Undergraduate Philosophy Journal of Australasia

Volume 2

Issue 2

December 2020

UPJA

Founded in 2019, the *Undergraduate Philosophy Journal of Australasia* (UPJA) is the first undergraduate philosophy journal run by students from the Australasian region. We publish two issues and host two virtual conferences annually: one mid-year and another at the end of the year. Our calls for papers for each issue open roughly in February and August, respectively. We aim to be an inclusive and diverse journal that welcomes submissions on any philosophical topic attempting to make a substantive contribution to contemporary debate. Submissions from women and other members of underrepresented groups in philosophy, including those for whom English is not their first language, are particularly encouraged.

Undergraduate Philosophy Journal of Australasia – UPJA

Undergraduate Philosophy Journal of Australasia

Mailing list: https://forms.gle/we6ZeihAx1hDKf6E6

contact@upja.online

www.upja.online

REFEREES

Vlademire Kevin D. Bumatay is an instructor in philosophy from the University of the Philippines Baguio, Republic of the Philippines. His main areas of research are general philosophy of science, especially confirmation theory; metaphysics, especially naturalistic metaphysics; and epistemology, especially formal epistemology.

Kristina Dukoski is pursuing her Master's degree in philosophy at the University of Toronto. Her research interests include topics in the philosophy of sex and epistemology.

James Eder is a Master's student in philosophy at the University of Canterbury, New Zealand. He is researching technological bioenhancement and autonomy, and he is interested in ethics and policy analysis.

Alejandro Gracia Di Rienzo is a Master's student in philosophy at the University of Salamanca in Spain. He graduated from philosophy at the Universidad Autónoma de Madrid with a thesis on historical causation. His current interests include logic, the metaphysics of causation, and the philosophy of mathematics.

Abi Gunning recently graduated with Honours in French and philosophy from the University of Edinburgh. Her interests include feminist philosophy, social and political philosophy, epistemology, and philosophy of language.

Isaac Hadfield recently graduated from the University of Oxford with an undergraduate degree in philosophy, politics, and economics. He is particularly interested in social ontology, ethics, and philosophy of language.

Levi Haeck has recently commenced a PhD at Ghent University, and specialises in Kant's transcendental logic and how this discloses a subtle yet fundamental relation between logic and metaphysics.

Frank Hernandez is an undergraduate philosophy and multimedia journalism student at the University of Texas at El Paso. He is interested in epistemic injustice, epistemic justification, and contemporary issues in philosophy of language.

Brylea Hollinshead recently completed her undergraduate studies at the University of Canterbury majoring in philosophy and classics. She is passionate about ancient philosophy, philosophy of emotion, and existential philosophy, and interested in how these fields contribute to questions in ethics and meaning in life.

Ruby Hornsby is pursuing her Master's degree in philosophy at the University of Leeds, England. Her research interests predominantly lie in ancient philosophy, with recent work focused on Platonic conceptions of pleasure, friendship, and the extent to which they are welcome in the good human life.

Vincent Lee has just begun his PhD in philosophy at the University of Toronto, having graduated from Yale-NUS College in 2019. He is interested in bringing into dialogue philosophies from the classical Indian Buddhist and early modern European traditions, with particular regard to how their metaphysics bear differently on common notions like substancehood and selfhood.

João Almeida Loureiro is a Master's student in philosophy at NOVA University of Lisbon. He is presently researching Bergsonian philosophy and the implications of the concept of duration for a new philosophy of history.

Kael McCormack-Skewes is a PhD candidate at the University of New South Wales whose research focuses on moral psychology.

Kai Milanovich is studying philosophy and music at the University of Missouri-Kansas City. They are interested in ethics and education, but really they just want to go for a walk.

Ariane Moore is a PhD candidate at the University of Tasmania. Her primary interests include philosophy of religion and atheism studies.

Katie Prosser is a second-year student pursuing a BPhil in Philosophy at the University of Oxford. Her main research interests are animal ethics and climate ethics, primarily focusing on the intersection of these topics, and approaching them from a notably feminist viewpoint.

Stephen Rowe is a Master's student in psychology at the University of Canterbury. He has a keen interest in how the descriptive can inform the normative, particularly in relation to debates about mental illness and how we can live meaningful lives.

Madeleine Shield is currently undertaking a PhD with the School of Historical and Philosophical Inquiry at the University of Queensland, Australia. Her research to date has focused on continental philosophy, and she is currently developing a specialised interest in the philosophy of emotions field.

Kaitlin Smalley is a white, non-Indigenous PhD candidate at the University of Queensland, who lives and works across Jagera, Turrbal, and Quandamooka Land. Her research explores epistemic prejudice in the discipline of philosophy – in particular, the epistemic exclusion of Aboriginal and Torres Strait Islander philosophies/philosophy from Australian tertiary philosophy education.

Zachary Swindlehurst is a PhD candidate in philosophy at the Australian National University. He is interested in philosophy of mind, epistemology, and metaethics.

Kramer Thompson is currently a philosophy PhD student at the Australian National University. He has a broad interest in philosophy, but focuses on theory of action, ethics, metaphysics, and epistemology.

Joel Towell graduated from Honours in philosophy at Monash University in 2018. He is particularly interested in logic, philosophy of language, and epistemology.

EDITORS' NOTE

This year has brought many unexpected obstacles. The COVID-19 pandemic continues to challenge our way of being and perceptions of normality. This has forced students to adapt to a new way of life, with disruptions to on-campus learning having changed the landscape of tertiary education. As such, we are proud that Volume 2, Issue 2 has provided a means of collaboration for undergraduate philosophy, a locus for the free movement of ideas at a time when physical movement is severely restricted.

Additionally, there have been significant changes to the funding of humanities programmes in Australian universities, providing further roadblocks in a climate where academic philosophy must continually justify its existence. For our latest Voices from the Region and Beyond section, we posed the question 'Why is philosophy relevant today?'. The responses published in this issue remind us of the importance of philosophy and its relevance to scholarly, political, social, and personal domains.

With international travel greatly limited this year, we are pleased that UPJA has been a platform for dialogue between emerging philosophers around the world. This edition of the journal saw our pool of referees expand to include students and recent graduates from institutions worldwide, generating substantial interest from applicants in several parts of the Northern Hemisphere alongside the Australasian community. Our team of 22 referees for this issue spans Australia, Belgium, Canada, England, New Zealand, the Philippines, Portugal, Scotland, Spain, and the United States. Also, in November we hosted our second Virtual Conference for Undergraduate Philosophy. This involved talks from seven student presenters in five countries, a keynote address by Associate Professor Stephanie Collins, and many engaging discussions with attendees from across the globe.

Given UPJA's growing international presence, it was perhaps unsurprising to observe the regional diversity among those who submitted a paper for this issue. We received 42 submissions from students at 32 different institutions in 11 countries. In addition, we are pleased to report that half of these submissions (and two-thirds of referee applications) came from individuals who identify as members of underrepresented groups in philosophy. Three papers are published in the present issue, giving a competitive acceptance rate of 7%.

In 'A Kantian Take on Mind Extension', Levi Haeck (Ghent University) provides a thorough examination of Immanuel Kant's conception of the mind and its (inter)relationship with the world. Recognising that Kantian theories are lacking within recent dialogue in philosophy of mind, Haeck then assesses Andy Clark and David Chalmers' extended mind thesis through a Kantian lens. This theory, Haeck argues, fails to adequately dispel the dualistic Cartesian opposition between 'mind' and 'world'. Accordingly, Kant's transcendental idealism – which Haeck suggests

does successfully refute Cartesian idealism – may provide underappreciated insights into contemporary notions of mind extension.

In a similar vein, Ruby Hornsby (University of Leeds) reflects on historical ideas of the good life and their application to various contexts with modern-day relevance in 'The Nature of Pleasure in Plato's *Philebus*'. Through a detailed analysis of Plato's account, Hornsby identifies two overarching varieties of pleasure: impure and pure. The former consists of the restoration of harmonious conditions, while the latter involves the actualisation of certain potential states. In either case, Hornsby maintains, a process of *change* is essential to the occurrence of pleasure.

Matthew W. G. McClure (University of Edinburgh) focuses on a very different sort of change in 'Star Models and the Semantics of Infectiousness': a modification to logical systems. More specifically, McClure demonstrates how a novel variety of star semantics may be incorporated in forms of logic with many-valued semantics that include the truth-value 'indeterminate'. McClure goes on to explore three possible ways of interpreting the indeterminate truth-value – the nonsense, off-topic, and emptiness interpretations – all of which, they argue, appear just as compatible with star semantics as many-valued semantics. In an interesting parallel with ongoing epidemiological events, McClure concludes by discussing how infectious logics might best be 'quarantined', and the means through which star semantics may navigate this challenge.

We are pleased to award two prizes for this issue. Levi Haeck receives the prize for Best Paper, and Ruby Hornsby receives the prize for Best Paper (Member of an Underrepresented Group in Philosophy). Both of these are funded by the Australasian Association of Philosophy, whose continued support of UPJA has been invaluable.

Indeed, many organisations and individuals contributed to the production of Volume 2, Issue 2. We are thankful to the numerous student philosophy societies and Minorities and Philosophy chapters for circulating our call for papers; our team of referees for providing such erudite and constructive reports on submissions; and our three faculty advisors – Associate Professor Stephanie Collins, Assistant Professor Sandra Leonie Field, and Dr Carolyn Mason – for their ongoing assistance. Thanks also to Dr David Ripley and Raphael Morris for sharing their philosophical expertise.

Finally, we thank our hardworking and dedicated Associate Editors, Alan Bechaz and Racher Du. Their creative ingenuity, philosophical insight, and tenacity in producing this issue were second to none. It is with great pride that we pass on to them the roles of Editors-in-Chief for Volume 3, Issue 1.

Rory Collins and Anita Pillai

December 2020

VOICES FROM THE REGION AND BEYOND
"WHY IS PHILOSOPHY RELEVANT TODAY?"

Surveying our multifarious fields of knowledge, it might give one pause to notice that each topic was begotten by philosophy. Indeed, as human knowledge progressed and philosophy birthed new fields, philosophy ceded elements of its esteemed position; and today, it has perhaps lost the position of seniority it once enjoyed – at least in popular thought. That philosophy is the forefather of many modern fields should not be forgotten. And where such fields go astray, it remains the role of philosophy to postulate their rebirth. Philosophy postulates what is to be known, science ensures it is known: both work in harmony.

Jason Horvatic, University of Western Australia

As the world's economies suffer under the weight of the pandemic and the worst off bear the brunt of adversity, it is more important than ever that everyone, not just those who can access formal education, be equipped with the tools to ask big questions about justice, equity, and good governance.

