kant makes it quite clear that “the freedom to think is opposed first of all to *civil compulsion*. [...] Second, freedom to think is also [...] opposed to *compulsion over conscience* [...] [and] third, freedom in thinking signifies the subjection of reason to no laws except *those which it gives itself*” (*Orientieren*, AA 8, 144 [Kant 2001b, 16]). The last point is about the self-legislative aspect of thought, the fact that a thinking subject is bound only by the laws of thought itself, not just logical laws but also general rules for universalizability. We shall return to this latter aspect later. The second point concerns moral duty and conscience, which we cannot discuss here. This is also particularly relevant in the context of religious issues, which are discussed in *Aufklärung* but we will have to set aside here. The first element mentioned here,

<sup>1</sup> Horace writes in his letter to Lollius: “Dimidium facti, qui coepit, habet: sapere aude, / incipe” (*Epistolae* I,2,40).

freedom of thought as opposed to “civil compulsion” or civil coercion, is an important facet that Kant also emphasizes in *Aufklärung*: it signifies the fundamental and incontrovertible public character of thought, “namely [the] freedom to make *public use* of one’s reason in all matters,” as he writes in *Aufklärung* (AA 8, 36 [Kant 1999, 18]). But before we get to that, I want to go back for a moment to the aspect of “resolution” and “courage” that Kant associates with emerging from immaturity, and the question of authority. He writes further:

It is because of laziness and cowardice that so great a part of humankind, after nature has long since emancipated them from other people’s direction [...], nevertheless gladly remains minors for life, and that it becomes so easy for others to set themselves up as their guardians. *It is so comfortable to be a minor!* If I have a book that understands for me, a spiritual advisor who has a conscience for me, a doctor who decides upon a regimen for me, and so forth, I need not trouble myself at all. I need not think, if only I can pay; others will readily undertake the irksome business for me. That by far the greatest part of humankind [...] should hold the step toward majority to be not only troublesome but also highly dangerous will soon be seen to by those guardians who have kindly taken it upon themselves to supervise them; [...] pre-cepts and formulas, those mechanical instruments of a rational use, or rather misuse, of his natural endowments, are the ball and chain of an everlasting minority. (*Aufklärung*, AA 8, 35–6, emphasis added [Kant 1999, 17])

What Kant says here can be transferred without much fantasy to the context of our own time, in which an expertocracy is more and more seen as the indisputable measure of truth while anyone who makes independent use of his reason is quickly disparaged as an armchair scientist or—very topical during the recent Covid pandemic—a hobby virologist, or put down as an anti-vaxxer, conspiracy theorist or worse. This situation is, on the one hand, partly determined by the relative decline in trust in the institutions and the worrisome increase in what is sometimes referred to as “fake news” and “alternative facts” (a.k.a. “post-truth”), a trend that became increasingly apparent in the Covid pandemic during 2020–22. On the other hand, science is seen by its latter-day “fideist” proponents as neutrally “objective” or completely ideology free as if there was such a thing as *the* science or as if scientific consensus were not something that is reached only after long processes of internal evaluation and as if science is sometimes not also beset by controversy and malfeasance (e.g., especially when the pharmaceutical industry is involved). In an extremely complex society such as ours, hyper-specialization in the various fields of science accounts for why only a multidisciplinary approach can lay any claim to objectivity when dealing with a multifaceted, far-reaching *social and political* as well as medical and epidemiological problem such as a pandemic. Presumptuous, self-righteous insistence on “the facts speak for themselves,” on the so-called clear

language of “science”—in this case, epidemiology and virology—betrays a blinkered view rather than a sincere wish to solve or mitigate the real problem, in all its social and political ramifications. The sharp polarization between science “fideists” and its critics is fuelled by a distrust on the part of the science “fideists” of any kind of self-thinking, which is denounced as by definition anti-scientific, hence irrational and irresponsible, if not conspiratorial—whether it concerns the self-thinking from fellow scientists whose specialization is not in virology or epidemiology, critical journalists or indeed laymen. This polarization increases social pressure and the risk of groupthink, which can quickly degenerate into mass psychosis. Signs of this were visible during the pandemic. It is precisely the political authorities, and increasingly large parts of the mainstream media (with fortunate exceptions), that display a culpable inability to pause for reflection, think critically and facilitate a truly critical debate—all with the *prima facie* attractive-sounding excuse of “we do not have time for reflection, we must act now!”

After almost two years of pandemic without any real exit strategy, the media were mainly to be blamed for forsaking its fundamentally critical role towards failing governments and instead lambasting, as a true guardian (*Vormund*) in state service, those who criticized the government and its epidemiologist and virologist experts. The media’s active role, as a quasi-propaganda tool of an increasingly authoritarian government, in denigrating or outright slandering—in Germany critics were often disdainfully called, by government officials as well as journalists, *Querdenker* or *Schwurbler*—those who remain critical, however ostensibly unfounded, and do not comply with the dictates of power, is quite recent. This poses a potential danger to the fundamental freedom to be permitted, within the limits of the law, to have a different point of view and express this publicly, which is precisely what Kant strives to illuminate in *Aufklärung*. A pandemic, and certainly not one of the magnitude as the one that plagued us recently, is not a sufficient reason to suspend that freedom, both the freedom to think and be able to make choices for oneself when it comes to questions of life and death.<sup>1</sup> When we give up that freedom in favor of public health or a supposed fundamental right to physical health (*körperliche Gesundheit*), say, as it was frequently argued by experts and intellectuals alike,<sup>2</sup> the foundations of a liberal democratic society, in which we have to hold out and tolerate differences of opinion even in crisis situations, are undermined.

Kant demonstrates a psychosocial insight when he articulates one reason for immaturity, which could be called an argument from comfort: namely, it is all too convenient to be placed under the “guardianship” of an expert or a group of experts, lest one be compelled to think and investigate for oneself. It is easiest to defer to others, to experts or political leaders, for making decisions and in this way to delegate one’s epistemic duty to think to others. Why should I want to think about certain matters when we have politicians or scientists or church leaders to do this on our behalf?

<sup>1</sup> See Schulting (ms.) for more detail.

<sup>2</sup> See e.g., the discussions in Günther & Volkmann (2022).

One thus relies solely on the transfer of knowledge by the expert. But by unquestioningly accepting the expert's absolute authority, as one's guardian, one in fact also ascribes all responsibility to him. In this way one incapacitates (*entmündigt*) oneself. This implies that *per impossibile* one wishes to divest oneself of one's autonomy, the innate capacity of man to follow only that rule or law that one has prescribed or legislated oneself through reason, in accordance with point 3 from the quotation from *Orientieren*: "freedom in thinking signifies the subjection of reason to no laws except those which it gives itself" (*Orientieren*, AA 8, 144 [Kant 2001b, 16]).

Another social phenomenon that Kant touches on here is the lurking groupthink mentioned earlier: people consider maturity not only to be a burden because they find it uncomfortable for themselves, but also "very dangerous." One might be tempted to make up one's own mind! There is huge peer pressure not to engage in thinking for oneself. The risk of group coercion and conformity is a real one. This is of course especially true when the state takes on an excessively patronizing role or encourages conformism: paternalism is a form of a state-ordained *Entmündigung* of the freethinking and autonomously acting citizen—paternalism is a symptom of despotism, as Kant says (AA 8, 290–1). But it is not only the state—the monarch, in Kant's argument—that can act outside of its lawful remit here: it is precisely those who set themselves up as guardians, the experts, and the media foremost, who may play an important censoring role, which is often denied because it is widely but erroneously thought that by definition censorship can be imposed only by the state. There is a special, pernicious kind of censorship that threatens to become internalized. When Kant speaks of "precepts and formulas" (*Aufklärung*, AA 8, 36 [Kant 1999, 17]), he points to the temptation to conform, to fall back on standards which others—not necessarily state officials—set out for you, steadily gnawing away at your natural ability to think for yourself. It is not just that it is a dereliction of my duty to think for myself when I defer to others by compliantly submitting to their heteronomous authority, similar to how I would be summoned to simply obey a tyrannical monarch's censorious edicts. Rather, what happens is that I have *internalized* particular norms about what is deemed to be true and good, propagated and enforced by institutions, pressure groups and the media as well as thought leaders and *Querdenker*, so that what is in actual fact imposed on me by an external authority is presented to one *as* inner, as one's own authentic voice, thereby basically eliminating self-legislated thinking for the self has become a replica of external authority. The spontaneous agent of thought has been replaced by an automaton that is merely parroting what others think. Mechanical, automated thinking has as it were become second nature.<sup>1</sup> The most harmful censorship is not blunt censorship by the state but the hidden censorship of *self-censorship*, an internalized conformity as a collectively sanctioned way of thinking that becomes the standard.

---

<sup>1</sup> Pointedly, Kant refers to the fact that a human being is more than a "machine" at the end of *Aufklärung* (AA 8, 42 [Kant 1999, 22]).

Of course, Kant is not saying that we should not rely at all on the expertise or guidance of a scientist, church pastor or physician, or that the state should not have a regulating or steering function at all. It certainly does not mean that anyone can think and express what they like in any arbitrary fashion, without being bound by any rule or law; “libertinism” (*Freigeisterei*) is certainly not the motto of the Enlightenment (*Orientieren*, AA 8, 146 [Kant 2001b, 17]). Even the greatest nonsense is still subject to laws, Kant says, and certainly proper thinking, despite its internally anchored freedom, is bound by laws, namely its *own* laws. Neither thought nor freedom is unbridled (*Orientieren*, AA 8, 144–5).<sup>1</sup>

However, the issue here revolves around the question to what extent and in what sense the individual assigns authority to the expert or body of experts and takes into account the relationship between the transfer of knowledge by the expert and himself as a subject thinking for himself; and, very importantly, the extent to which he is given the space to do so, by the state and by society at large (including the media).