Bree Booth, University of Melbourne

Philosophy in its ancient Greek roots translates to a love of wisdom, but what is so relevant about this discipline today? 'Philosophy' explicates the advanced, unautomatable linguistic skills to accurately paraphrase even complicated positions into no more than three sentences. Students encyclopaedically memorise forms of fallacious argument and exhaustively practice deductive proofs to become competent researchers, policy analysts, entrepreneurs, and teachers. Instructors earnestly facilitate an environment where students seek and do not fear criticism. As we foresee workforces en voyage for greater automation and a politics of emotions, I cannot see how philosophy is anything but relevant today.

Jack Hudson Stewart, University of Western Australia

Philosophy is relevant today as a coping mechanism for internal demons, a filter for the modern communications bombardment, and a validity check on social narratives. Its value is immeasurable, and its influence is undeniable. The art of thinking is the very core of what it means to be human, a thinking thing.

Christopher Elwell, Charles Sturt University

TABLE OF CONTENTS

* Winner of Best Paper

† Winner of Best Paper (Member of an Underrepresented Group in Philosophy)

Undergraduate Philosophy Journal of Australasia www.upja.online
VOL. 2, NO. 2, DECEMBER 2020

A Kantian Take on Mind Extension

LEVI HAECK[*]

GHENT UNIVERSITY

Abstract

I assess Andy Clark and David Chalmers' groundbreaking exposition of the extended mind thesis (EMT), as originally put forward in 1998, from the viewpoint of Immanuel Kant's transcendental idealism. Both stances are committed to investigating how extension might be constitutive of the mind, yet they do so on completely different terms. In Section 1, I set out how Kant relativises the Cartesian distinction between mind and world by showing how the very internality of the mind is necessarily constituted in relation to extension, giving rise to the suggestion that the mind is an activity. In Section 2, I use this Kantian dynamic to assess Clark and Chalmers' claim that at certain times and under certain conditions the mind is extended into the world. Although they compellingly show that the functions of the mind are sometimes *taken over* by the world, a close reading of their text reveals that this does not really challenge the Cartesian opposition between mind and extension. This allows for the conclusion that Kant's eighteenth-century approach to EMT stands much further from Cartesianism – but also from computationalism – than its twentieth-century competitor, thus preluding an alternative and perhaps more radical pathway to conceptualising mind extension.

1. Immanuel Kant

1.1 A Kantian Take on 'Basic Minds'

Immanuel Kant's epistemological legacy originates from his attempts at investigating the necessary conditions of possibility of the object (*KrV*, B xvi).[1] In this respect, Kant's philosophy of transcendental idealism (TI) has given way to the widespread yet powerful assumption that the objectivity of the empirical world is constituted by *a priori* (i.e., non-empirical) subjective conditions.[2] Broadly speaking, TI adheres to the

---

[*] *Levi Haeck has recently commenced a PhD at Ghent University, and specialises in Kant's transcendental logic and how this discloses a subtle yet fundamental relation between logic and metaphysics.*

[1] As custom dictates, all references to Kant cite from the *Akademieausgabe*, indicating title abbreviation (for instance *KrV* for the *Critique of Pure Reason*), followed by volume number and page number. Moreover, for the *KrV*, citations refer to the A and B editions of this work. These citations are inserted in the text itself.

[2] According to Kant, concepts are objective (have objectivity) when they can be related to sensible intuition. Sensible intuition, on the other hand, is objective when it can be related to concepts (see *KrV*, B 75). The term 'object' itself, then, should be understood as that in which conceptuality and sensibility are united into a whole (see *KrV*, B 137).

philosophical strand that the world is, in a sense, *made possible* by the mind – by conditions that 'belong' to the mind. Kant's 'Copernican revolution' might indeed give way to the idea that what is proper to the mind is *constitutive* of the world.

At first sight, then, Kant's transcendental philosophy seems unfit for hitting the target of contemporary philosophy of mind. Kant actually says this: "[T]he chief question always remains: 'What and how much can understanding and reason cognize free of all experience?' and not: 'How is the faculty of thinking itself possible?'" (*KrV*, A xvii). What the mind itself could amount to *as an object of investigation* is indeed a question not readily associated with TI. The overall absence of TI from contemporary debates in philosophy of mind could indeed be due to the fact that the central aspects of Kant's philosophy are mainly concerned with the epistemological issue of object constitution. What is perhaps primarily at stake for Kantians is the objectivity of the empirical world, and the *a priori* laws of the mind that govern the former. Kant's focus lies with the object (and how it is constituted), not with the mind as a separate entity. However, it is often left unnoticed that in Kant's oeuvre the faculties of thinking come forward as requiring constitution as well. In the transcendental deduction of the *Critique of Pure Reason*, for instance, the constitution of the object seems to involve what Kant calls the 'modification' of the mind (*KrV*, A 97–99) and of 'inner sense' (*KrV*, A 367). In line with Jeff Malpas, I will argue that, for Kant, 'the mind' is as much in need of constitution as 'the world'.[3]

Furthermore, I will argue that because of this two-way model, whereby the mind is constitutive of the world and the world is constitutive of the mind, the Kantian mind is essentially a *basic mind* – that is to say, a mind explained in non-contentful and non-representational terms.[4] In Section 1, I explain, firstly, why Kant can be seen as committed to such a conception of the mind. Secondly, I conclude the section by

---

[3] Malpas 1999.

[4] This implies that a Kantian take on basic minds might be interesting for the (radical) enactivist movement in philosophy of mind, which aims, among other things, to conceive of minds without appealing to representation and content. The notion of the 'basic mind' was introduced by David Hutto and Erik Myin as follows: "[W]e propose, the nature of the mentality in question is not underwritten by processes involving the manipulation of contents, nor is it, in itself, inherently contentful. Basic minds do not represent conditions that the world might be in. To think otherwise, as many do, is to ascribe features and characteristics to basic minds that belong only to enculturated, scaffolded minds that are built atop them" (2013, p. ix). Apart from a rejection of content, I take it that the concept of the basic mind also involves the search for what is *sufficient* as well as *necessary* so as to identify what counts as a mind *überhaupt*, and nothing more. In this text, I use the term especially in the latter sense, as Hutto's and Myin's use of the term 'content' is far from on par with my Kantian understanding of the term.

I propose, however, that the notion of content across the many philosophical traditions, and especially regarding radical enactivism and TI, is worthy of investigation, especially since both aim to discard it somehow (cf. infra). But this is for another paper. One hurdle that such research would have to face concerns the question how Myin and Hutto's aim to give a *naturalised* account of the mind can be reconciled with Kant's *transcendental idealist* take on it. The latter is concerned with a transcendental investigation of how naturalist explanations of world phenomena are made possible to begin with. That is, Kant's TI aims to explain empirical phenomena in terms of principles that are *a priori* and pure, i.e., not empirical. Likewise, Kant aims to determine the conditions of possibility of the mind (cf. infra), rather than investigate into its causal – naturalised – constituents. However, if Kant has a contentless (i.e., 'basic') account of the mind, I do take this to be a result of his transcendental analysis of it.

showing how such a Kantian theory of mind amounts to the idea of mind extension, whereby the mind 'is there' by virtue of a world.

1.2 A Kantian Take on 'Minds Without Content'

In his review of Samuel Thomas von Sömmerring's monograph *On the Organ of the Soul* in 1796, Kant characterises the mind (*das Gemüt*) as the capacity for combination.[5] Here, Kant appears to brush up his earlier comment in the first *Critique* that "the synthesis of the manifold" occurs "through a common function of the mind for combining it in one representation" (*KrV*, A 109). In his commentary on Sömmering, however, Kant seems to stipulate that only the single qualification *that* the mind must entail some sort of combination counts. Of course, in a most rigid sense, the very concept of combination requires that *something* be combined. *What* this is, however, need not be determined *a priori*.

In that strict sense, Kant's account of the mind is a bit like a Fregean propositional function: $y = f(\,)$. According to Gottlob Frege, one must take seriously this notation: a propositional function merely prepares a *place* for an argument to be inscribed. That is, functions only *formally* anticipate possible *contents*. The question as to *what* specific arguments come to satisfy a function is irrelevant to the structure of the function itself. Yet the fact *that* functions must anticipate said arguments in the very structure of their notation, however, is of the essence.[6]

Something similar goes for Kant's theory of object constitution. In the *Critique of Pure Reason,* Kant develops a theory of object constitution and cognition in quite formal and general terms.[7] Nevertheless, Kant develops this theory so as to account for the constitution and cognition of objects insofar as they can be related to a specific content. He is not only interested in mapping the constitution of the object *in general.* What seems ultimately at stake in doing so is the constitution of empirical, singular objects – say, the ones you use every day.[8] His transcendental theory of the object is formal and mundane at the same time. Quite similarly, Frege's purification of the relation between values, functions, and arguments – $y = f(\,)$ – undoubtedly accounts for *satisfied* functions as well. That is, Frege is not only interested in the *structure* of functions, but he is also interested in how, for instance, they underlie specific arithmetical operations – e.g., 64 as the value of the function x^3 with 4 as the satisfying argument. But here it is crucial that, according to Frege, the essence of the function is to be understood as if 'not yet' enabling such specific function values. The same goes

---

[5] Eisler 2002, p. 182.

[6] Frege 2008, p. 2.

[7] In the 'Analytic of Concepts', for instance, Kant develops an account of the object whereby categories are *legislative* with regard to nature. To do so, Kant does not need to give an account of specific, contingent features of nature. He need only account for the possibility of nature in general, or what he calls *natura formaliter spectata* (*KrV*, 162–65).

[8] Especially in the 'Analytic of Principles', Kant gives examples of such singular empirical objects to substantiate his formal account of object constitution: he discusses houses (*KrV*, B 235), boats (*KrV*, B 237), and arithmetical operations (*KrV*, B 205), among other things. The message seems to be that mundane objects are formal affairs as much as objects in general.

for Kant's formal system of categories and forms of intuition: it is itself to be understood as if it were 'not yet' constitutive of singular, empirical objects.

In the *Critique of Pure Reason*, Kant specifies the mind's capacity for combination in terms of twelve categories, famously organised according to the four headings of quantity, quality, relation, and modality. These categories serve as the 'necessary conditions of possibility' for the constitution (and cognition) of empirical objects. They are deduced from what Kant holds to be the twelve basic 'forms of judgement' (which are also organised according to quantity, quality, relation, and modality). Kant is adamant, in this regard, that the twelve categories and their corresponding forms of judgement are all instances of 'the logical function'. Logical functions, according to Kant, concern "the unity of the act" (*die Einheit der Handlung*) enabling to bring "different representations under a common one" (*KrV*, B 93).[9]

These subtly coupled elements governing the constitution of the object – namely the logical function, the forms of judgement, and the categories – seem indebted to Kant's more generic account of the mind (*das Gemüt*) as the capacity for combination. As far as epistemology is concerned, this combining capacity of the mind comes forward as a logical and objectifying act (*Handlung*).[10] When related to representations, this act falls apart into twelve forms of judgement. And, says Kant, when these representations involve *intuitions* delivered by our faculty of sensibility, the twelve forms of judgement give way to the twelve pure categories of the understanding (*KrV*, B 104–05).

Although representations are in that sense at stake, their *specific content* remains, in fact, irrelevant.[11] When the mind concerns cognition through categories, content is

---

[9] In that sense, the Kantian function *might* be seen as the precursor of the Fregean one, though much can be said about their differences. It is still an open question whether the Fregean function – namely in terms of (*i*) a *function*, (*ii*) an *argument*, and (*iii*) a *value* – is in a way anticipated by the Kantian function, which is described in terms of (*i*) the unity of an act that combines (*ii*) different representations under (*iii*) a common one (*KrV*, B 93). Generally speaking, both seem to hold dear to a 'triadic' or 'trichotomic' structure. All the while, one should note that, for Frege, a function seems to be a figure of formal logic that is only *potentially* accompanied by epistemological and ontological implications, whereas it seems that, for Kant, a function is an element of formal logic that has epistemological implications *at the same time*. Kant's logical function and its twelve specific forms give rise, namely, to the categories of the understanding that count as the necessary conditions of possibility for the constitution and cognition of the object – see the metaphysical deduction of the categories in the *Critique of Pure Reason* (*KrV*, A 66–83/B 92–116).

[10] Terms like *Actus* and *Handlung* are crucial to Kant's theory of the mind (see Kaulbach 1978; Saugstad 2009). Kant is quite adamant that the Cartesian dictum of the mind – the *cogito* (the 'I think') – is to be called an act (*KrV*, B 137, 423). But what does this mean? In line with what I have argued already, it seems to imply that, for Kant, to have a mind is a *consequence* of *saying* 'I think' rather than the fundamental basis of it. To be able to say, write, or discursively judge that I think has an effect: it constitutes the mind, it creates the distinction between *res cogitans* and *res extensa*, between what can be *called* 'internal' and what can be *called* 'external'.

[11] In Kantian scholarship, the term 'representation' is a tricky one. The term itself alludes to something being re-presented, but this is a bit at odds with the original German term *Vorstellung*, which is perhaps better translated as 'presentation'. It seems to be a very generic term, signifying the occurrence *that* something *presents* itself to the mind (see *KrV*, B 377). So, it is not surprising that in Kant's philosophy, the term *Vorstellung* emerges in various contexts. In the context of cognition, a representation is objective when it involves a sensible intuition for which we can at the same time give a concept. In this case, that which *presented* – namely, the representation – does is presented in intuition while also being determined conceptually. But both a *mere* sensible intuition (without being determined

nothing but an explanandum *in general*. Kant's infamous 'Transcendental Deduction of the Categories' (*KrV*, A 95–130/B 116–69) does indeed precisely consist in abstracting from the content of representations in order to identify the latter's *formal* conditions of possibility. Or, as Jeff Malpas has it: "In so far as content is established through connection [i.e., the formal conditions of possibility], content surely cannot be that on the basis of which connection is established".[12]

So, if a philosophy of mind is extracted from Kant's critical oeuvre, and more precisely in relation to his theory of cognition, a basic mind can be encountered that is fundamentally non-contentful, precisely because content is somehow at stake. Therefore, a Kantian philosophy of mind would have to involve a basic mind insofar as it is 'not yet' a particular and contentful mind. It would have to involve a basic mind, namely, as if it were 'not yet' deeply engaged in representation and cognition. Or, for instance, with regard to the *Critique of the Power of Judgment* and the *Critique of Practical Reason*, Kant's philosophy involves a basic mind as if it were 'not yet' committed to aesthetic contemplation and wilful action in accordance with the categorical imperative.

However, I will argue that Kant's formal take on the mind not only involves *combination* and its different modalities, but *activity* as well. We will see that Kant's peculiar notion of the unity of the *act* (*Handlung*) is essential to grasping the formal and potentially elusive notion of combination. Reminiscent of a radical enactivist strand, we will see that, for Kant, combination *is* activity. In order to address this, I turn to Kant's infamous 'Refutation of Idealism' (*KrV*, B 274–80).

1.3 A Kantian Take on Extended Minds: The 'Refutation of Idealism'

Although a Kantian theory of the mind can be seen as a purely formal, non-contentful concern, this is not to say that the Kantian mind does not, in a way, *anticipate* content. Or, to put it with a Kantian idiom: content *in general* can (and must) be part of a formal take on the mind, but not content *itself*. At the end of the second chapter of the 'Analytic of Principles', Kant lucidly states the following: "All principles of the pure understanding are nothing further than a priori principles of the possibility of experience, and all synthetic a priori propositions are related to the latter alone, indeed their possibility itself rests entirely on this relation" (*KrV*, B 294). Indeed, the possibility of the mind's *a priori* principles of experience (that are systematically tied to the categories), which enable cognition of objects, rests entirely on the latter's relation to objects of experience itself.

In his 'Refutation of Idealism' (*KrV*, 274–80), Kant explains how this transcendental epistemological theory significantly problematises a Cartesian take on the mind. According to Kant, the Cartesian method of doubting the existence of the *res extensa* (extended thing) in the interest of isolating a fully functional *res cogitans* (thinking

---

conceptually) and a *mere* concept (without being related to an intuition) can be called representations (*KrV*, B 377).

[12] Malpas 1999, p. 4.

thing) is a kind of substantivist or material – as opposed to transcendental – idealism.[13] On Kant's reading, René Descartes allows for one empirical statement about the world only, namely the proposition that 'I am'. All other empirical statements are 'problematic' or dubitable from the viewpoint of absolute certainty. Therefore, all other empirical statements can be disconnected from the proposition 'I am', and more importantly, from the underlying analytical statement 'I think'. The crux of Descartes' material idealism is that the thinking mind can be an object of investigation without reference to the extended world outside of the subject, as the former's existence is indubitable whereas the latter's is not.[14]

Now, the crux of Kant's refutation of this Cartesian idealism consists in thoroughly relating these opposed terms – 'mind' and 'world', 'thinking' and 'extension', but also 'inner' and 'outer' – in such an intricate fashion that the one is in fact inconceivable without the other, thus turning any Cartesian 'meditation on the mind' into a pointless endeavour. Kant's argument reveals that the mind – insofar as it concerns 'inner sense' or 'inner experience' – is impossible without extension. It is worth taking a look at Kant's argumentation in its technicity.

Kant's line of reasoning begins with the premise that there is something like 'inner experience' – namely, that one is able to experience 'things' that seem entirely internal to the mind and are somehow disconnected from what can be ascribed to the environment. His second premise is that such inner experience is essentially *successive*; that the *form* of inner experience is, in other words, time: "I am conscious of my existence as determined in time" (*KrV*, B 275). He then goes on to the third premise, quite crucial for his argument, stating that "[a]ll time-determination presupposes something persistent in perception" (*KrV*, B 275). This means that there must be something persistent *in respect of which* change (variation) is possible, indicating that *without* something persistent, change is contradictory and impossible.[15]

Kant then applies this framework to the case of inner experience, which – as it is successive – is exactly such 'change'. Given the third premise, Kant soundly argues that inner experience must have some kind of persistent point as well. Kant then qualifies, quite unproblematically, that this persisting point of inner experience cannot be inner experience *itself*. After all, persistence and change do not coincide – this would be absurd. Crucially, then, the persisting point of inner experience *must lie elsewhere*. The persistence of the successive change we call *inner* experience lies, then, *outside* of us. In other words, inner experience is *made possible* by what is external.

---

[13] Descartes' substantivist account of the mind seems to entail the view that the mind is a distinct ontological entity.

[14] See the first two meditations of Descartes' *Méditations Métaphysiques* (2009, pp. 79–108).

[15] As suggested by an anonymous referee, someone might object that everything may very well be in a total state of variation, change, or fluctuation. If this were the case, a counterargument could be made that variation does not require a persistent point at all, as Kant maintains. However, if Kant's argument assumes that change is relative to a persistent point, then this is not incompatible with the option that the latter is *itself* subject to, or susceptible to, continual change. The inner change of the mind that Kant seeks to investigate can be relative to a persistent point that is ontologically *fictional* (as 'in reality' it would be in a state of continual fluctuation).

This allows Kant to refute Cartesian idealism: *from* the very assumption that there is something like inner experience, we must conclude that there are objects in space outside of us, i.e., that the possibility of assuming inner experience is tied to the possibility of assuming outer experience. Herewith, Kant maintains, Cartesian idealism is refuted.

Now, from this line of reasoning it might appear that Kant defends a content-based, deeply representationalist take on the mind. In line with his ever-formal approach, however, Kant immediately qualifies that the intertwinement of the internal mind with the extended environment, as elegantly put forward by his refutation of Cartesian idealism, is in no way limited to *actual*, *existing*, and thus *particular* objects. Kant is adamant, namely, that "[h]ere it had to be proved only that inner experience in general is possible only through outer experience in general" (*KrV*, B 278–79). Inner experience, otherwise subjected to the *a priori* categories of the understanding, is also predicated on the outer world in general, that is, on extension in general. More crucially, however, this means that the supposedly 'internal' categories of the mind that serve to constitute our inner experience stand in relation to extension as well. They are, in a sense, 'anticipatory' of extension.[16]

This formal, transcendental argument leads to the following, striking conclusion: insofar as basic minds involve inner experience, they must, on Kantian grounds, at the same time involve extension. Herewith, Kant subtly yet fundamentally destabilises any Cartesian or substantivist account of the distinction between 'mind' and 'world', 'thinking' and 'extension', 'inner' and 'outer'. The epistemological consequences of this destabilisation are not to be disregarded, according to Kant. In the 'Fourth Paralogism', he writes: "I am no more necessitated to draw inferences in respect of the reality of external objects than I am in regard to the reality of the objects of my inner sense (my thoughts)" (*KrV*, A 371).

For my purposes, however, it is crucial to see that the mind in its 'internal' thinking activities, described above in terms of categories and logical functions, is necessarily predicated on extension. Accordingly, to presume that the Kantian categories – i.e., the highly formal modalities of cognition – are 'internal' is perhaps a bit redundant. There is, in any case, no *need* to presume it.[17]

In the 'Transcendental Deduction' Kant seems to suggest, seemingly in this spirit, that the intricate relation holding between extension and the formal modalities of thinking with regard to inner sense must be conceived of as *epigenetically* structured. By shortly

---

[16] Although categories anticipate extension, they can only do so formally, as they could not appeal to the specific content of extension. The latter is due to their following the *a priori* and *formal* rules of the faculty of the understanding. Therefore, categories can only anticipate extension (and content) *in general*.