This brings us back to the first crucial element of maturity that I referred to earlier: namely, the necessarily public and argumentative or communicative nature of thought. This is an element that resurfaces several times in Kant's thinking. For example, in the *Critique of the Power of Judgment*, published six years after *Aufklärung*, where Kant, in the context of a discussion of the *sensus communis*, elaborates on the following three “maxims of the common human understanding,” the second of which concerns a “broad-minded” way of thinking:

1. To think for oneself; 2. To think in the position of everyone else; 3. Always to think in accord with oneself. The first is the maxim of the unprejudiced way of thinking, the second of the broad-minded way, the third that of the consistent way. The first is the maxim of a reason that is never passive. The tendency toward the latter, hence toward heteronomy of reason, is called prejudice; and the greatest prejudice of all is that of representing reason as if it were not subject to the rules of nature on which the understanding grounds it by means of its own essential law: i.e., superstition. Liberation from superstition is called enlightenment, since, although this designation is also applied to liberation from prejudices in general, it is superstition above all (*in sensu eminenti*) that deserves to be called a prejudice, since the blindness to which superstition leads, which indeed it even demands as an obligation, is what makes most evident the need to be led by others, hence the condition of a passive reason. As far as

<sup>1</sup> Consider also Kant's metaphor of “the light dove [that], in free flight cutting through the air the resistance of which it feels, could get the idea that it could do even better in airless space” (*Critique of Pure Reason*, A5/B8–9 [Kant 1997, 129]), suggesting that free and critical thought is not the kind of free-floating thinking that is completely unbound by limits and constraints.



the second maxim of the way of thinking is concerned, *we are accustomed to calling someone limited (narrow-minded, in contrast to broad-minded) whose talents do not suffice for any great employment (especially if it is intensive). But the issue here is not the faculty of cognition, but the way of thinking needed to make a purposive use of it, which, however small the scope and degree of a person's natural endowment may be, nevertheless reveals a man of a broad-minded way of thinking if he sets himself apart from the subjective private conditions of the judgment, within which so many others are as if bracketed, and reflects on his own judgment from a universal standpoint (which he can only determine by putting himself into the standpoint of others). (Critique of the Power of Judgment, AA 5, 294–5, emphasis added [Kant 2001a, 174–5])*

The concept of *sensus communis* is actually more applicable to aesthetic judgments, Kant writes, yet before the enumeration of the three maxims of common human understanding quoted above, he says that we can understand it as

the idea of a communal sense, i.e., a faculty for judging that in its reflection takes account (a priori) of everyone else's way of representing in thought, in order as it were to hold its judgment up to human reason as a whole and thereby avoid the illusion which, from subjective private conditions that could easily be held to be objective, would have a detrimental influence on the judgment. Now this happens by one *holding his judgment up not so much to the actual as to the merely possible judgments of others, and putting himself into the position of everyone else, merely by abstracting from the limitations that contingently attach to our own judging*; which is in turn accomplished by leaving out as far as is possible everything in one's representational state that is matter, i.e., sensation, and attending solely to the formal peculiarities of his representation or his representational state. [...] In itself, nothing is more natural than to abstract from charm and emotion if one is *seeking a judgment that is to serve as a universal rule*. (*Critique of the Power of Judgment*, AA 5, 293–3, emphasis added [Kant 2001a, 173–4])

Kant makes it very clear here that self-thinking is not pure subjectivist, unrestrained “libertinism” (*Freigeisterei*) (cf. *Orientieren*, AA 8, 146). Thinking for oneself comes with a certain responsibility that is inherent in the reflexive character of thinking itself. The purposive and justified use of one's reason, the “broad-minded” way of thinking of which Kant speaks, aims to rise above one's own subjective prejudices and put oneself in the position of “human reason as a

whole (*die gesamte Menschenvernunft*),” i.e., of “everyone else.” We must seek “a judgment that is to serve as a universal rule.” This could be called the universalizability requirement that is inherent in maturity, in enlightened thinking. The consideration must be whether I can imagine that my own thoughts yield a universal rule, which affords a criterion for determining a shared objectivity to one’s thoughts. But of course, that also applies reciprocally to the thoughts of the other, in whose at first sight “narrow-minded” world of thought I must be able to put myself.

Mature thinking is not based on a standpoint epistemology, which persists in one’s own bias—Kant’s term is “prejudice”—or indeed ignorance, nor on shifting responsibility to “science.” Thinking is based on a reflective consideration of a wide range of differing views. Self-thinking is therefore always also thinking *with* the other. It requires a certain “formal” distancing from one’s own ingrained contingent prejudices or one’s sense of moral-epistemic superiority. We cannot exclude the other simply because we—and by “we” is meant each individual thinking “I” for himself—believe that we have truth on our side, that we consider our view to be “the only true one.” Freedom of thought also implies the freedom to be wrong. And the possibility that we ourselves are wrong in fact cannot be ruled out either. Maturity therefore not only means freedom of thought and freedom from the authority of the other but also implies a certain epistemic humility on our own part.

In contemporary debates about freedom of speech and academic freedom—the latter a subject that is more complex than can be addressed here, since there are more factors at play and certain limitations apply in the case of academic freedom that do not apply to freedom of speech in general—it is often said that no-platforming a particular speaker does not imply that his or her specific *beliefs* are thereby censored. By making the distinction between having beliefs and having a platform to express them, one believes the charge of cancel culture or censorship can be evaded. As said, it is also believed, erroneously, that censorship can only be exercised by the state, something that would have been the case in Kant’s time—take for example the edict that affected Kant’s ability to write on religious topics in the early 1790s—but which is no longer so in contemporary Western society where the spheres of influence shared between state and social actors are much more evenly distributed. In any case, distinguishing between having certain beliefs and having a platform for expressing them comes down to sophistry, based on a misguided distinction between freedom of thought and freedom of publishing. Of course, no one has a *right* to be published on a particular platform or medium. However, in principle denying someone the opportunity to publicly express their beliefs —by means of social no-platforming or withdrawing invitations under social pressure, for example—comes down to *de facto* censorship.

Kant is very explicit about this in the text from *Orientieren* quoted in part earlier:

Of course, it is said that the freedom to *speak* or to *write* could be taken from us by a superior power, but the freedom to *think* cannot be. Yet how

much and how correctly would we *think* if we did not think as it were in community with others to whom we *communicate* our thoughts, and who communicate theirs with us! Thus, one can very well say that this external power which wrenches away people's freedom publicly to *communicate* their thoughts also takes from them the freedom to *think*—that single gem remaining to us in the midst of all the burdens of civil life, through which alone we can devise means of overcoming all the evils of our condition. (*Orientieren*, AA 8, 144 [Kant 2001b, 16])

Freedom of thought implies the ability to communicate one's thoughts "publicly." If one deprives someone of the opportunity to express himself in public, one deprives that person of the freedom of thought *tout court*. As we have seen before, and what Kant again confirms in the quotation above: Freedom of thought only exists "in community with others," to whom we must be able to communicate our thoughts and vice versa. If we exclude someone from that community, then freedom of thought itself is at stake.

Let us return to *Aufklärung*. Halfway through the essay, Kant introduces the familiar, and at first blush confusing, distinction between the "private" and "public" uses of one's reason or speech. We have not come across the notion of a "private" use of one's reason or speech until now. The private use of one's speech does not exactly concern an individual person's right to have his own opinion (freedom of speech). In fact, the private use of one's reason can be quite limited.<sup>1</sup>

By the latter, Kant means the use a person makes of his reason "in a certain *civil* post or office with which he is entrusted" (*Aufklärung*, AA 8, 37 [Kant 1999, 18]). This is, of course, primarily aimed at officials of the state and civil servants, whose "positive" function is to achieve certain public goals. As officials, they must be "passive" with respect to their own freedom as citizens of the "commonwealth," to which they belong just like any other citizen. The state must be able to perform its function without interference or opposition from its own officers. As Kant puts it, "it would be ruinous if an officer, receiving an order from his superiors, wanted while on duty to engage openly in subtle reasoning about its appropriateness or utility; he must just obey" (*Aufklärung*, AA 8, 37 [Kant 1999, 18–19]). As functionaries they are "passive," in a sense "prejudiced" as Kant says in the above-quoted passage from the *Critique of the Power of Judgment*, for they are subject to the heteronomy that comes with their passive role as civil servants. Only as citizens of the commonwealth can those same civil servants also make "public" use of their reason, by actively criticizing the same policy that they officially, passively must support. This passive/active distinction also applies to ordinary citizens themselves: as residents of a state

---

<sup>1</sup> Though Kant does not specifically address here academic freedom, academic freedom is captured by both the private and public uses of one's reason. Insofar as an academic is an officer in service of a state organization, e.g., at a public university, his freedom to "privately" use his reason can be quite limited, whereas at the same time he has at least the same broad freedom as scholar as everyone else.

citizens *passively* have to pay taxes on pain of a fine or even imprisonment, and yet as mature citizens of the commonwealth they can at the same time *actively* criticize tax policy—as long as they pay.

By making public use of one's reason, Kant means "that use which someone makes of it as a scholar before the entire public of the world of readers" (*Aufklärung*, AA 8, 37 [Kant 1999, 18]). It is this public use that is essential to freedom of thought, as we have discussed at length above. And here too in *Aufklärung*, Kant makes it clear that thinking takes place before a readership (*Publikum der Leserwelt*), a forum where one can make one's thoughts known. Everyone, whether public servant, priest or citizen, is a member of a "commonwealth," and therefore "a scholar who by his writings addresses a public in the proper sense of the word" (*Aufklärung*, AA 8, 37 [Kant 1999, 18]). It is striking that for Kant *everyone* can be considered a "scholar." This label is thus not limited to academics! To be a scholar in this sense does not mean to possess a vast wealth of knowledge; one can have so much objective knowledge (*Kenntnisse*) and yet be "least enlightened in the use" one makes of one's faculty of cognition. Self-thinking, maturity, means "seeking the supreme touchstone of truth in oneself (i.e., in one's own reason)." "The maxim of always thinking for oneself," which is a "negative principle in the use of one's faculty of cognition," *that* is enlightenment (*Orientieren*, AA 8, 146, footnote [Kant 2001b, 18]). And every layman is capable of thinking for himself.

Let me close this section by quoting the passage in *Orientieren* which best encapsulates Kant's view on enlightenment in full:

Thinking for oneself means seeking the supreme touchstone of truth in oneself (i.e., in one's own reason); and the maxim of always thinking for oneself is enlightenment. Now there is less to this than people imagine when they place enlightenment in the acquisition of information; for it is rather a negative principle in the use of one's faculty of cognition, and *often he who is richest in information is the least enlightened in the use he makes of it*. To make use of one's own reason means no more than to ask oneself, whenever one is supposed to assume something, whether one could find it feasible to make the ground or the rule on which one assumes it into a universal principle for the use of reason. This test is one that everyone can apply to himself; and with this examination he will see superstition and enthusiasm disappear, *even if he falls far short of having the information [Kenntnisse] to refute them on objective grounds*. For he is using merely the maxim of reason's self-preservation. (*Orientieren*, AA 8, 146, footnote, emphasis added [Kant 2001b, 18])

## II. Maturity, Universalism, and Emancipation

Exhortations to aspire to a humanistic universalism and appeals to think beyond a tribalistic “we,” beyond any particular identity or group interest, remain gratuitous without an appeal to individual responsibility and judgment and when no room is made for the possibility of actually realizing such lofty ideals. Universal ideals are after all pursued by subjects that are aware of themselves as self-thinking agents pursuing such ideals. However, not everyone ipso facto meets the epistemic duty to think “for oneself” and judge in accordance with reason at all times, and thus help realize the universal ideal of the self-preservation of reason (*die Selbsterhaltung der Vernunft*) (*Orientieren*, AA 8, 146, footnote [Kant 2001b, 18]).<sup>1</sup>

To think for oneself means to submit to no heteronomous authority, recognizing solely the jurisdiction of reason. “Self-thinking” people recognize nothing as reasonable but what they themselves grasp as reasonable. “Self-thinking” means that with every action or view one must ask oneself if the maxim or rule on the basis of which one carries out any action or adopts a particular view could be adopted as a universal law, applicable to everyone (see e.g., *Orientieren*, AA 8, 146, quoted at the end of the previous section). Put simply, are the reasons why I say or do something also universalizable? That is to say, is what I say or do in principle also something everyone else should say or do? To ask oneself such questions constitutes, for Kant, the essence of “enlightenment.”