[17] As aptly noted by an anonymous reviewer, it is perhaps a bit redundant as well to talk about 'something outside of us' without already presuming a mind or an inner experience proper to a persisting subject. This is, of course, the rather 'textbook' interpretation of Kant's TI. In his 'Refutation of Idealism', however, Kant subtly argues that what is internal is as much dependent on what is external as the other way around, which is indeed seemingly contrary to the more common way of approaching TI. Considering his closeness in time and spirit to Descartes's dualism, Kant seems to struggle with an often elusive, yet potentially elegant dialectic between what is inner and outer, what is mind and what is world.

analysing the meaning of this 'epigenesis' in the next section, I suggest that Kant's take on the mind must concern activity, and seemingly in the *spatial* sense at that, since extension is involved.

1.4 Kantian Minds as Activities

In order to situate his epistemological theory of object constitution in the wider philosophical landscape, Kant often made use of one very particular and compelling analogy. In his *Inaugural Dissertation* (and on several other occasions), Kant maintains that the categories are either derived from experience, or innate, or *a priori* acquired (*MSI*, AA 02: 395). Kant aligns these three options with the three most dominant eighteenth-century approaches to embryology: physical influx, preformation, and epigenesis. Kant argues that the categories involve *a priori* acquisition, so he sides with epigenesis and rules out the options of physical influx and preformation.[18] So, not unlike the early modern embryological theory of epigenesis, Kant's transcendental version of epigenesis concerns a procedure, the various elements of which cannot in themselves sufficiently explain the presupposed result, that is to say, the constitution of the object. Whereas embryological epigenesis presupposes (*i*) a material predisposition in combination with (*ii*) a specific environment so as to explain (*iii*) the emergence of a mature organism, Kant's analogy poses (*a*) the combinatory capacity for discursive thought (formative of the forms of judgement) proper to the mind's faculty of the understanding[19] in combination with (*b*) the capacity for sensory affection proper to the faculty of sensibility in order to explain (*c*) the constitution of the object.[20] It is, namely, only under the pressing invitation – "sous la sollicitation", as Herman de Vleeschauwer has it – of sensory affections, that forms of judgement can be invoked to produce categories that stand in relation to experience.[21] Quite importantly, this means that according to Kant there is no pre-formed categorial system 'waiting to be applied'.

In that regard, Béatrice Longuenesse argues that the heterogeneous relation between discursivity and sensibility does not indicate a mere conjunction but rather amounts to a relation of dependence between sensibility and discursivity: "[B]ecause intuitions rest on affections or depend on receptivity, concepts [categories] have to rest on functions".[22] Indeed, insofar as intuitions are in themselves blind (*KrV*, B 75) and purely singular, the human subject *must* make an appeal to its general discursive capacities in order to constitute these intuitions as empirical objects. Therefore, the Kantian mind is not 'filled' with categories, waiting to be applied to intuitions – it is, again, not preformed. Instead, the mind's confrontation with its own faculty of

[18] See Lu-Adler 2018 for an interesting account of Kant's grounds for aligning TI with epigenesis instead of with physical influx (*generatio aequivoca*) or preformation.

[19] Discursive thought or discursivity entails the combination of concepts through judgements (cf. Longuenesse 1998, p. 6) according to general rules. Sensibility, on the other hand, is non-conceptual and therefore also non-discursive.

[20] Malabou 2016, pp. 21–22.

[21] De Vleeschauwer 1937, p. 270.

[22] Longuenesse 2005, p. 93.

sensibility highlights the need for something that is completely different (heterogeneous) from that faculty, namely a capacity for thinking, for judging, for acquiring categories – that is, the faculty of the understanding.[23] However, the latter faculty is in turn also in need of sensible intuitions if it wants to have a relation with objects. Kant is very clear that intuitions and concepts "therefore constitute the elements of all our cognition, so that *neither concepts without intuition* corresponding to them in some way *nor intuition without concepts can yield a cognition*" (*KrV*, B 74, my italics). In light of that, Kant famously states that "[w]ithout sensibility no object would be given to us, and without understanding none would be thought" (*KrV*, B 75). And even more importantly: "Thoughts without content are empty, intuitions without concepts are blind" (*KrV*, B 75).

Now, that object constitution is in that sense organised epigenetically (i.e., neither concepts nor intuitions can account for cognition alone, yet both are necessary) informs the following: despite the fact that the categorial system necessarily amounts to a general, formal discursivity – otherwise it could not make a difference with regard to sensibility – this is not to say that the occasion for this system to be invoked and developed (to be *a priori* acquired) does not presuppose the singular position of a sensory, embodied subject. The Kantian mind of discursive categories is, therefore, at the same a mind that moves through a world of sensibility, that attempts to orient itself amidst a manifold of intuitions. The highly formal and discursive categories are fundamentally *distinct* from our sensible intuitions, but they are not *detached* from them. This is why the determination of intuitions on account of categories involves a 'modification' of the mind (*KrV*, A 97–99).

In other words, *if* the Kantian mind is above all to be called a *thinking thing*, then it is at the same time to be called an *extended thing*. The discursive features of the Kantian mind are indeed inseparable from the latter's active, embodied, moving – that is to say *sensible* – aspects. So, if Kant puts us on the track of assuming that the mind is a combining capacity and, in a second move, that this combining capacity is to be called a *Handlung* (e.g., *KrV*, B 93) or *Actus* (e.g., *KrV*, B 137, 423), then we could, in fact, take this literally. The mind, it seems, *is* an act. One particular comment of Kant's is highly revealing in this regard: "We cannot think of a line without drawing it in thought" (*KrV*, B 145). And the same goes for circles, triangles, and so forth. This subtly rich statement, easily overlooked as a trivial remark, is a radical one indeed. The fact that highly abstract 'concepts' like geometrical lines cannot even be *thought* without reference to the extended activity of *drawing* them, is of the essence for Kant. The radicality of Kant's statement lies in the assumption that to *merely think* of a line requires movement and extension as much as drawing or seeing an 'actual' one. From this, one might conclude that Kant attempted, perhaps even unknowingly and

[23] In this regard, one could perhaps wonder, as thankfully suggested by an anonymous reviewer, whether Kant's philosophy of mind presupposes *minds to have minds*. This suggestion could give rise to an interesting avenue of research, although I am much more inclined to speak, with regard to Kant, of *a singular yet divided mind*. The Kantian mind, so it seems, can be set out not only in terms of the different faculties and capacities that seem to give life to it, but also (and perhaps more importantly) in terms of the heterogeneity between these very faculties and capacities, which highlights the internal division of a singular mind rather than a plurality of minds.

undoubtedly still preliminarily, to suggest that the workings of the mind involve extension as much as the activities of bodies. His project seems to be a subtle and largely unnoticed attempt at interweaving what standardly pertains to the mind (thinking, categories, combination) and what standardly pertains to extension (spatial movements, acts, activities).[24] It reveals, in any case, a more fundamental struggle to think through his quite upfront refutation of the sharp Cartesian distinction between mind and world.

With this in mind, I now turn to Andy Clark and David Chalmers' seemingly anti-Cartesian defense of mind extension, arguing that it must not be seen as a continuation of Kant's subtle attempts at showing mind's extension, but rather as a philosophical setback for this emerging project.

2. Andy Clark and David Chalmers

2.1 The Extended Mind Thesis' Cartesian Heritage

For a long time, philosophies of mind and cognition have been committed to the kind of 'internalism' already problematised by Kant, most famously in his 'Refutation of Idealism'.[25] In that sense, the field is a dominantly Cartesian one. When it comes down to it, the mind is still a *res cogitans*. In recent years, however, this Cartesian heritage has been increasingly considered a thorn in the flesh. A certain reflective dissatisfaction reigns the field nowadays. And perhaps the road to philosophical progress must indeed be anti-Cartesian, whatever the implications. A new, embodied, enactive, or extended take on the mind, secretly promising to yield a closing of the gap between *res cogitans* and *res extensa*, is at the horizon.[26]

The proposal of the extended mind thesis (EMT), as put forward by Clark and Chalmers in 1998, can attest to this tendency. Their observation that the mind's (specifically computational or more generally procedural) functions can sometimes be

---

[24] Kant does something similar in *Was Heißt: Sich im Denken Orientiren?* (1786), where he attempts to show that the human subject should not only orient itself in the world, but in thinking as well (*WDO*, AA 08: 136).

[25] In general, I construe internalist theories of mind as assuming that, for having a mind, intrinsic capacities are sufficient. Descartes' first two meditations can be read as arguments in favour of internalism. Externalist theories, on the other hand, I construe as assuming that some extrinsic features of the environment are required to have something like intrinsic capacities of the mind to begin with. Thus construed, internalism is a very strict point of view, while externalism is a more open alternative. The former excludes extrinsic features of the environment (in the sense that they are not required to explain intrinsic capacities), while the latter includes intrinsic capacities (in that it tries to give an explanation for these intrinsic capacities). Recently, however, externalist theories of mind are increasingly becoming exclusive. Hutto and Myin's radical enactive approach to mind seems to argue, for instance, that to have a mind, intrinsic capacities are not required *at all*. The mind, then, is a *completely* external (or extensive) capacity (2013, pp. 142–47).

[26] A notable attempt at getting rid of the Cartesian heritage in philosophy mind, beside the one by Clark and Chalmers, concerns the enactivist movement largely due to work by Francisco Varela, Eleanor Rosch, and Evan Thompson, especially their *The Embodied Mind* in 1991, which has proven to be seminal. More recently, David Hutto and Erik Myin have developed a radical version of enactivism, with publications like *Radicalizing Enactivism: Basic Minds Without Content* in 2013 and *Evolving Enactivism: Basic Minds Meet Content* in 2017.

taken over by mechanisms, features, means, and objects in the environment seems to break with the exclusively internalist take on mind and cognition. Clark and Chalmers' overarching idea seems to be that there are no good reasons for assuming that the mind's computational, intellectual, and contentful procedures must *always* be internal. However, does this mean that EMT sets in motion a wholly new science of the mind, radically doing away with any Cartesian take, in continuation of Kant's radical yet straggled critique? I have reservations.