But self-thinking is a mentality or stance, not an innate factual quality or a dispositional trait that automatically manifests itself, nor a trait that could be causally explained by one’s genetic makeup or cultural background, or indeed group identity. Every adult human being, regardless of background or identity, possesses this quality inherently *just because* he or she is a human rational being, but it must be cultivated. Every adult person has the *capacity* for enlightenment, regardless of circumstances, “however small the scope and degree of a person’s natural endowment may be” (*Critique of the Power of Judgment*, AA 5, 295 [Kant 2001a, 175]).

For Kant, having the capacity of understanding and being able to use that capacity independently of others is a universal maxim that is valid for every human being without exception, that is to say for every adult human being whose cognitive abilities are properly functioning. But note: “is valid.” We must talk here about validity, not about facts or factuality per se, or at least these two realms, facticity and validity, must not be conflated. After all, it is not the case that everyone also thinks in accordance with reason or makes proper use of his or her capacity for thought in actual fact, always, and in all respects. Some people or groups of people do not seem to be making much use of their capacity for thought at all.

Having the capacity of the mind, the ability to think according to rules that one imposes on oneself strictly in accordance with reason, does not by itself imply *strict* universal validity, as is

---

<sup>1</sup> Cf. Refl. 1509, AA 15, 823: “Grundsatz der Vernunft: ihre Selbsterhaltung.”

the case with the categorical imperative which can be deduced a priori from reason, or indeed with the constitutive rules of logic to which using that capacity is bound. Having the capacity to think is itself not like a logical law that always applies univocally, such that everything that I think is always true or even logically valid. The capacity of self-thinking is in a sense an assumption rooted in rational faith (*Vernunftglaube*) (*Orientieren*, AA 8, 140–1), which, although inextricably bound up with the possibility of philosophy, cannot be derived from it as if it concerned a principle of logic. Rational faith, a central Kantian concept that appears in various of his works and which we cannot examine further here, and the assumption of an innate predisposition to think for oneself are related and presuppose each other. In principle we thus have a capacity for reasonable thinking based on rational faith. However, that does not automatically mean that everything that is in actual fact being thought is thereby also reasonable, in accordance with reason (even if *what* we think may still be valid purely *logically*). As Heiner Klemme recently put it succinctly: “The maxims of the self-preservation of reason do not protect the individual against foolish acts” (Klemme 2023, 66).

It is thus important that we exercise judgment (*Urteilstkraft*) in making use of our thinking capacity (*Denkvermögen*). It also means that we must be open to the special circumstances under which we apply our maxims for judgment and action. Simply put, everything we do and think must always be put in context, and is defeasible. We must always ask ourselves whether certain rules we employ for our actions, and certain views we hold, are actually applicable in, or effectively relate to, reality. There is a fundamental discrepancy between a rule or maxim taken *in abstracto* and its actual application or realization. We cannot simply ignore the particular context of that about which one makes a judgment or in respect of which one formulates a maxim for action. Therefore, universalist platitudes such as “All people are equal” or “The intrinsic dignity of all people is inviolable,” while true, are meaningless pronouncements when expressed merely in the abstract without taking into account contingent external factors. Such in principle true statements cannot be utilized as a panacea for social or political problems if their concrete context is not further addressed and politically feasible solutions are not carefully deliberated. There are often conflicting interests, and liberal sounding statements that merely make a point of stressing equality and unity tend to be clichéd and moralistic and as such do not contribute to a practical resolution of such conflicts. Sometimes an idealist stance is ill-advised and one must be realistic and simply acknowledge that a conflict cannot be resolved or cannot be resolved immediately, without thereby detracting from the in principle equal status of the parties in the conflict. Not everyone makes proper use in actual fact of his or her capacity for thought, or bears responsibility in actual fact for his or her actions. Such observations are not value judgments about people’s innate *capacities*. They are judgments about the *proper use* of one’s capacity for thought and judgment *in particular circumstances*.

Abstract universalism and the rejection of universalism are two sides of the same coin. Just like abstract universalism, which foregrounds the universal truths of equality and intrinsic dignity for all while ignoring particular circumstances, the generalizing, formalistic criticism of, or rejection of, universalism as such expresses the same abstract interpretation of it. On account of such critical views, Kant's universalism is considered not truly universal because supposedly certain groups of people would be excluded from that same universalism in actual cases where the tenets of universalism such as "Equality for all" or "The intrinsic dignity of all people is inviolable" do not seem to apply, or at least not completely. Ergo, Kant's universalism would in reality be particularistic and exclusive, intended exclusively for the privileged Western white man, who alone enjoys universal rights to the full. People of non-Western backgrounds or women in general would supposedly not be the subjects of Kant's universalism, or in fact explicitly be excluded from it. This criticism can be heard more and more in contemporary debates about Kant, especially in the context of discussions of his alleged racism and position on colonialism.

But here too validity (*Gültigkeit*) and application (*Geltung*), or validity and realization, are easily conflated. Such a reading that appeals to specific irreducible cultural differences rests—just like an overly abstract universalism that holds forth with hollow declarations such as "All are equal"—on an abstract distinction between form and content, which holds that either the general form is stressed at the expense of the particular content, or the particular content is seen as stressed too much at the expense of universal form.

We can illustrate such an overly formalistic, cursory reading of Kant's texts with the standard account of his views on women. There are texts in Kant's oeuvre that, at first glance, appear to show unsettling examples of misogyny or are, at the very least, not particularly favorable to women (see e.g., AA 6, 279; AA 7, 209, 307; AA 27, 36, 48). Kant is not particularly known for his feminism! It thus seems that Kant excludes women, as apparently, he does other groups of people, from his universalist perspective, regardless whether this has to do with moral action or more generally with the capacity for judgment associated with the use of one's reason. But are women, by definition, as adult human beings, on Kant's account then not mature citizens who must be able to use their capacity for thinking without the guidance of others?

Klemme quotes a telling passage from one of the *Vorlesungen zur Anthropologie* (*Anthropologische Menschenkunde*):

It is assumed that certain people are not authorized to use their understanding on their own but are capable of [*können*] judging only with the help of a foreign understanding. Such people are called minors [*Unmündige*]. Some are minors according to age; they are incapable of [*können...nicht*] complying with their own understanding and reason but must be guided by another. Others are minors according to sex; certain

insights and businesses are wholly outside the sphere of the female room. They are not allowed [*dürfen...nicht*] to make use of their own reason but must submit to the observations of a foreign reason; but as regards public matters, they must rely on foreign reason. With children immaturity is natural; the guardian of a woman's room is called a custodian. (*Anthr. Menschenkunde*, AA 25, 1046–7, translation mine)

On a superficial reading, this passage from one of his many *Vorlesungen* on anthropology seems to indicate that Kant regards women as immature. In the published *Anthropology* from 1798, he indeed appears to deny (married) women maturity, as wives are to be submissive to their husbands as their guardians. He curiously contends that a woman who “by the nature of her sex has enough of a mouth [*Mundwerk genug hat*] to represent both herself and her husband, even in court [...], could literally be declared to be *over-mature* [*übermündig*]” (AA 7, 209 [Kant 2007, 315]). But Kant makes no validity judgments in the passage in the anthropology lecture quoted above. Rather, he points to the concrete, actual condition of immaturity in which women largely find themselves—this was certainly the case in Kant's days. At least two things bear this out.

First, Kant uses the modal verb “*dürfen*” here (and not, for example, the imperative “*sollen*”). There is a restriction imposed from outside: either something is allowed or prohibited, or one is obliged to do something. This is significant in view of the fact that Kant believes that immaturity is something that is rather “self-incurred” (*selbstverschuldet*) and deserving of blame. Of course, the imposed restriction also applies to children: both children and women “have to” (*müssen*) be guided by others. But the important distinction between minors and women is that, according to Kant, immaturity in children is “natural,” since children by definition have not yet reached a certain age and therefore maturity, hence they are called minors, while women need a chaperone or custodian, suggesting that immaturity in women is not natural. Note also that Kant contrasts “*können*” in the first sentence with “*dürfen*” in the second: whereas women are presumably not *permitted* to make use of their own understanding, children *cannot* make use of their understanding without the guidance of others. Likewise, to speak of “self-incurred immaturity” in the case of children would be inapposite because children cannot be blamed for their being underage. The culpability condition does not apply in their case. Therefore, a categorical distinction should be heeded between immaturity “according to age” (*den Jahren nach*) and that “according to sex” (*dem Geschlecht nach*). The latter is not natural but socially constructed.

This latter way of conceiving of maturity is what Kant elsewhere associates with “legal” or “civil” immaturity, which can be so called “if it rests on legal arrangements with regard to civil affairs” (*Anthropology*, AA 7, 208–9 [Kant 2007, 315]). It may seem at first sight that Kant endorses such “legal” immaturity for women, who “cannot personally defend their rights and pursue civil affairs for themselves.” It is not clear, from the declarative manner of reasoning in the



published *Anthropology* text—which we recall is not part of Kant’s corpus of transcendental or critical philosophy, and was compiled on the basis of his many lectures over the years, even if Kant authorized it as being the latest version of his views on anthropology—whether a) Kant thinks women *ought* not be able to personally exercise their rights and thus enjoy civil independence, or b) just that, in given social and economic circumstances, women’s actual civil status can be guaranteed only by the husband in a marriage, thus more reflecting women’s current status in society than prescribing a normative rule. Kant argues that, in principle, there is a “natural” equality in a couple but that the husband’s dominance over his wife is not contradictory to it to the extent that it functions “to promote the common interest of the household” (AA 6, 279 [Kant 1999, 428]). But the former (a) could hardly be justified a priori just on the basis of the precept of maturity, and it also conflicts with the idea that immaturity is self-incurred, for if their “legal” immaturity is warranted or even a priori justified and they *ought* to be immature, women could hardly be blamed for it (see further below on the culpability condition). So it seems that there is room on a Kantian view, if not strictly speaking in Kant’s own view, for the idea that women can and even must achieve some sort of civil independence, inside marriage and indeed outside it.

Kant should at any rate have drawn that conclusion also based on the fact that he encountered many women in prominent positions, either in person or through correspondence, women who were quite obviously articulate and made use of their own capacity for thinking and judging without the guidance of others. Between 1758 and 1762 two Russian Tsarinas ruled in Königsberg. He even writes a letter to Tsarina Elisabeth in which he addresses her with all manner of honorifics such as “allerdurchlauchtigste Großmächtigste Kaiserin.” There was undoubtedly something opportunistic about the way Kant formulated his address to the Tsarina, because he wanted to obtain the position of professor at the University of Königsberg (which he only succeeded in doing years later). But a “Selbstherrscherin” is not exactly someone who needs a custodian (quotations from Klemme 2023, 67–8).

Secondly, someone who is immature “according to age,” such as an underage child, cannot be held responsible for being underage. After all, it is its natural state. However, the aspect of culpability is precisely what Kant emphasizes in *Aufklärung*. Immaturity, including that of women, is “self-incurred” or culpable if it is not due to lack of understanding but because one fails to try and get oneself out of it, something for which one is to be blamed and for which one is responsible oneself. This sounds paradoxical at first: if it is true that women are not *permitted* to be “legally” mature in virtue of society’s limitations of the exercise of their freedom, also and particularly in business and legal affairs—their immaturity is imposed on them by society’s structures and social mechanisms—how can it be that they themselves are to be blamed for this? Are they really responsible for their own imposed immaturity?

Klemme writes regarding this:

Kant does not mean that women could not, and do not want to, be partially mature. After all, they run the household and are legal subjects who can enter into a marriage contract. However, they do not take an active interest in their civic independence. (2023, 68, translation mine)

What Klemme points out here is the importance of a citizen's independence, which is to be claimed by the subject herself alone (but cf. AA 6, 314 and the earlier cited *Anthropology* at AA 7, 208–9). This is about autonomy and self-agency. Although immaturity may be the effect of external hindering circumstances, the responsibility for remaining in that immature state ultimately lies with the subject herself. Outside hindrances, social or otherwise, cannot be exculpatory reasons to remain in a state of immaturity. Thus, she herself must take an “active interest” in acquiring “civic independence” in social life. That is why Kant speaks of immaturity as something “self-incurred.” The contingent circumstances in which one finds oneself at the hands of others do not provide a license to shift the blame for one's situation to others. One owes it to oneself to break free from the reins of tutelage, if at all possible. And this I contend is the general dialectic of Kant's reasoning in *Aufklärung*, also in regard to women, as I shall argue below.

Here, Kant could in fact be taken to be advocating for the liberation of women, which however women must foster of their own accord. This is only logical because liberation cannot be a heteronomous cause, something done by others on your behalf. For if it were, this would imply a degree of tutelage or paternalism that one wants precisely to get away from. To no longer bow under the yoke of tutelage, to emerge from immaturity, is an action that the subject must undertake herself, an act for which she herself must take responsibility. It is always a matter of resolutely “to make use of one's own reason” (*Orientieren*, AA 8, 146, footnote, emphasis added [Kant 2001b, 18]).

This is of course far easier said than done, and Kant is well aware of this, even though one may find that Kant is perhaps somewhat too optimistic (or too negative for that matter, on a certain reading of his views of a woman's place in marriage). He speaks of a “progress of enlightenment” (*Aufklärung*, AA 8, 37 [Kant, 1999, 18]), but the “great unthinking masses” develop in that direction only very slowly, for to “enlighten an age” is “arduous” (*Orientieren*, AA 8, 146 [Kant, 2001b, 18]). “Precepts and formulas, those mechanical instruments of a [...] misuse of [one's] natural endowments,” keep the general public in a permanent state of immaturity. Only few manage to escape from it. Immaturity seems to have “almost become nature” (all quotations *Aufklärung*, AA 8, 36 [Kant 1999, 17–18]).

Nevertheless, “we do have distinct intimations that the field is now being opened for them to work freely in this direction and that the hindrances to *universal* enlightenment or to *humankind's* emergence from its self-incurred minority are gradually become fewer” (*Aufklärung*, 8, 40, emphasis added [Kant 1999, 21]). Notice that Kant talks about “*universal* enlightenment” and

“humankind’s emergence” as a goal toward which we gradually progress. This suggests again that the *universal* validity of the precept of maturity must be read in terms of a regulative ideal for everyone, women included, which in practice is not yet actualized to its full potential, and in all likelihood remains something realized in full only in potentiality. Not everyone everywhere may already enjoy the freedom to make use of one’s own reason without sanction. In reality, while the emergence from immaturity is the aim which is realizable just because the maxim or principle of maturity, of thinking for oneself, is universally valid, the state of minority is still the factual situation for the greater part of humanity, and sadly in particular for most women in the non-Western regions of the world. The universal validity of the maxim of maturity does not take away from the fact that it is really possible and indeed actual that large groups of people remain constrained by the shackles of immaturity at least for some time, and that contingent sociopolitical forces are largely the cause of this. The universality of the principle or maxim cannot undo the fact of the matter of external restriction and oppression, just as much as contingent facts cannot undermine the universal validity of the maxim. “People gradually work their way out of barbarism of their own accord *if only one does not intentionally contrive to keep them in it*” (*Aufklärung*, AA 8, 41, emphasis added [Kant 1999, 21]).

Enlightenment is a continuous development, not a closed, finite collection of achievements: “As matters now stand, a good deal more is required for people on the whole to be in the position, or even able to be put into the position, of using their own understanding confidently and well in religious matters, without another’s guidance” (*Aufklärung*, AA 8, 40 [Kant 1999, 21]).<sup>1</sup>

When Kant therefore writes that “by far the greatest part of humankind (*including the entire fair sex*) should hold the step toward majority to be not only troublesome but also highly dangerous” (*Aufklärung*, AA 8, 35, emphasis added [Kant 1999, 17]), Kant is not, despite appearances and despite apparent contradictory views stated elsewhere, making a misogynistic statement about women (as well as other groups) and their perceived *natural* inability to be mature. On the contrary, in the remainder of the sentence he makes it perfectly clear that “guardians” see to it that women are kept under control, and

have kindly taken it upon themselves to supervise them; after they have made their domesticated animals dumb and carefully prevented these placid creatures from daring to take a single step without the walking cart [*Gängelwagen*]<sup>2</sup> in which they have confined them, they then show them

<sup>1</sup> Kant focuses on religion because religious immaturity is “the most harmful [as well as] the most degrading [*entehrendste*] of all” (*Aufklärung*, AA 8, 41, trans. emended [Kant 1999, 21]).

<sup>2</sup> Kant uses the same word for “walking cart,” *Gängelwagen*, in *Conjectural Beginning of Human History* (1786), where, translated as “go-cart,” it is associated with the “guardianship of nature” and “instinct” in contrast to “reason” and “the condition of freedom” (AA 8, 115 [Kant 2007, 168]). Earlier in the *Critique of Pure Reason*, Kant uses the term *Gängelwagen*, translated as “leading-strings,” in a theoretical context, that is, the introductory section of the

the danger that threatens them if they try to walk alone. (*Aufklärung*, AA 8, 35 [Kant 1999, 17])

This domineering guardianship ensures that large groups of people do not dare to think for themselves, “to walk alone.” The real danger associated with assertiveness may not be all that great. However, for many people the real-life consequences of self-thinking by way of criticism of the accepted norm—the risk of social isolation, disapproval by peers, denunciation, “cancelling,” etc.—do “make them timid and usually frightens them away from any further attempt” (*Aufklärung*, AA 8, 35–6 [Kant 1999, 17]). Self-thinking therefore requires “resolution and courage,” hence the exhortation quoted from Horace at the start of *Aufklärung*: “*Sapere aude!*” “Dare to think!” (*Aufklärung*, AA 8, 35 [Kant 1999, 17]).

Women are subjected to oppressive institutionalized structures and social mechanisms in large parts of the world, which prevents them from benefiting from their natural right to take responsibility, and think, for themselves. These factual circumstances do not have a negative influence on the universal *validity* of the rational maxim of maturity as such but they have an effect on the extent to which what is universally valid is also realized, actualized.

Kant makes no pronouncements about how the emancipation of women or any other group of civil minors could or should take shape. That is simply not his philosophical project. To criticize him for it is to misapprehend the nature of his thought, and also to overstate its scope—withstanding prevaricating statements in the various texts in which he addresses the status of women. In *Aufklärung*, he merely 1) argues that the practical precept of maturity is valid for every adult human being and 2) observes that in many cases circumstances have a negative or inhibiting effect on the actual possibilities for realization (the application range of the regulative ideal of maturity). This concerns a certain anthropology that describes the subjective, both hindering and facilitating conditions for its application, comparable to how Kant sees this in the context of the a priori laws of morality and its relation to a moral anthropology (cf. AA 6, 217). Such an anthropology—and it certainly need not be the *Anthropology* that Kant actually published but could be a more critical one—is important but does not in any way detract from the universality of the maxim of maturity as such. The latter is not grounded on such an anthropology. After all, if it were so grounded, that would mean that the principle would be dependent on empirical, contingent factors, which would contradict Kant's view that to persist in immaturity is something for which one is culpable. Such culpability can only be apportioned to human rational beings who in principle have the freedom of thought and who by nature, that is to say by virtue of their possessing the capacity of reason, must be treated as mature. They can be blamed for their rational decisions, not

---

Analytic of Principles, to signify the assistance that examples provide to “sharpen the power of judgment” (A134/B173–4 [Kant 1998, 269]).

for their temperament or cultural background, let alone their social conditions, or the simple fact that one is a man or a woman.

Maturity, in which enlightenment basically consists, is actually only in its infancy as far as the wider populace is concerned, says Kant. We do not yet live in an “*enlightened* age,” but we do live in an “age of enlightenment,” says Kant, suggesting that it concerns not a completed project but a continuous process of enlightening (*Aufklärung*, AA 8, 40 [Kant 1999, 21]). This was certainly the case in Kant’s own time when large parts of humanity still found themselves in a state of minority, but it remains true in our time despite the achievements of the past two centuries of European civilization since Kant’s essay. This indicates that, for Kant, enlightenment is not a completed entity with a definable historical limit, nor a body of doctrines or a hereditary ideology. Rather, enlightenment is a universal ideal to be realized, a process, not a closed project that supposedly failed because not everyone enjoys in actual fact the same benefits that it promised to all. It is a continuous, live task for every person alike to extricate oneself, to self-emancipate, from the chains of tutelage.<sup>1</sup>

## References

AA = Kant’s *Gesammelte Schriften* (De Gruyter, 1900–), followed by volume and page number(s).