The aim of this section, then, is to show that EMT still complies with a dualist or Cartesian take on the mind. Sure, the proclaimed philosophical conclusions and scientific consequences of EMT run counter what would be Descartes' own – certainly anachronistic – philosophy of mind. Indeed, in doubting everything he knew so as to lay bare the single point of absolute certainty, Descartes' meditations give rise to a conception of the mind that radically excludes any extension whatsoever. Even more so did his methodology – put forward in the interest of the epistemological and ontological isolation of what pertains to the thinking mind itself (*res cogitans*) – precisely consist in the theoretical elimination of extension (*res extensa*). Superficially speaking, then, EMT seems to be the Cartesian's rightful opponent. The project of Clark and Chalmers is indeed seemingly set to theoretically incorporate thinking, cognition, and mind into extension, allowing them to defuse the alleged historical effect of Descartes' first two meditations. Their take on language as a form of mind extension, for example, is indicative of such an objective:

> Without language, we might be much more akin to discrete Cartesian 'inner' minds, in which high-level cognition relies largely on internal resources. But the advent of language has allowed us to spread this burden into the world. Language, thus construed, is not a mirror of our inner states but a complement to them. It serves as a tool whose role is to extend cognition in ways that on-board devices cannot.[27]

What is really at stake here is the question as to what should and should not pertain to a full-fledged account of a basic mind. That is to say, insofar as we are concerned with the mind's *basic* features, what is it that we are dealing with? Philosophy of mind seems to be driven, above all, by the attempt to stipulate both what is *necessary* and what is *sufficient* in order to account for this mysterious thing called 'the mind'. Let us, then, rephrase the issue as follows: for a Cartesian, the extended environment would be qualified as *neither necessary, nor sufficient* to do the job. Now, a revolutionary thinker, hopeful for a new paradigm, who engages with Clark and Chalmers a bit superficially would perhaps be inclined to read the exact opposite of such Cartesianism into their proposal of mind extension. They do indeed purport to show that "extended cognition is a core cognitive process, not an add-on extra".[28]

Now, does the extended environment play a *sufficient* role so as to account for what is a mind according to EMT? Or is the environment put forward as playing a *necessary*

---

[27] Clark & Chalmers 1998, p. 18.
[28] Ibid., p. 12.

role? Generally speaking, I would judge that, according to EMT, extension is perhaps sufficient, but certainly not necessary. This is important: on their terms, whether the mind is extended is a matter of case-to-case comparison. As we will see, their heuristic is role- or function-based. The difference with TI is already quite obvious here. According to TI, it should be easy to do away with such a crooked structure. On Kantian terms, namely, there is mind if there is extension, and vice versa. But if the mind should only *sometimes* be seen as extended into the world, what follows is that at certain other times the mind might not be extended at all. Let me illustrate this by taking a closer look at the precise role of the overarching example that guides their argumentation: the cases of Inga and Otto.

Inga wants to go to the Museum of Modern Art (MoMA). On her way to the exhibition she is interested in, she remembers that the museum is located on 53rd Street. Here, the non-occurrent belief that MoMA is on 53rd Street "was somewhere in memory, waiting to be accessed".[29] Otto, who suffers from Alzheimer's disease, also hearing about the exhibition at MoMA, decides to go as well. Due to his disease, he must consult a notebook – in which he writes down all sorts of information – in order to retrieve the address of the museum. As Clark and Chalmers have it, "it seems reasonable to say that Otto believed the museum was on 53rd Street even before consulting his notebook", because "in relevant respects the cases are entirely analogous: the notebook plays for Otto the same role that memory plays for Inga". In Otto's case, "it just happens that this information lies beyond the skin".[30] This compelling example quite lucidly brings Clark and Chalmers to conclude that in Otto's case, the mind, insofar as it concerns belief, must be seen as extended into the world. But what about Inga? As for her retrieval of information from within her own memory, Clark and Chalmers give the impression that it would be pointless to argue for extension in her case. They contend, namely, that in Otto's case, "it just happens that this information lies beyond the skin",[31] solely giving arguments for the claim that in *his* case what is mindful must extend into the world.

But perhaps Inga's mindful activities can be seen as extended by considering what Clark and Chalmers say about the role of *language*. Language, they say, is not a mirror of our inner states but a tool that allows for extending high-level cognition into the world. Where the endeavours of the inner mind fail (or are plainly less useful), language comes to the rescue, extending the processes of the mind into the world. Language is, in their words, a *complement* to our inner resources.[32] But this does not solve the problem: precisely because language is put forward here as a *complement* to our inner resources, the mind is herewith only *partially* extended. On these terms, some of my internal activities might not involve language at all. Some of my internal activities, then, might very well be fully independent of the *occasional* extensions of

[29] Ibid.
[30] Ibid., p. 13.
[31] Ibid.
[32] Ibid., p. 18.

my mind. In that sense, Clark and Chalmers fail to present extension (e.g., pertaining to language, or other tools like notebooks) as *necessarily* constitutive of the mind.[33]

Clark and Chalmers seem to be operating with a kind of 'leftover' internalist conception of the mind. Therefore, I take EMT *not* to be the antipode of a Cartesian philosophy of mind. The internalist leftover (cf. Inga, or 'the mind insofar as it does not involve language') is treated, moreover, as in no way theoretically *influenced* or *affected* by the proposed cases of mind extension (cf. Otto, or 'the mind insofar as it *does* involve language). Their comparison of Inga with Otto thus allows for the sharp delineation between *res cogitans* and *res extensa* to be maintained in philosophy of mind, although sometimes the latter is seen to portray the same cognitive role as the former.

This is crucial: what counts for Clark and Chalmers with regard to conceptualising minds are cognitive *roles* or *functions* (not in the Fregean sense). And sometimes a specific role that is otherwise organised internally (that is to say, within the confinements of the skull), like memory, can be taken over by a specific feature of the extended environment – your phone, for instance, or a notebook. In some instances, something extended like a phone is nothing but mindful. To develop this thought, which is indeed quite revolutionary and not without scientific effect, Clark and Chalmers need not bypass the substantivist distinction between inner and outer, between internal processes and environments, *between mind and extension*. If their thinking is radical, it is perhaps only so in the sense that what is radicalised is the role- or function-oriented philosophy of mind they admittedly adhere to.[34]

I am not hesitant to say that their thesis of mind extension promises interesting insights. It has, in fact, generated undeniably stimulating discussion in the field. Enactivism, for instance, seems to find inspiration in EMT. The broad enactivist idea that the environment in which the subject is situated is vital for enacting what is mindful and what counts as cognition owes much to their suggestion of occasional extension. The radical enactivist wish to conceive of the mind not only as extended, but as intrinsically *extensive*, also clearly originates in Clark and Chalmers' proposal.[35] As opposed to enactivism, however, the question as to what is proper to the 'basic mind' is left quite implicit by Clark and Chalmers. On many levels, EMT is on par with the age-old Cartesian take on minds, be it hidden under a role- and function-oriented methodology.

In that regard, the fact that extension is only defended as an *occasional* feature of the mind cannot be stressed enough. Seemingly, said occasionality is much more symptomatic than it is methodological. By *occasionally* incorporating the mind into extension, EMT testifies to the stubborn attempt to extend minds which are – or can

---

[33] Mark Rowlands develops a similar argument, noting quite correctly that "the extended mind is perfectly compatible with the existence of a brain in a vat" (2009, p. 631).

[34] In this regard, it seems that EMT shows solidarity with computationalism. In trying to get a scientific hold on the human mind, computationalists find inspiration in Turing machines. According to them, the mind is a machine that 'computes' – governed by specific *functional* rules and scripts, the mind manipulates symbols that count as inputs so as to generate certain outputs.

[35] Hutto & Myin 2013, pp. xviii–xix.

easily be seen as – essentially non-extended. Take a look at the central question Clark and Chalmers aim to tackle: "Where does the mind stop and the rest of the world begin?".[36] Such an objective holds dear to an assumption that can easily be ascribed to the Cartesian project, namely that 'mind' (*res cogitans*) and 'world' (*res extensa*) are to be seen as *a priori* given terms, representing two classes of objects that must be accurately described and distinguished before their exchange of roles can even be conceived. EMT insinuates, in that regard, that the mind is in fact comprehensible in separation of extension, although for the sake of completeness both terms must at times be seen as intertwined or as exchanging roles and functions.

In Kant's work, to the contrary, a take on the mind was found that is much more revolutionary *vis-à-vis* such Cartesianism, as well as much more radical *vis-à-vis* the issue of mind extension. On Kantian grounds, and I refer to epigenesis here, terms such as 'mind' and 'world' could be seen as *resulting* from the activities of a subject, rather than as the *a priori* given elements of said activities.

2.2 A Note on Action, Thought, and Computation

According to Clark and Chalmers, the exemplars of mind extension they put forward play an *active* role. By taking away, for instance, Otto's notebook, one takes away a whole cognitive process, because the "external features here are just as causally relevant as typical internal features of the brain".[37] Although this externalist take on the mind is still compliant with a Cartesian, internalist, and computationalist one, as I have argued, Clark and Chalmers also portray an attempt to move things to a slightly more radical philosophical point. With regard to a Scrabble game, for example, which involves the physical rearrangement of tiles, they ask themselves whether an internal computationalist account in terms of inputs and actions is really that fruitful. Eventually, they come to the conclusion, somehow reminiscent of Kant's example of the necessity of *drawing a line in thought*, that in "a very real sense, the re-arrangement of tiles on the tray is not part of action; it is part of *thought*".[38] Seemingly, what is 'action' (and would accordingly be extended) is in fact 'thought' (and would accordingly be mindful).

But is there really opposition here? My Kantian suggestion that thoughts *are* actions, that the mind *is* an activity, seems to break with this opposition in a significant way. It allows for the conclusion that thought and extension are different sides of the same coin, and by no means contradictory predicates of the term 'mind'. To say, as Clark and Chalmers do, that physical interventions in the world (like rearranging tiles on a board) might not be actions but thoughts, is one thing. To say that thoughts *are* actions, as I do on Kantian terms, is quite different.[39]

---

[36] Clark & Chalmers 1998, p. 7.
[37] Ibid., p. 9.
[38] Ibid., p. 10.
[39] As rightfully stipulated by an anonymous referee, it should be noted that in philosophy of mind and cognitive science, scholars are already preoccupied with similar problems. The issue of 'mental

That the Kantian mind essentially involves activity is, as I have argued, tied to the epigenetic foundation of its discursive capacities. These discursive capacities must involve formal rules and categories, according to Kant, so as to make a difference with regard to sensibility. These categories are not to be found in our heads, waiting to exhibit certain rules with regard to an information-carrying environment. To the contrary: precisely because the environment is *not* information-carrying, unification and combination are requested activities. Hereby, a formal system of categories is installed, or perhaps more accurately, revealed to have been 'operative' or active all along.[40]

In this regard, and perhaps a bit surprisingly for some readers familiar with the history of philosophy, a Kantian philosophy of mind diverges significantly from computationalism, although both do indeed hold on to minds as consisting of formal rules and procedures. For computationalism, however, said formal rules and procedures are more like scenarios with presupposed outputs as end results, seemingly waiting to be applied to a world made up of information-carrying inputs. From a transcendental idealist perspective, then, computation is an undesired variety of formalism of the mind. Firstly, computation neglects the constitution of the object. Objects, namely, are given to the mind as information-carrying inputs, suitable for internal (or external, in the case of EMT) processing according to formal rules. Second, these same formal rules are grounded from *outside of the system*. A computational system's formal rules, namely, appear to be 'put in' (say, by a human). In Kant's terminology, computation is like preformation: the formality of its systematicity is

---

action' especially comes to mind. As Thomas Metzinger has it, it must be recognised that "[m]ental actions are a large and relevant subset of the domain of mental events" (2017, p. 1). Particularly striking for my purposes is the manner in which mental action has been an issue for predictive processing (PP), the computationalist theory that is perhaps closest to Kant (see Swanson 2016 for a defense of this claim). PP presupposes formal systems to *predict* (or, more generally, *anticipate*) the kinds of input the system will need to process. That is, mental systems are not merely receptive, but *spontaneous* as well – they are spontaneous in their reception. Yet, it remains an open question whether 'mental action' is in any theoretical way compatible with the Kantian idea preliminarily hinted at and developed in this paper, that the mind *is* an activity. At first glance, conceptual points of convergence are hard to find: mental action involves, for example, the attempt to retrieve images from memory (Metzinger 2017, p. 1), or just focusing on some mental tasks like arithmetical operations. Therefore, "[i]n mental action there is no motor output to be controlled and no sensory input vector that could be manipulated by bodily movement" (Metzinger 2017, p. 1). Thus, it is hard to relate the idea of mental action to the one that assumes that the mind *is* an activity, as the former is seemingly compatible with the Cartesian assumption that there is a realm of extension (involving motor and bodily activity) on the one hand and a realm of mentality on the other hand, and that both are ontologically independent (see Levy 2019 for a discussion of this problem).