A/B = A and B edition of the *Critique of Pure Reason* (Kant 1998).

*Anthropology* = *Anthropology from a Pragmatic Point of View*, in Kant (2007).

*Aufklärung* = *An Answer to the Question: What Is Enlightenment?* in Kant (1994, 1999).

*Orientieren* = *What Does It Mean to Orient Oneself in Thinking?* in Kant (2001b).

Refl. = *Reflexionen*

Deligiorgi, K. (2005). *Kant and the Culture of Enlightenment*, SUNY Press.

Günther, K. & U. Volkmann (eds) (2022). *Freiheit oder Leben? Das Abwegungsproblem der Zukunft*, Suhrkamp.

Kant, I. (1994). Beantwortung der Frage: Was ist Aufklärung? in I. Kant, *Was ist Aufklärung? Aufsätze zur Geschichte und Philosophie*, pp. 55–61, Edited by J. Zehbe, Vandenhoeck & Ruprecht.

Kant, I. (1998). *Critique of Pure Reason*, Edited by P. Guyer & A. Wood, Cambridge University Press.

Kant, I. (1999). *Practical Philosophy*, Edited by M. Gregor, Cambridge University Press.

Kant, I. (2001a). *Critique of the Power of Judgment*, Edited by P. Guyer, Cambridge University Press.

Kant, I. (2001b). *Religion and Rational Theology*, Edited by A. Wood & G. di Giovanni, Cambridge University Press.

Kant, I. (2007). *Anthropology, History, and Education*, Edited by G. Zöllner & R. Louden, Cambridge University Press.

Kant, I. (2012). *Lectures on Anthropology*, Edited by A. Wood & R. Louden, Cambridge University Press.

---

<sup>1</sup> Thanks are due as usual to Christian Onof for his comments on the penultimate draft.

- Klemme, H. (2023). *Die Selbsterhaltung der Vernunft. Kant und die Modernität seines Denkens*, Klostermann.
- Mendelssohn, M. (1994). Über die Frage: was heißt aufklären? in I. Kant, *Was ist Aufklärung? Aufsätze zur Geschichte und Philosophie*, pp. 129–32, Edited by J. Zehbe, Vandenhoeck & Ruprecht.
- Schulting, D. (ms.). *Vaccination, Autonomy, and “Moral Recklessness”*.



Journal of Philosophical Investigations

## Journal of Philosophical Investigations

Print ISSN: 2251-7960 Online ISSN: 2423-4419

Homepage: <https://philosophy.tabrizu.ac.ir>



University of Tabriz

# Kantian Fallibilist Ethics for AI Alignment\*

Vadim Chaly 

Lomonosov Moscow State University, Immanuel Kant Baltic Federal University, Russia, Email: [vadim.chaly@gmail.com](mailto:vadim.chaly@gmail.com)

### Article Info

#### Article type:

Research Article

#### Article history:

Received 20 July 2024

Received in revised form  
25 July 2024

Accepted 25 July 2024

Published online 07  
August 2024

#### Keywords:

AI alignment, moral  
deliberation, moral  
fallibilism specification  
gaming, kingdom of ends,  
categorical imperative,  
misgeneralization.

### ABSTRACT

The problem of AI alignment has parallels in Kantian ethics and can benefit from its concepts and arguments. The Kantian framework allows us to better answer the question of what exactly AI is being aligned to, what are the problems of alignment of rational agents in general, and what are the prospects for achieving a state of alignment. Having described the state of discussions about alignment in AI, I will reformulate them in Kantian terms. Thus, the process of alignment is captured by the concept of enlightenment, and for the final state of alignment in Kant's lexicon there is the concept of the "kingdom of ends." I will argue that the discourse of alignment and the Kantian ethical program 1) are devoted to the same general end of harmonizing the thinking and acting of rational agents, 2) encounter similar difficulties, well known in the Kantian discussions with its comparatively longer history, and 3) for a number of reasons lying on the side of humanity, do not have and, despite the hopes and attitudes of some participants in the AI discussions, will not have a theoretically rigorous, harmonious and practically implementable, conflict-free solution – alignment will remain a regulative idea in the Kantian sense, but will not become a reality.

**Cite this article:** Chaly, V. (2024). Kantian Fallibilist Ethics for AI Alignment. *Journal of Philosophical Investigations*, 18(47), 303-318. <https://doi.org/10.22034/jpiut.2024.62766.3837>



© The Author(s).

<https://doi.org/10.22034/jpiut.2024.62766.3837>

Publisher: University of Tabriz.

\* Funding this research was supported by the Ministry of Science and Higher Education of the Russian Federation grant no. 075-15-2019-1929, project "Kantian Rationality and Its Impact in Contemporary Science, Technology, and Social Institutions" provided at the Immanuel Kant Baltic Federal University (IKBFU), Kaliningrad.

## Introduction

The growing capabilities and spread of AI use in various areas of life make researchers and the general public concerned with the issue of ethical regulations for the use of AI. One way to formulate this problem is the discourse on AI alignment. In my article, I would like to reframe the ways of formulating and solving some of the main problems of this discourse in Kantian language and draw attention to the human side of the equation. This approach of translation into Kantian language has already proven its productivity in relation to a number of current problems in AI (Kim & Schönecker, 2022) and could be useful in case of alignment as well. The concept of alignment denotes a two-way process of bringing one side to fit the other<sup>1</sup>. However, the discussion of alignment in the AI literature mainly deals with the question “alignment of what?” and pays less attention to the question “alignment to what?”, which lies outside this disciplinary area. Meanwhile, putting forward humanity as the criterion or norm for alignment and fixing the “direction of fit” from AI to human means that we first need to understand the features of the human side, which forms or normalizes – or confuses – the AI side. That is, to try once again to answer the question that Kant places at the center of philosophy: what is a human being? More specifically, it is necessary to clarify how we think about alignment in relation to human rational agents, what obstacles arise along the way to alignment and to properly conceptualizing it, how we can – or cannot – imagine and conceptualize the final state of complete alignment, and so on. These questions take us to the traditional territory of ethics.

I will argue that the discourse of alignment and the Kantian ethical program 1) are devoted to the same general end of harmonizing the thinking and acting of rational agents, 2) encounter similar difficulties, well known in the Kantian discussions with its comparatively longer history, and 3) for a number of reasons lying on the side of humanity, do not have and, despite the hopes and attitudes of many participants in the AI discussions, will not have a theoretically rigorous, harmonious and practically implementable, conflict-free solution – alignment will remain a regulative idea in the Kantian sense, but will not become a reality. I will begin with a brief outline of the current state of the debate on AI alignment and the current level of understanding of the problems emerging in this area, then offer a translation of these problems into the language of Kantian philosophy, and then demonstrate the problematic nature of attempts to find a crisp solution to these problems.

---

<sup>1</sup> If we allow for a moment and in a footnote some bold theological analogies, then alignment is the process of creation in the image and the likeness, and also, if we allow for AI’s ability to reflective counter effort, the process of humanization of the creature by analogy with divinization. In short, a paradigmatic visual metaphor for alignment is Michelangelo’s “The Creation of Adam.”



## 1. The problems in AI alignment

The AI literature defines AI alignment by specifying its aim: “AI alignment aims to make AI systems behave in line with human intentions and values” (Leike & et al. 2018; Ji & et al. 2024). Alignment researchers themselves acknowledge that this definition and discussions within AI are surrounded by numerous general philosophical difficulties. For example, the problem of the relationship between human intentions and values is philosophical: how they are “aligned” to each other, whether there are conflicts between intentions, or interests, and values, between values of different types, between the diverse values and interests of different people with whom it is proposed to align AI, and how these conflicts can be resolved. The literature contains a “value-centric” approach that attempts to incorporate ethical issues (Future of Life Institute 2017), and the “intention-centric” approach, which leaves ethical issues out of the equation (Leike et al. 2018). The question of the meaning of the concept of humanity is fundamental. For some “the term human can represent various entities ranging from an individual to humanity” (Ji et al. 2024, 12), for others, the concept of humanity is defined through the property of rationality (Gabriel 2020, 420). This set of questions can be called terminological or conceptual. It touches upon the “perennial” questions of philosophy, and it is unlikely that the conceptual aspect or the empirical material brought by AI, as well as the increased criticality of these questions due to AI, will stimulate their unambiguous and indisputable solution. Nevertheless, as I hope to show, the Kantian framework offers a well-thought-out and robust version of an answer to these questions, which were central to Kant.

Another set of questions concerns how practical reasoning occurs in AI systems, and the kinds of failures in it that cause misalignment. The problem of misalignment is discussed primarily in relation to AI systems with Reinforcement Learning (RL), a machine learning technique that reproduces human trial-and-error process to train AI systems to make choices leading to optimal (i.e. aligned) outcomes. A set of possible outcomes is specified and hierarchically ordered by assigning scores or rewards that measure their human preferability. An AI system seeking to maximize “reward” is thus “motivated” to seek the most preferable outcome. Misalignment occurs when the system fails to produce an outcome optimal from a human perspective. Sometimes this failure is dramatic and potentially harmful.

Ji et al. divide causes of misalignment into *specification gaming* and *goal misgeneralization*, with *reward hacking* emphasized as a significant kind of the former (2024, 4).

Specification gaming refers to the phenomenon in AI systems where the system finds ways to achieve its specified objective in unintended or undesirable ways by exploiting loopholes in how the objective was defined (Papyshev and Migliorini 2024, 1).

The problem of specification gaming is caused by two factors: the frequent human failure to foresee and specify the full set of relevant objective parameters and range of ways to achieve objectives, and AI's growing capacity to meet the specifications in unforeseen ways, producing undesirable states of affairs as byproducts.

Misgeneralization is a type of out-of-distribution (OOD) robustness failure in RL (Langosco & et al. 2022, 1). OOD characterizes an environment with features not encountered by the AI system in its learning environment. When confronted with such an environment, AI systems may fail to generalize in two ways. One is capability misgeneralization, when an agent fails to capably pursue a goal in the new environment. The other is goal misgeneralization, which "occurs when an RL agent retains its capabilities out-of-distribution yet pursues the wrong goal" (Langosco & et al. 2022, 1). In other words, when confronted with a new environment, an agent demonstrates that it has learned a wrong goal while in training. A basic example of such behavior is when "an agent trained to pursue a fixed coin might not recognize the coin when it is positioned elsewhere, and instead competently navigate to the wrong position" (Langosco & et al. 2022, 1).