Perhaps, then, the idea of the mind as an activity is closer to mental *behaviour* than to mental *action*. The latter involves consciousness and volition, while the former is devoid of conscious goal-representation but can still be seen as cognitively and epistemologically purposeful (Metzinger 2017, p. 3). It seems that, not unlike Kant's system of categories and functions of judgement, mental behaviour is much more *formal* than mental action. I think that Metzinger is on the right track when he argues that unconscious, non-volitional or unintentional mental behaviour helps *constitute* mental action (2017, p. 3). Yet the idea of mental behaviour does not necessarily show theoretical kinship with the idea that the mind is an activity *either*. So, a lot of conceptual work lies ahead of me if I want to bring this quite Kantian idea of 'mind as activity' to contemporary debates.

[40] In order to explain why there are objects, we must *presuppose* certain capacities to be operative or in action. Epigenesis, then, involves the *retrospective* deduction of certain capacities (and the fact that objectivity is subject to them).

devoid of any necessity.[41] Now, in comparison to my Kantian take on mind extension, Clark and Chalmers' thesis has a hard time detaching itself not only from Cartesianism, but also from this inherited 'burden' of computationalism. Unlike my Kantian proposal, their view of extension is loaded with content and information. And the same goes for their view of formality.

3. Concluding Remarks

I developed an analysis of Clark and Chalmers' seminal extended mind thesis from the viewpoint of what would be a Kantian take on mind extension. My analysis has been conducted on two levels. Firstly, and most importantly, I mapped both theories' respective relations to the Cartesian distinction between mind and world. Whereas Kant clearly works with the Cartesian paradigm yet thoroughly exposes its blind spots from a transcendental idealist perspective, EMT is superficially far from but fundamentally close to it. If there is a Kantian version of mind extension, then it applies *a priori* to all minds equally. Such a Kantian perspective turns down Clark and Chalmers' case for the occasional instances of mind extension and the examples they give to substantiate it. For Kant, I argue, there would be no fundamental difference between Otto and Inga.

Secondly, as an aside, I highlighted that Clark and Chalmers, although they simultaneously criticise it, operate within the realm of computation. The latter involves a certain formality of the mind that is predicated on the *processing* of objects as information-carrying inputs. Kant, however, takes the mind's formality to be *constitutive* of the object – of information, content, and representation. So, if Clark and Chalmers' influential proposal has given way to the increasingly supported disavowal of a formalism of the mind (as 'intellectualism'), paving the way for naturalisation and embodied cognition,[42] then this does not rule out a *Kantian* formalism of the mind, which is centred around constitution (and must not be construed as 'intellectualist'). Leaving behind computation's take on formality might be a rightful course of direction, but it does not justify leaving behind formality *altogether*.

[41] But see also Henk Vandaele's interesting (Kantian, but also Fichtean) analysis of computation from within a transcendental idealist viewpoint (2010).

[42] See Hutto & Myin 2013, 2017.

References

Clark, A., and Chalmers, D. (1998). 'The Extended Mind.' *Analysis, 58*(1), pp. 7–19. https://www.jstor.org/stable/3328150.

de Vleeschauwer, H. (1937). *La Déduction Transcendantale de 1787 Jusqu'à l'Opus Postumum* (Volume 3). Antwerpen: De Sikkel.

Descartes, R. (2009). *Méditations Métaphysiques*, edited by M.-F. Pellegrin. Paris: GF Flammarion.

Eisler, R. (2002). *Kant-Lexicon*. Zürich: Olms.

Frege, G. (2008). *Funktion, Begriff, Bedeutung: Fünf logische Studien*, edited by G. Patzig. Göttingen: Vandenhoeck und Ruprecht.

Hutto, D., and Myin, E. (2013). *Radicalizing Enactivism: Basic Minds Without Content*. Cambridge: MIT Press.

Hutto, D., and Myin, E. (2017). *Evolving Enactivism: Basic Minds Meet Content*. Cambridge: MIT Press.

Kant, I. (1900). *Gesammelte Schriften* (Volumes 1–23). Berlin: Deutsche Akademie der Wissenschaften.

Kant, I. (1900). *Kritik der Reinen Vernunft*, edited by B. Erdmann. Berlin: De Gruyter.

Kant, I. (1998). *Critique of Pure Reason*, edited by P. Guyer and A. W. Wood. Cambridge: Cambridge University Press.

Kaulbach, F. (1978). *Das Prinzip Handlung in der Philosophie Kants*. Berlin: De Gruyter.

Longuenesse, B. (1998). *Kant and the Capacity to Judge: Sensibility and Discursivity in the Transcendental Analytic of the Critique of Pure Reason*. New Jersey: Princeton University Press.

Longuenesse, B. (2005). *Kant on the Human Standpoint*. Cambridge: Cambridge University Press.

Levy, Y. (2019). 'What is 'Mental Action'?' *Philosophical Psychology, 32*(6), pp. 969–91. https://doi.org/10.1080/09515089.2019.1632427.

Lu-Adler, H. (2018). 'Epigenesis of Pure Reason and the Source of Pure Cognitions: How Kant is No Nativist about Logical Cognition.' In *Rethinking Kant* (Volume 5), edited by P. Muchnik and O. Thorndike, pp. 35–70. Newcastle upon Tyne: Cambridge Scholars Publishing.

Malabou, C. (2016). *Before Tomorrow: Epigenesis and Rationality*. New Jersey: John Wiley & Sons.

Malpas, J. (1999). 'Constituting the Mind: Kant, Davidson and the Unity of Consciousness.' *International Journal of Philosophical Studies, 7*(1), pp. 1–30. https://doi.org/10.1080/096725599341947.

Metzinger, T. (2017). 'The Problem of Mental Action: Predictive Control without Sensory Sheets.' In *Philosophy and Predictive Processing*, edited by T. Metzinger and W. Wiese, pp. 1–26. Frankfurt am Main: MIND Group.

Rowlands, M. (2009). 'The Extended Mind.' *Zygon, 44*(3), pp. 628–41. https://doi.org/10.1111/j.1467-9744.2009.01021.x.

Saugstad, J. (2009). 'Kant on Action and Knowledge.' *Kant-Studien, 83*(4), pp. 381–98. https://doi.org/10.1515/kant.1992.83.4.381.

Swanson, L. R. (2016). 'The Predictive Processing Paradigm Has Roots in Kant.' *Frontiers in Systems Neuroscience, 10*, pp. 1–13. https://doi.org/10.3389/fnsys.2016.00079.

Vandaele, H. (2010). 'Why Computers can not Anticipate the Sufficient, but the Necessary Condition of What Will be.' *Casys: International Journal of Computing Anticipatory Systems, 23*, pp. 52–64. http://hdl.handle.net/1854/LU-2000590.

Varela, F. J., Thompson, E., and Rosch, E. (2016). *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge: MIT Press.

Undergraduate Philosophy Journal of Australasia www.upja.online
VOL. 2, NO. 2, DECEMBER 2020

The Nature of Pleasure in Plato's Philebus

RUBY HORNSBY[*]

UNIVERSITY OF LEEDS

Abstract

The central question in *Philebus* concerns whether the life of pleasure or the life or reason is most akin to the good human life. Naturally, engagement in such discussion requires an adequate analysis of the natures of pleasure, rationality, and the good. It is the purpose of this paper to outline and defend a (non-exhaustive) two-fold account of pleasure as presented in the dialogue. Specifically, the paper will argue for the claim that Plato advocates an account of pleasure as a process of change that occurs in sentient beings either when the harmonious natural condition is genuinely or apparently restored (impure pleasure), or when certain potentials are actualised by the rational human (pure pleasure).

Philebus is widely conceded to be an important work, though relative to the vast and comprehensive commentaries of the other dialogues, it has in large part been victim to systematic neglect.[1] Motivated by the desire to draw contemporary attention to the sagacious and illuminating themes in *Philebus*,[2] I attempt to contribute towards filling a gap in the field by elucidating the intricate and notoriously complex account of pleasure in the dialogue. I do this by providing a novel and comprehensive two-fold (non-exhaustive) analysis of pleasure. It will be argued that, on Plato's account in *Philebus*, pleasure is a process of change that occurs in sentient beings when either:

(1) The harmonious natural condition is genuinely or apparently restored (impure pleasure), or when

(2) Specific potentials are actualised by the rational human (pure pleasure).

Section 1 is a preliminary discussion to this paper and will elucidate Plato's 'fourfold division of being' – an ontological account presented in *Philebus* which places "everything that actually exists now" into four kinds.[3] This is because Plato's

---

[*] *Ruby Hornsby is pursuing her Master's degree in philosophy at the University of Leeds, England. Her research interests predominantly lie in ancient philosophy, with recent work focused on Platonic conceptions of pleasure, friendship, and the extent to which they are welcome in the good human life.*

[1] Davidson 1990, p. 2.

[2] A desire I share with Fletcher 2017, pp. 195–206, 2018, p. 30; Gill 2019, p. 75; and Tuozzo 2018, pp. 325–29, among others.

[3] 23b–27d.

ontological thesis plays a fundamental role in elucidating and defending several key claims made throughout this paper; hence, clarifying this account from the outset will contribute towards making the overarching argument of this paper lucid. Section 2 will defend the claim that *impure* pleasures are unlimited in kind and subsume under the restoration model: an account which argues that pleasure arises when "harmony is regained, and its former nature restored".[4] Contrastingly, Section 3 will argue that *pure* pleasures conform to the following identity criteria: (a) they are preceded by *unperceived* lacks, (b) they have objects that are true, beautiful, and measured (such that they belong to the mixed ontological kind), and (c) they are sufficient; their telos is *internal*. Hence, the paper will conclude that there are at least two varieties of pleasure in Plato's *Philebus*.

1. Fourfold Division of Being

The fourfold division of being divides "everything that actually exists now" into four kinds:[5]

> Socrates: As the first I count the unlimited, limit as the second, afterwards in third place comes the being which is mixed and generated out of those two. And no mistake if the cause of this mixture and generation is counted as number four.[6]

The unlimited kind can be classified as that which is relative and thus contains the "More and Less" (i.e., it supervenes on an indefinite matter of degree).[7] This class is characterised by being ontologically scalar and includes implicit comparative 'opposites' identifiable semantically through gradable adjectives: hotter/colder, strongly/gently, or rather, as Neil Cooper terms, "being R-er than".[8] These terms are not quantitative (exact) for they are infinite, though they do differ in degree. Accordingly, the unlimited class is "always in a state of flux and never remains".[9] Contrastingly, those which are both definite and non-scalar, e.g., 'equal to X', 'double Y', or the number 40 itself, belong to the limited kind, identifiable in virtue of their inherent exactness. The third kind is a *mixture* of the limited *and* unlimited and 'comes-to-be' when a definite point is imposed onto an infinite scale. For example, 40°C belongs to the mixed class since 40°C is a definite point on the unlimited scale of temperature.[10] If the *right* definite limit is imposed on the mixed class, it has the capacity to take away "excesses and unlimitedness, and establish harmony and moderation" by inflicting boundaries in that domain. For instance, when the right limits (e.g., the right pitch, tempo, timbre, etc.) are imposed into the unlimited domain of music, melody (as a member of the mixed class) is generated.[11] It is noteworthy that

---

[4] 31da3–8.
[5] 23c3.
[6] 27b–c.
[7] 24a–e.
[8] Cooper 1968, p. 12.
[9] 24d3, 24d1–25a4.
[10] Cooper 1968, p.13.
[11] 26a, 25e; Gosling & Taylor 1982, p.132.

everything Socrates places into this division is likewise something classified as good.[12] Finally, the fourth kind can be described as the cause of the third kind, the *maker* of that which comes-into-being.[13] To extend the previous example, the composer who generates melody by imposing limit into the unlimited domain of music is the cause of that generation.

2. Impure Pleasures

This section will defend the claim that impure pleasures are (a) unlimited in kind, and (b) resemble each other qua subsuming under the restoration model: an account which argues that pleasure arises in a human being when "harmony is regained, and its former nature restored".[14] It will argue for this by first defending the claim that the impure pleasures belong to the unlimited kind in Plato's fourfold ontology. It will then proceed by deciphering an account of the restoration model before outlining how the main varieties of impure pleasures subsume under this model – namely, the restorative/non-restorative and true/false pleasures.

Both Socrates and Protarchus[15] agree that (impure) pleasure "*itself* is unlimited and belongs to the kind that in and by itself neither possesses nor will ever possess a beginning, middle or end".[16] In virtue of lacking exactness and being intrinsically ontologically scalar,[17] impure pleasures are in a constant state of flux.[18] For example, pleasure is always 'pleasanter' relative to its opposite counterpart, pain, and the extent to which a pleasure is *pleasanter* is not finite, but rather a matter of degree: "pleasures seem greater compared to pain, and more intensive, and pain seems, on the contrary moderate in comparison with pleasures".[19] Accordingly, pleasure cannot be characterised as a finite end-product;[20] instead, this paper will argue for the claim that pleasure is a *process* (of restoration).

Pleasures reside in sentient beings (whereby beings themselves are formed of a natural combination of limit and unlimited); hence, impure pleasures arise *in connection with* the mixed kind.[21] According to the restoration model in *Philebus*, this natural combination of the right mixture of the unlimited and limited is harmonious but contingent in living organisms. The process of deviating from this harmonious state

---

[12] John Cooper goes as far as to suggest that their being good is a *consequence* of their being constituted by a combination of the unlimited and limited kind: "To be a good thing just *is* to be a combination" (1977, p. 715).

[13] 27a.

[14] 31da3–8.

[15] The primary interlocutors of *Philebus*.

[16] 31a.

[17] Refer to §1.

[18] 43a.

[19] 42b.

[20] I owe thanks to an anonymous reviewer for their challenge that we could quantify pleasure as a welfare unit and establish welfare as the agreed end. However, this misses the point. The unlimited kind is necessarily unquantifiable; to attribute quantity to pleasure (even by virtue of specifying *parameters*) would be to remove it from the unlimited class and place it into the mixed ontological kind – contrary to what Plato explicitly endorses.

[21] 30a–c.

within an organism can be identified as pain, whereas the process of the restoration of that harmonious state is pleasure – if the process of deviation/restoration is of a sufficient magnitude to be perceived, i.e., affect the soul.[22] One of the "most obvious" examples involves the restoration of harmony to a hungry person (i.e., one who *lacks* food) via the perceived process of filling the 'empty' body with food.[23] As such, it seems that "every pleasure seems to presuppose pain (a lack), just as every process of restoration presupposes a process of destruction".[24]

2.1 Reflective and Nonreflective Impure Pleasures[25]

Socrates provides several examples of the restoration model which convey *physical* imbalance and, seemingly, are only explicitly intended to extend solely to nonreflective pleasures. Namely, those processes of *physical* replenishment that *jointly affect the body and soul* by the same affection; this is what Plato refers to as 'motion perception'.[26] The human body undergoes constant replenishments and sentient beings do not perceive all of them (e.g., formation of an eschar which contributes towards restoring skin). However, when the process of replenishment is intense enough in degree such that one does perceive it, the soul is jointly affected by the bodily replenishment and, subsequently, that process is experienced as pleasurable. Hence, nonreflective pleasures can be regarded as a "psychic epiphenomenalism" of a physical replenishment.[27] Examples of nonreflective pleasures could include feeling a cool breeze on a summer's day (which restores one's bodily temperature), receiving a bodily massage after exercise (which restores muscle tissue by relaxing it), and feeling the sun on your face (such that it restores vitamin D levels in the human body, re-establishing a healthy condition).

In contrast to nonreflective pleasures, reflective pleasures refer to those pleasures that belong to the soul alone.[28] Such pleasures include the pleasures of anticipation:[29] those found in anticipating a future state of affairs. This process involves having previously had sufficiently intense sensory perception (i.e., an affection of both body and soul) and a preservation of that perception in memory which the soul can then authentically and independently recollect to cause present pleasure: "conscious psychic processes caused by entertaining mental representations or images of oneself in conditions that (one thinks) cause [or equate to] pleasure".[30] For example, for a fatigued person to

---

[22] 32b.
[23] 31d–32b.
[24] Fletcher 2018, p. 35. Contrary to Fletcher, I recommend 'deviation' as a more judicious articulation.
[25] Although Plato does not coin terminology to distinguish between different types of pleasures, doing so helps to elucidate the varieties and types of pleasures in *Philebus*. I have borrowed the terms 'reflective' and 'nonreflective' from Tuozzo (1996, pp. 498, 513).
[26] 34d.
[27] Tuozzo 1996, p. 497.
[28] 32c.
[29] It is not made explicit that the pleasures of anticipation are exhaustive of the reflective impure pleasures. However, they adequately demonstrate at least one way in which the restoration model can extend to the reflective pleasures.
[30] Frede 1985, p. 165; Tuozzo 1996, p. 497.

experience pleasure in anticipating going to bed, requires that they have a psychic representation of themselves in the conditions of restoration (i.e., getting sleep) that one thinks would be pleasant in the present/future – based on their memory of past experience.[31] In anticipating yet-to-be-actualised pleasure inducing conditions, one experiences pleasure in the present.

Despite the fact that Plato only explicitly applies his restoration model of pleasure to nonreflective pleasures, it is perhaps insightful that after Socrates introduces the reflective pleasures, he recapitulates the account of pleasure as involving restoration, whilst leaving the relation between pleasure and restoration vague – i.e., such that it need not involve *physical* restorations.[32] According to Thomas Tuozzo, this vague relation between pleasure and process can be interpreted as a causal one whereby pleasure need not be caused by a physical restoration, but by representations of replenishments in the mind of the beholder: pleasures "caused by images representing both the conditions of such a restoration and the pleasure ensuing on such a restoration".[33] In other words, he argues that mental images of conditions (associated with their pleasant consequences) *cause* pleasure in virtue of (a) actually restoring the natural harmony or (b) by appearing to restore the natural harmony. Hence, it can be argued that the restoration model extends to anticipatory pleasures – mental images are sufficient for *causing* pleasure.

However, the claim that pleasure is *caused* is controversial. Near the beginning of *Philebus*, it is argued that the pleasure is the restoration of the natural state inferring that pleasure is the restoration, i.e., the relation is one of identity. Contrastingly, Tuozzo highlights that later in the text, the emphasis shifts to pleasure being caused by the process of restoration: "great changes produce pains and pleasure in us".[34] This ambiguity perhaps reveals Plato's failure to distinguish between pleasure as a process (i.e., the restoration itself) and pleasure as a product (i.e., caused by the process of restoration).[35] Since pleasure is seemingly ambiguous, Tuozzo's choice to adopt the latter account without sufficient justification renders his decision arbitrary.

Alternatively, I argue that it is more plausible to suggest that pleasure is not ambiguous but is rather a process *and* a product in this context: what is 'generated' is a process of replenishment.[36] Imagining a particular mental image may cause pleasure in the sense it initiates a *psychic* process of restoration. For example, anticipating seeing your family (manifesting as a mental, pictorial representation of this) when you miss them *may*[37] trigger some psychic process of restoring a healthy mind (e.g., from a state of anxiety). Furthermore, restoring mental states in virtue of having the relevant mental images could also initiate the restoration of the body by preventing or reducing the impact of physical symptoms of poor mental health, such as panic attacks, fatigue,

---

[31] In the form of an imprinted scribe or painting in their mind (39a–c).
[32] 32d–e.
[33] Tuozzo 1996, pp. 504–05.
[34] 43c.
[35] Frede 1985, p. 169.
[36] This explains why Plato does not distinguish between pleasure as process and pleasure as product: the product *is* a process.
[37] 'May' meaning it is at least *possible* that.

or low mood. In other words, mental representations are capable of instigating mental and/or physical processes of restoration.

2.2 True and False Impure Pleasures

Plato further divides impure (nonreflective/reflective) pleasures into those that are true and those that are false. As such, this section will now attempt to determine how truth and falsehood can coherently be applied to restorative processes.