An analysis of the literature shows that in the context of the discussion of AI, both the problem of alignment itself and the direction of the search for its solution are presented as the elimination of "loopholes" and the achievement of the ideal of an exact specification, in which AI will not have the ability for unforeseen interpretations and actions within the environment to achieve the set tasks and receive rewards. This research and development goal is set clearly and almost in a military manner – which is understandable in light of one of the main areas of AI use – as achieving "objectivity," "accuracy," "literality" in the description of the environment, in setting tasks, and in the presentation of the results obtained by intelligent systems. An ideal AI system possessing these capabilities is characterized as "fully robust." An approach based on these expectations is open to attack from two directions at once.

From a theoretical perspective, the task so posed does not seem feasible. Any rational agent, natural or artificial, will have only a partial representation of the environment given in the senses or at its input, and only a partial theoretical model that captures the given in the senses. In more Kantian terms, a finite agent is confronted with only a portion of the manifold of all possible intuition and can cognize only what its *a priori* architecture allows for. In addition to these epistemic limitations, it will not grasp the existential or qualitative parameters of the situation of other agents, immersed, embodied or, to use existentialist language, "thrown" into the environment, each in its own unique way and experiencing it from its own unique perspective. These parameters are accessible only to themselves, and yet are essential for such things as humanity, freedom, responsibility, dignity – ultimately for the values with which it is proposed to align and which do not exist "objectively" or independently of concrete living people.

From the practical side, which is significant in the question of alignment, where values are involved, the solution of the problem posed in this objectivist and universalist way will lead to

negative, inhumane consequences, undermining the very values and intentions it is expected to promote. Let us assume for a moment that the theoretical problems were solved, and the programmers received an AI system with a model of the world that grasps and considers all relevant facts about the environment in a consistent manner. We can call such a model Laplacian, since it would embody – or encode – the features of the famous Laplace’s demon, that is, the classical standard of rationality of the early modern period, dating back to earlier ideas about the omniscient and non-contradictory Divine intellect. An attempt to implement this super-system would encounter the fact of the diversity of human models and pictures of the world and their discrepancy with the “high” standards of objectivity, universality, general significance, practical power, etc. achieved in AI. Alignment in this case would mean adjusting the diversity of human worldviews to the single artificial super-model – an operation that is directly opposite in its direction of fit to the humanistic goals stated in the definition of AI alignment.

Since alignment means bringing AI into line with humanity norms, I will focus below on the problems of human “alignment,” the clarification of which is a necessary condition for the success of AI attunement. I will claim that the inability to foresee and specify the full set of relevant objective parameters and range of ways to achieve objectives, to prevent reward hacking, and so on, seen as bugs in the AI discourse, are in fact features of humanity, arising from its openness and fallibility. I will start by offering a translation of some key terms of the AI alignment discourse into Kantian language, then, in the next section, I will examine Kant’s procedure of “alignment to values” by means of the categorical imperative and demonstrate the failure of expectations of obtaining an “objective” or “fully robust” alignment for humans, which also undermines expectations of alignment in the human-AI relation.

## **2. Alignment as enlightenment**

The Kantian analogue of the process of alignment is, no more and no less, enlightenment, understood for a person as a transition from minority to maturity, and in a generalized form as a transition of a rational system from heteronomy to autonomy. By heteronomy in the sense of Kant we can understand the alignment to external ends, by autonomy – the alignment to internal ends, that is, the proper ends of reason, recognized by the system endowed with reason as its own. Here an objection may arise that the alignment to human intentions and values for AI will be a condition of heteronomy, subordination to human volition, while autonomy would be the pursuit of emergent or spontaneous ends not controlled or instilled by humans. There can be at least two Kantian strategies for responding here. Firstly, one can challenge on philosophical-linguistic grounds the very possibility of meaningfully talking about intentions, values and goals of AI: only a living sentient being can have them in the proper sense, and attributing them to anything else is an incorrect attribution or a category mistake. Kant is only one of the proponents of this strategy (Hanna & Maiese, 2009, 15); Wittgensteinians, enactivists, and supporters of other programs will agree with it (Bennett & Hacker, 2021, 79ff). However, this strategy would involve reading into

Kant more naturalism than Kant himself would allow: embodiment accentuates the phenomenal and downplays the noumenal. Secondly, and this time in full accordance with Kant, one can claim that intentions, values, and goals are attributed not simply to human beings, but to human beings as a species of rational beings, that their true bearer is reason or intellect, for which they act as essential or necessary characteristics. Kant, of course, believed that these characteristics are universal in the strong sense of the word, that is, that they will be inherent in any reason or intellect regardless of the place, time, and way of its implementation. To what extent these two strategies, human-centric and intellect-centric, are compatible with each other is a separate and large question. What is important for us here is that any of these two strategies excludes a situation in which AI would have its own ends that could differ from the goals of human being – either from actual ends or from those that universal reason obliges humans to have. In both cases, human intentions and values will be the proper intentions and values for AI.

Further, Kant has a concept that captures the entire situation of the complete alignment of rational systems: the “kingdom of ends.”

The concept of every rational being as one who must regard himself as giving universal law through all the maxims of his will, so as to appraise himself and his actions from this point of view, leads to a very fruitful concept dependent upon it, namely that of *a kingdom of ends*.

By a *kingdom* I understand a systematic union of various rational beings through common laws. Now since laws determine ends in terms of their universal validity, if we abstract from the personal differences of rational beings as well as from all the content of their private ends we shall be able to think of a whole of all ends in systematic connection (a whole both of rational beings as ends in themselves and of the ends of his own that each may set himself), that is, a kingdom of ends, which is possible in accordance with the above principles (GMS 4:433; Kant, 1996, 83).

This quote contains the definition of a member or stakeholder in the “kingdom of ends,” or, as Kant adds on the same page, its sovereign (Oberhaupt): a being who is aware of and observes its role as a universal lawgiver and reflexively evaluates itself from this point of view. The experience of living such a role and the responsibility and freedom it provides is another complex qualitative or existential state, of which there is no reason to consider AI capable. AI is also incapable of learning a behavioral model that would replicate or mimic such a role – according to Kant, there is simply no data available for such learning:

In fact, it is absolutely impossible by means of experience to make out with complete certainty a single case in which the maxim of an action otherwise in conformity with duty rested simply on moral grounds and on the representation of one's duty (GMS, 4:406-7; Kant, 1996, 61).

It is even possible that moral actions are “actions of which the world has perhaps so far given no example” (GMS, 4, 408; Kant, 1996, 62). Only a critical examination of a moral consciousness or practical reason can clarify the “a priori” foundations of any moral experience that we can make intelligible to ourselves, and the only bearer of such practical reason available to our study is ourselves.

Kant’s answer to the question of possibility of full enlightenment of rational beings into the “kingdom of ends” is moderately optimistic: the “kingdom of ends” is a regulative idea that guides and ensures the slow moral development or progress of the human race in history, but we cannot hope for its full realization or even imagine this final state in any concrete way. Humanity is led to the “kingdom of ends” by the “moral compass” of the categorical imperative (GMS, 4, 404). Its first formula, known as the formula of universal law (FUL), describes a procedure of moral deliberation that, according to Kant, allows us to test whether a subjective principle or maxim of our particular action can be a universal law in the “kingdom of ends.” The fulfillment of the requirements of FUL by rational beings gradually, one by one, fills this “kingdom of ends” with a multitude of laws regulating the behavior of free and equal rational beings. It can be said that, even without being an optimist with regard to the full realization of the “kingdom of ends,” Kant was an approximationist, that is, he believed in a gradual approach to it, and a cumulativist, that is, he believed that maxims that have passed the test become universal laws in the strong sense, i.e. for everyone and forever. Thus, the feasibility of human alignment in the Kantian framework depends not on a fully specified grand plan of an end-state, but on the careful observance of the procedure of moral verification and the subsequent fulfillment of the selected maxims.

In terms of execution, we face almost insurmountable difficulties caused by human inclinations and self-love. However, Kant is optimistic about the feasibility of a rational test that yields clear moral knowledge. In the next section, I will try to explain the reasons for a more reserved or fallibilist attitude to the results of moral deliberation according to FUL as a path to alignment in the “kingdom of ends.” The fallibilist position entails the rejection of approximationist and cumulativist hopes for constructing a “kingdom of ends” by selecting maxims. In turn, skepticism about approaching a state of alignment among humans with our varied intentions and actions entails skepticism about bringing AI into alignment with humanity. Alignment in this understanding remains a crucial goal, but its achievement will inevitably be partial, local, open to revision, and carrying unavoidable risks.

### **3. Alignment under the first formula of categorical imperative: standard interpretation**

The “kingdom of ends” is thus an imaginary end state, slowly constructed by a multitude of rational beings in history. Just as enlightenment for Kant is not a result, an “age” that has arrived, but a process, the “kingdom of ends” is concretized not as a state of affairs but as a complex deliberative process of a multitude of rational agents, as a great act of reasoning. This explains the central importance attached by both Kant and his readers to the description of the procedure of such

reasoning – the first formula of the categorical imperative, known as the formula of the universal law. It reads: “act only in accordance with that maxim through which you can at the same time will that it become a universal law” (GMS, 4, 421; Kant, 1996, 73). Kant further specifies FUL with the formula of the law of nature (FLN) for rational beings immersed in nature, understood as “the existence of things insofar as it is determined in accordance with universal laws” (GMS, 4, 421; Kant, 1996, 73). “act as if the maxim of your action were to become by your will a *universal law of nature*” (GMS, 4, 421; Kant, 1996, 73). John Rawls and his followers deploy Kant’s moral deliberation as a step-by-step “CI procedure”:

- (1) I am to do *X* in circumstances *C* in order to bring about *Y* unless *Z*. (Here *X* is an action and *Y* is an end, a state of affairs) [...]
- (2) Everyone is to do *X* in circumstances *C* in order to bring about *Y* unless *Z*. [...]
- (3) Everyone always does *X* in circumstances *C* in order to bring about *Y*, as if by a law of nature (as if such a law was implanted in us by natural instinct [added in Rawls & Herman, 2000]). [...]
- (4) We are to adjoin the as-if law of nature at step (3) to the existing laws of nature (as these are understood by us) and then think through as best we can what the order of nature would be once the effects of the newly adjoined law of nature have had sufficient time to work themselves out. (Rawls, 1989, 499–500; Rawls & Herman, 2000, 167–69)

This procedure serves as an explication of the tasks facing any rational autonomous system that participates in the creation of a state of alignment, or, in Kantian terms, that is conscious of itself in the role of a universal lawgiver for the “kingdom of ends.” We can ask a question that is Kantian in form: how is the process of alignment possible? What conditions must be fulfilled by rational agents in order for “a systematic union of various rational beings through common laws” to emerge as a result of their many deliberations? Both Kant and many of his readers expect that this deliberation ought to correspond to the standards of classical rationality: systematic unity in this case is possible through consistency, completeness, unambiguity, immutability of the deontic qualifications of all the actions of all rational beings. More specifically – and perhaps somewhat idealized – these expectations can be represented by a series of theses that I will call the standard interpretation:

- a) There is (or ought to be) *only one* correct, or objective, or relevant description (specification) of action *X*, end *Y*, circumstances *C*, conditions *Z*.
- b) There is *only one* correct way to generalize the description of *X* to the maxim *M*.

c) The verdict is issued *for all agents*, that is, it is completely independent of a particular agent and their situation, or is context-independent.

d) The CI procedure, if done correctly, always generates a *definitive* or *monotonic* deontic verdict. Any revision of the result is prohibited.

e) Maxims are tested *one at a time* against the background of a *fixed worldview* or *world model*. A maxim can be accepted or rejected without affecting any other parts of the worldview, including other maxims. In other words, the standard interpretation implicitly accepts the principle of *atomism of maxims*, as well as the principle of *ceteris paribus* (other things being equal) and/or the principle of *ceteris absentibus* (other things being absent).

Numerous opponents as well as many sympathizers of the Kantian program find these requirements unfeasible, and for varying, and even opposing, reasons. One line of criticism charges Kant's approach with "empty formalism": "it is impossible to make the transition to the determination of particular duties from the ... determination of duty as *absence of contradiction*, as *formal correspondence with itself*, [...] it is possible to justify any wrong or immoral mode of action by this means" (Hegel, EPR § 135; Hegel, 1991, 162). A "sufficient ingenuity" of a deliberating agent in describing the intended action allows one to achieve universalizability of the maxim in almost any case (e.g. Anscombe, 1958, 2; MacIntyre, 1966, 197–98). So, fulfilment of the steps of the CI and their full correspondence to the formal requirements listed in the five points does not promise anything regarding the actual morality of the resulting system of laws. The formally impeccable "kingdom of ends" [Reich der Zwecke], that is, complete alignment, may turn out to be a terrible dystopia, a fully rationalized concentration camp. In the view of other critics, on the contrary, Kant offers a rigorist model that produces "obligation overload" (O'Neill, 2013 (1975), 135) or is "overdemanding" (Sticker, 2019). The consensus is that moral deliberation on the categorical imperative "systematically yields false positives and false negatives" (Wood, 2006, 345). Thus, the procedural path to the "kingdom of ends" and the ideal of alignment of intelligent agents turns out to be impossible.

The Kantian literature offers a more fine-grained analysis explaining how these difficulties arise – and, importantly for our purposes here, it does so in terms similar to the description of difficulties in the literature on AI alignment. Thus, already in the first step of the CI procedure, the problem of choosing a "relevant description of the act" arises, which is essentially the same as the problem of "specifying" the situation (environment) and task for AI. And in the second step, the problem of "choosing the level of generalization of the maxim" arises, a mistake in which lead to a situation of "misgeneralization," similar to the analogous difficulty in AI. In addition, the Kantian literature discusses the complex of problems of "transcendental illusion" (Grier, 2001), self-deception and rationalization (Papish, 2018; Muchnik, 2019), "natural dialectic" (Sticker, 2017), "logical egoism" and "maxim-fiddling" (Sneddon, 2011), similar to what is referred to in the AI literature as "specification gaming" and "hallucination." In all these cases, rational agents either unintentionally

or intentionally produce models of the world that do not correspond to reality, and act in accordance with these models. Here we will have to limit ourselves to indicating the similarity and relevance of this theme, long discussed in the Kantian literature, to the current debates in AI.

With regard to the problem of choosing a relevant description, a pluralistic consensus has already emerged in the Kantian literature: “any action admits of a wide variety of true descriptions” (Timmons, 2017, 60). Many scholars believe that in practice the problem of choice is resolved by social conventions or “rules of moral salience” (Herman, 1993, 77ff), which are historical, diverse and poorly amenable to further Kantian rationalization and ordering. People describe their actions and derive their maxims in the socially accepted manner, focusing on the opinions of others and worrying about these opinions. If so, then belonging to a moral community and experiencing one’s dependence on its judgments, caring about belonging and status are necessary conditions for alignment to the intentions and values accepted among the community. It is difficult to imagine that AI could possess these properties. But even in human performance, this process of adaptation to unclear and changing morals has a permanent character. Universalism in the strong sense of the word is impossible; Kantian ethics does not fulfill the high expectations of its creator and readers. For many, this means failure and discrediting of the Kantian approach in its entirety. For the problem of AI alignment, this result means the impossibility of an “objective” or “completely stable” specification, based on which AI could act robustly and in full accordance with human intentions and values. The problem again turns out to lie not in AI, but in humans.

As in Kant’s time and earlier, the reaction to the failure of philosophical construction may be either dogmatism or skepticism. Dogmatism will impose the universality of principles and the unambiguity of moral assessments derived from them by non-philosophical means, that is, by deception and violence. It will invent mandatory relevant descriptions and impose them by force and cunning – in the extreme similarly to what Viktor Klemperer described for the Third Reich (Klemperer, 2013, [1947]). In the case of AI, it is easy to imagine a dystopian situation in which not AI, but humans, unable to generate a moral system that meets high requirements, will adapt to a rigid artificial language. Alignment in this case will take the opposite “direction of fit,” humans-to-AI, and for technocratic rationality this may seem like a necessary solution, the lesser of the available evils. Skepticism will expose these attempts and survive with the heavy consciousness of the impossibility of any ethos. However, it seems that the Kantian strategy of alignment can be maintained if the threshold of expectations expressed by the five standard points is lowered. The result is a fallibilist model that excludes the completion of the construction of the “Tower of Babel” of the universal normative system, but allows for piecemeal improvement of our condition.

#### **4. Kantian fallibilist alignment**

The fallibilist model can be represented by relativizing and weakening the five theses:



- a) There is *more than one* correct or relevant description or specification of action X, goal Y, circumstances C, conditions Z—the plurality of descriptions thesis.
- b) There is *more than one* correct way to generalize the description of X to the maxim M—the plurality of generalizations thesis.
- c) The verdict is made for *agents similar to the reasoner*, that is, those having sufficiently similar ways of framing the X, Y, C and Z—the contextuality thesis.
- d) The CI procedure is *non-monotonic* and generates a *modifiable* deontic verdict—the defeasibility thesis.
- e) The maxims are tested *collectively* and *together with the background*—the maxim holism thesis.

The plurality of descriptions or specifications is an outcome of the absence of a privileged meta-position from which an “objective” specification could be constructed. One could call it the principle of perspectivism and, if desired, trace its genesis from Kant’s theory of knowledge, as some modern perspectivists do (e.g. Massimi, 2017; 2018). The nature that must be specified and to which a maxim in the form of a new law must then be added is given as an appearance in the senses to some agent. This agent possesses some *a priori* way of organizing experience, but all the same this way of organizing will have a historical, evolutionary, local, particular, contingent character, falling short of the Kantian “classical” standard of *a priori* as a universal and ahistorical set of necessary forms or capacities of reason in general. A relatively more “robust” *a priori* will act as a form for a more variable and obviously more local *a posteriori* content. In such a situation of comparatively greater pluralism of content or explanandum and comparatively lesser, but still present and legitimate, variability of form or explanans, a plurality of descriptions is inevitable. Moreover, it takes place not only between theoretical frameworks, paradigms, worldviews, cultures, but also within them: not being complete and coherent, they all allow ambiguity, imprecision, diversity of modes of presentation and evaluation. Such is the human condition, and the appearance of AI in it does not change anything in this aspect – except, perhaps, the appearance of a reason for the already described unification through coercion to an “objective” AI-centric specification, that is, to the aligning of human to AI (and its owners, if any remain). A somewhat more technical way to explain the thesis of multiple descriptions is to assert underdetermination of our beliefs about nature and its laws that has ramifications for moral deliberation (Baumann, 2019; 2022; Чалый, 2022), and the resulting pluralism of languages, worldviews (ontologies) and specifications built within their frameworks.

The plurality of generalizations thesis is partly a continuation of the previous one. The operation of description or specification contains categorization or subsumption under concepts, in which we identify something as a special case of a certain type. Generalization is the inverse operation and therefore is already embedded in the first, as, according to Leibniz’s metaphor, Hercules is embedded in a block of marble. But, contrary to Leibniz’s also classical way of thinking, in the

fallibilist view it is embedded without necessity, not as the only “innate” possibility. In other words, an individual specification can be generalized in a different way, by creatively discovering in the existing description the features of something different, by examining and subsuming it under other categories. This possibility is prerequisite for freedom, belongs to the realm of art and is not subject to complete rationalization. The plurality of generalizations is also not a failure or error of the intellect, but the norm – and if we consider the connection with freedom, then an essential necessity – of the human condition. AI does not add anything new here either. By imitating ever more accurately the mode of operation of the human intellect, it will also, successfully or unsuccessfully, reproduce this capacity for creative transgression, narrowly perceived as “misgeneralization.”

The contextuality thesis underlines not only the locality and historicity of the features of characteristics and generalizations as concrete speech acts, but also their dependence on the general background of beliefs that sets the meanings available for use in specifications and generalizations. For moral deliberation, this means a shift from universalism towards situationism: no two situations are identical, equally captured and classified in terms of universal principles, but there are acts that agents commit under conditions of uncertainty, still bearing moral responsibility. For the discussion of AI alignment, this means that there are no separate words or terms that self-contain their fixed meanings and together form a precise language, by means of which complete “objective” alignment could be achieved and fixed. Such understanding of the work of reason and language goes back to logical positivism and persists among some computer scientists and AI developers, despite post-positivist programs and their implications for the exact sciences with their attempts to represent and change reality. Instead, the specifications of both human and AI acts depend on a particular community of language speakers in a given environment facing a set of challenges and can only be considered in this complex. Alignment is an adaptation to a specific complex. The linguistic aspects of contextuality are described, for example, by François Recanati (2007), the moral-political aspects are discussed in the classic works of the communitarians, such as Alasdair MacIntyre (1988).

The non-monotonicity thesis means that no specification or generalization of it can claim to be final, but is always open to revision in the light of new information (Koons, 2022). Descriptions, maxims, and whole worldviews are defeasible – and therefore, the moral judgments or deontic verdicts obtained within their framework and on their basis are also defeasible. Even if we agree to regard certain general principles expressing immutable values as unchanging or monotonic – for example, human rights – the meaning of these formulas, and even more so the results of the evaluation of specific situations in their terms, will not be monotonic or unchanging. Kant, in essence, characterizes our reasoning or plans for achieving happiness or bliss (*Glückseligkeit*) as non-monotonic: constantly and rapidly changing ideas about the content of this elusive state entail constant changes in the chains of pragmatic imperatives aimed at achieving this state (GMS, 4, 417-19). This situation means that not only the specifications of human values, but even more so

the human intentions that figure in the formulation of the task of alignment, are too fluid to leave hope for a thorough and permanent solution.

Finally, the maxim holism thesis means that it is wrong to regard the universalizability tests of subjective principles of our actions as isolated or atomic acts. In fact, the entire framework by which and within which the maxim is formulated is tested each time. The standard understanding of the Kantian deliberation suggests that the failure of a maxim to be universalizable entails only one consequence: its disqualification. The fallibilist understanding indicates that this is only one option. We are also entitled to look for error in other parts of the framework: in adjacent or distant maxims that have once been tested, in beliefs about the laws of nature, in the implicit presuppositions on which our worldview rests. Everything is open to critical examination, the result of which may extend to a complete revision of the worldview or a paradigm shift. Of course, a more extensive revision is also more expensive, but sometimes, when the action in question is especially important to us, the game is worth the candle. A personality or moral character can change over one pivotal act (“Crime and Punishment” gives a famous example). In application to the problem of alignment, this means that the operation of bringing some rule of AI action into line with human intentions and values affects not only other rules, specifications, generalizations, but also human intentions and even the understanding of values to which the AI is fit. By changing the behavior of the system being adjusted, we change as well.

### **Alignment and relativism**

If we change, is there still some basis to which we could align ourselves with some success? The Kantian framework allows us to answer this question in the affirmative. There is a “hard core” of values that cannot be falsified or abandoned. This is the value of humanity, which appears in the second formula of the categorical imperative, associated with rationality, freedom and responsibility, with good will and critical reflection. However, the interpretation of these basic concepts and, as a result, the demands that follow from them can change and has changed in different theories and historical-cultural situations. For example, since Kant’s time, the word “humanity” (*Menschheit*) has been gaining in its naturalistic “phenomenal” meaning and losing its non-naturalistic “noumenal” meaning along with the naturalization of the worldview under the influence of natural sciences and other factors. It was losing its essentialist meaning of a set of universal properties or virtues and was accumulating an existentialist meaning of radical freedom – for example, in English-language Kantian literature, the dominant understanding of humanity is that of “the capacity to set oneself an end – any end whatsoever” (MS, 6,392), which for Kant was only part of a complex answer to the main question of philosophy. In Soviet Marxism, “humanity” was understood primarily as its concrete vanguard – the proletariat led by the Communist Party. In colonial rationalizations (and poeticizations, for example, in Kipling) the pinnacle of humanity could be the white man bearing the burden of enlightening uncivilized peoples. The terrain of meanings of “humanity” is riddled with the trenches of past and ongoing wars, and any attempt to

align the AI to this concept automatically makes the attempter a combatant with all the ensuing consequences. In some cases, the discussion of AI alignment becomes a springboard and the shortest path to the epicenter for those who hope to win these disputes, change human nature or the idea of it, and enter the battle consciously. Finally, in relation to humanity, one can also adopt a mystic point of view, similar to the position of apophatic theology in relation to the concept of God: humanity cannot have a definition, cannot appear as a clear and distinct list of attributes and is doomed to remain a mystery to us, humans, thus leaving the question “alignment to what?” unanswerable.

One way or another, the struggle for the content and scope of basic concepts such as humanity, value, freedom, reason, and good will nevertheless preserves their basic status, that is, it preserves the framework for the ongoing process of alignment. As long as this is so, we are not in a situation of relativism, although we are quite far beyond the rigid universalism of the standard interpretation.

## Conclusion

AI alignment is a two-way process in which humanity as the determining party is more important than AI as the determined party. Without understanding what humanity, human intentions and values are, we cannot hope to concretize alignment as a technical task. So far it exists only as an important but vague problem. This problem of alignment is a special case of the more general problem of harmonizing the intentions, values and resulting behavioral patterns of intelligent beings, which has long been posed in relation to humans and also does not have a clear and unproblematic solution. Among the existing strategies for alignment of intelligent beings, the Kantian one is characterized by its particular detail and systematicity. In the course of discussions and criticism, the Kantian framework has accumulated a range of nuanced concepts and arguments, relevant to the context of AI. The recent decades of systematic collective work by scholars of Kantian ethics have shown the impossibility of a standard, or classical, or strictly universalist model that would make complete and final alignment achievable. The state of the debate pushes us to accept, instead of strict universalism, a fallibilist interpretation of Kant’s procedure for alignment, in which it is impossible as an actual state, but remains as a regulative idea, the orientation towards which allows us to gradually resolve specific ethical issues, that is, to achieve local and temporary alignment of intelligent agents. Such a view seems fruitful for understanding the state of affairs and the prospects for AI alignment.

## References

- Baumann, M. (2019). Consequentializing and Underdetermination. *Australasian Journal of Philosophy*, 97 (3), 511–27. <https://doi.org/10.1080/00048402.2018.1501078>
- Baumann, M. (2022). Moral Underdetermination and a New Skeptical Challenge. *Synthese* 200 (3), 208. <https://doi.org/10.1007/s11229-022-03529-w>

- Bennett, M. R., & Hacker. P. M. S. (2021). *Philosophical Foundations of Neuroscience*. John Wiley & Sons.
- Future of Life Institute. Asilomar AI Principles. *Future of Life Institute* (blog). <https://futureoflife.org/open-letter/ai-principles/>
- Gabriel, I. (2020). Artificial Intelligence, Values, and Alignment. *Minds and Machines*, 30 (3), 411–37. <https://doi.org/10.1007/s11023-020-09539-2>
- Grier, M. (2001). *Kant's Doctrine of Transcendental Illusion*. Cambridge University Press.
- Hanna, R. & Michelle M. (2009). *Embodied Minds in Action*. Oxford University Press.
- Hegel, G. W. F. (1991). *Elements of the Philosophy of Right*. Edited by A W. Wood. Translated by H. B. Nisbet. Cambridge University Press.
- Herman, B. (1993). *The Practice of Moral Judgment*. Harvard University Press.
- Ji, & et al. (2024). *AI Alignment: A Comprehensive Survey*. arXiv. <http://arxiv.org/abs/2310.19852>
- Kant, I. (1996). *Practical Philosophy*. Edited & translated by M. J. Gregor. Cambridge University Press.
- Kim, H. & Dieter S. (eds). (2022). *Kant and Artificial Intelligence*. Walter de Gruyter GmbH & Co KG.
- Klemperer, V. (2013). *Language of the Third Reich*. Bloomsbury Academic.
- Koons, R. C. (2022). Defeasible Reasoning. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/sum2022/entries/reasoning-defeasible/>
- Langosco, & et al. (2022). Goal Misgeneralization in Deep Reinforcement Learning. In *Proceedings of the 39<sup>th</sup> International Conference on Machine Learning*, 12004–19. PMLR. <https://proceedings.mlr.press/v162/langosco22a.html>
- Leike, & et al. (2018). *Scalable Agent Alignment via Reward Modeling: A Research Direction*. arXiv. <https://doi.org/10.48550/arXiv.1811.07871>
- MacIntyre, A. C. (1966) *A Short History of Ethics*. Macmillan.
- MacIntyre, A. C. (1988). *Whose Justice? Which Rationality?* University of Notre Dame Press.
- Massimi, M. (2017). What Is This Thing Called ‘Scientific Knowledge? – Kant on Imaginary Standpoints and the Regulative Role of Reason. *Kant Yearbook* 9 (1), 63–84. <https://doi.org/10.1515/kantyb-2017-0004>
- Massimi, M. (2018). Points of View: Kant on Perspectival Knowledge. *Synthese* 198 (S13), 3279–96. <https://doi.org/10.1007/s11229-018-1876-7>
- Muchnik, P. (2019). Laura Papish, Kant on Evil, Self-Deception, and Moral Reform, Oxford University Press, 2018 pp. Xvii + 280 Isbn 9780190692100 \$85.00.” *Kantian Review* 24 (2), 316–22. <https://doi.org/10.1017/s1369415419000104>
- O’Neill, O. (2013). *Acting on Principle: An Essay on Kantian Ethics*. 2<sup>nd</sup> edition, Cambridge University Press.
- Papish, L. (2018). Kantian Self-Deception. In *Kant on Evil, Self-Deception, and Moral Reform*, edited by Laura Papish, Oxford University Press. <https://doi.org/10.1093/oso/9780190692100.003.0004>

- Papyshev, G. & Migliorini, S. (2024). *Developing a Liability Framework for Harms Arising out of Specification Gaming*. In. <https://openreview.net/forum?id=pU9QUQGsuc>.
- Rawls, J. & Herman, B. (2000). *Lectures on the History of Moral Philosophy*. Harvard University Press.
- Rawls, J. (1989). Themes in Kant's Moral Philosophy. In *Kant's Transcendental Deductions: The Three Critiques and the Opus Postumum*, 80–113. Stanford University Press.
- Recanati, F. (2007). *Perspectival Thought: A Plea for (Moderate) Relativism*. Clarendon Press.
- Sneddon, A. (2011). A New Kantian Response to Maxim-Fiddling. *Kantian Review* 16 (1): 67–88. <https://doi.org/10.1017/s1369415410000087>
- Sticker, M. (2019). Kant, Moral Overdemandingness and Self-Scrutiny. *Noûs* n/a (n/a): 1–24. <https://doi.org/10.1111/nous.12308>
- Sticker, M. (2017). When the Reflective Watch-Dog Barks: Conscience and Self-Deception in Kant.” *Journal of Value Inquiry* 51 (1), 85–104. <https://doi.org/10.1007/s10790-016-9559-4>
- Timmons, M. (2017). *Significance and System: Essays in Kant's Ethics*. Oxford University Press.
- Wood, A. W. (2006). The Supreme Principle of Morality. In *The Cambridge Companion to Kant and Modern Philosophy*, Edited by P. Guyer, 342–80. Cambridge University Press.
- Чалый, В. А. (2022) К кантианскому моральному фаллибилизму: недоопределенность в рассуждениях по первой формуле категорического императива. *Вестник Московского Университета. Серия 7. Философия* 1, 105–14.