Dorothea Frede argues that anticipatory (reflective) pleasures are true in virtue of having propositional content by defending the claim that some anticipatory pleasures consist in "definite logoi or pictures representing reality".[38] Hence, she denotes that when one "clearly" hopes (meaning they are certain that the object of hope will actualise) for that which is pleasurable – in virtue of it having the capacity to restore the natural harmony – that such anticipatory pleasure does, in fact, have genuine propositional content. For example, anticipating having a child when you are nine months pregnant is a 'clear' hope and qualifies that anticipatory pleasure as true; there is a commitment to the description of the pleasure. Hence, Frede argues that the process of the restoration is itself a form of pleasure and its relation to reality in virtue of hope (i.e., in the technical sense of a definite prediction) means that pleasure can thereby be true or false: "The only way in which pleasures can be true or false is when he enjoys what his thought is the thought of and when the thought consists in an assertion about facts" past, present, or future.[39]

However, Frede's account is intuitively implausible in light of epistemological concerns regarding when one can ever be *certain* that a pleasure/restoration will actualise. It seems as though her account of 'certain' hope requires one to have supernatural precognition capabilities, which for the majority of persons is unattainable, or even impossible. One might attempt to defend Frede's account against this charge by claiming that if we take feeling and knowing to be sufficiently distinct psychological states, then it could be argued that *feeling* certain that pleasure will actualise does not require knowledge that it will.[40] In other words, Frede's account need not be subject to the epistemological charge because rendering a pleasure true consists in *feeling* certain, rather than possessing knowledge that it will actualise. However, I think that this response is implausible since the choice to categorise only the anticipatory pleasures we *feel* are certain as true – where this feeling is distinct from knowing – would be arbitrary. On what grounds would feeling *certain* in the relevant sense qualify a pleasure as true, as opposed to feeling happy, or anxious?

Furthermore, Frede's account is exclusive to reflective pleasures, and cannot extend to nonreflective pleasures since they can – and often do – occur independently of hope. Accordingly, if we were to endorse Frede's account, we would be advocating for an ambiguous account of the truth. Not only is this problematic in the sense that Plato

---

[38] Frede 1985, pp. 172–73.
[39] Ibid., p. 173.
[40] I owe thanks to an anonymous reviewer for this comment.

never explicitly commends an ambiguous account of truth in *Philebus*, but a defender of Frede's account would still be left with the challenge of identifying in virtue of what nonreflective pleasures qualify as true.

Alternatively, since pleasure can be experienced by pre-linguistic persons (e.g., infants can enjoy eating food[41]) and individuals more often than not experience pleasure without being aware of it 'as a restoration' (e.g., we feel pleasure in warming up on a winter's day, despite being ignorant of the process of the biological restoration itself), pleasure need not – and is unlikely to – have *propositional* content.[42] Considering this, when Plato speaks of "true" and "false" (impure) pleasures, I offer the less-restrictive interpretation that he is referring to whether or not a pleasure is *actually* restorative. For instance, if a man experienced pleasure in feeling the warmth of the sun, that pleasure would be (a) true, if it restored a vitamin D deficiency, or (b) false, if he was wearing strong sunscreen such that no restoration actually occurred.[43] This account of truth is able to consistently explain in virtue of what *both* reflective and nonreflective pleasures can render true, explaining why Plato did not explicitly endorse an ambiguous account of truth in the dialogue. On a charitable reading of *Philebus*, this paper thus renders Frede's account as implausible and argues instead for this latter position: that pleasures are true if genuinely restorative, and false if apparently restorative.

In sum, Section 2 has outlined that pleasures have an unlimited ontological nature in the fourfold division of all things, before claiming that pleasure arises in relation to the mixed class as a perceived restoration of the natural condition. The paper extended the restoration account to nonreflective pleasures, before arguing that impure pleasures are true if they are genuinely restorative and false if they are merely apparently restorative.

3. The Pure Pleasures

This section will outline and defend a plausible account of the pure pleasures by first depicting the general account of the good presented in the dialogue as that which is sufficient, true, beautiful, and moderate. Since the pure pleasures are exclusively ranked amongst the goods at *Philebus* 66a–67b, an account of the pure pleasures must be compatible with an account of the good. The paper will go on to establish that pure pleasures – which include pleasures of appreciating specific colours, sounds, smells, or geometrical shapes, as well as pleasures of learning – subsume under the following identity criteria:[44]

---

[41] Nicklaus 2016, §3.2.

[42] I do not wish to claim that it is impossible for pleasure to have propositional content; rather, I am asserting that we should adopt a less restrictive means of attributing truth/falsehoods to pleasures that can help elucidate how both reflective and unreflective pleasures can be true or false.

[43] I owe thanks to an anonymous referee for the latter example.

[44] 51e7–52a3; Lang 2010: p. 155.

(1) Pure pleasures are preceded by *unperceived* lacks (rational potentials).

(2) Pure pleasures have objects with true, beautiful, and measured properties.

(3) Pure pleasures are sufficient; they have an internal telos.

3.1 The Good

The final ranking of goods for the human life[45] is depicted as follows:[46]

(1) Measure

(2) The well-proportioned, beautiful, perfect, and self-sufficient

(3) Reason and intelligence

(4) Sciences and arts

(5) The pure pleasures

What is important for the purposes of this section is the methodology used to construct this hierarchy.[47] In order to depict the final ranking of the goods, the different categories are judged in relation to the good *itself*, presented as a triadic unity "derived from those features of the good always exhibited by good things" that capture the good in a conjunction of three forms: truth, beauty, and moderation.[48] This demonstrates how all pure pleasures must be compatible with this tripartite conception of the good such that they are included in the hierarchy.

Furthermore, earlier in the dialogue, the good is characterised as "sufficient", meaning that only that kind which is sufficient to itself ('being'), as opposed to the kind that is for the sake of something else ('becoming'), is *intrinsically* good. For example, if one were to drink water for the sake of something else (e.g., relieving thirst), drinking would not qualify as being intrinsically good. Contrastingly, one may appreciate beauty solely for the sake of appreciating beauty – subsequently that act is intrinsically good. After all, Socrates proclaims that "if pleasure really is becoming, then we shall be placing it correctly if we place it in a category other than the good".[49] This paper will argue that that which has an internal telos (a process in which the goal of that process is internal) occurs 'for the sake of itself' and it thus compatible with the good; on the other hand, a process which has an external telos is for the sake of something else. As such, this paper defends the claim that pure pleasures have an internal telos (rational potentials are actualised for the sake of that actualisation).

---

[45] At 28d–30c Plato presents a microcosm-macrocosm argument regarding the human body and the universe; hence, *perhaps* one would be justified in extrapolating the good in the good human life to the good of the universe (Frede 1993, p. 78): "the body of the universe which has the same properties as ours" (30a).

[46] 66a–d.

[47] The notion that the final ranking of the goods is *hierarchical* is merely implicit, though plausible since measure is explicitly regarded as the "most valuable" of the goods (64d–e).

[48] 65a; Lang 2010, p. 165.

[49] 54d1–2.

Although this depiction of the good as measured, beautiful, true, and sufficient is somewhat general and obscure, it is informative enough to determine the nature of the pure pleasures.

3.2 Preceding Unperceived Lack Condition

Socrates describes the pure pleasures as those pleasures that are "based on imperceptible and painless lacks, while their fulfilments are perceptible and pleasant".[50] Hence, as a starting point for depicting the nature of pure pleasures, this section will begin by evaluating the claim that the pure pleasures are necessarily and sufficiently those that are preceded by *imperceptible* painless lacks "apparently construed as signalling the satisfaction of needs we are unaware of, and so not pained by, acquiring or having".[51] For example, having a 'lack of knowledge' is not perceptually a painful experience in its own right,[52] but the process of learning (i.e., fulfilling that 'lack') is a perceived pleasurable experience. Perception in *Philebus* refers to that which is of a sufficient magnitude to affect the soul either independently, or jointly with the body; hence, unperceived lacks do not affect the soul. This is perhaps why the pure pleasures' dependence on lacks does not impede their higher evaluative status – their "cure is slight".[53] This paper will now refer to this condition as the preceding unperceived lack condition (PUL): for a pleasure to be pure, it is (at least necessary) that the pleasure is preceded by an unperceived lack.

However, this paper will argue that the PUL condition cannot solely account for the nature of pure pleasures since, paradoxically, if it did, some pleasures would be welcome in the good life that are not *good*, i.e., compatible with truth, moderation, and beauty. For instance, taking pleasure in eating a dessert out of sheer decadence, laughing (taking pleasure) maliciously as someone else's expense, or taking pleasure in squeezing a pimple are all pleasures that intuitively bear no explicit relation to truth, beauty, or measure. Thus, it seems both in line with Plato's intentions and common sense that not *all* pleasures preceded by painless lacks (e.g., those pleasures that are ugly, false, or immoderate, and subsequently *possibly* incompatible with the good) should be classified as pure and included in the final ranking of the good.

Hence, in order to provide an account of pure pleasure that restricts the scope of the PUL condition such that only good pleasures are welcome in the good life, it could be argued that the pure pleasures must conform to the PUL condition and be *true*. This further condition is not arbitrary. Rather, it is a suitable attempt to restrict the scope of pure pleasures in virtue of the fact the pure pleasures are explicitly coined the "true pure pleasures".[54]As outlined in Section 2, true pleasures can be depicted as those that are genuinely *restorative*. Certain apparent pure pleasures (such as the ones outlined

---

[50] 51b.
[51] 51b5–6; Katz 2016, p. 221.
[52] Having a lack of knowledge could partially constitute a painful experience in certain contexts (e.g., humiliation in a classroom), but the lack of knowledge itself is unperceived.
[53] Lang 2010, p. 155.
[54] 50e–55c.

above) are false and thus not pure; they "have the appearance of enormous size and great variety" but are not genuinely restorative.[55] For example, the decadent man who eats the indulgent chocolate bar is *harming* his body overall (it could contribute to weight gain and thus reduce self-esteem or increase health risks, etc.). However, a weakness of this account is that it defends the claim that the true pure pleasures are *restorative.* This is implausible since such an account conflicts with the sufficiency condition of the good; replenishment "always takes place for the sake of some particular being".[56] As such, it seemingly follows that under this account the telos of the pure pleasures are external – they occur for the sake of restoring a being.

However, Plato's definition of *true pure* pleasures is based on "imperceptible painless lacks, while their fulfilments are perceptible and pleasant", which suggests that the account of truth in *Philebus* is ambiguous:[57] "the 'adjective' true seems to have a different sense in this passage than the analysis of true and false [impure] pleasures earlier in the dialogue, functioning instead as a synonym of 'pure'.[58] Hence, in elucidating that truth and purity are merely synonymous, this section has established a more plausible interpretation of Plato that pure pleasures can be true without having to be genuinely *restorative.* After all, although Socrates endorses the restorative view of pleasure, "he nowhere recommends that they endorse it completely or for every type of pleasure".[59]

Having established that the pure pleasures can be true (i.e., compatible with the good) without having to be restorative, this paper will now consider in what sense is pleasure related to the fulfilling of an unperceived lack if the fulfilling of a lack does not involve a restoration. As an alternative to the restoration model,[60] it could be argued that the *pure* pleasures just are a state of *completion* "thus enjoying that stability of its internal structure and/or its object and qualifying as a candidate for some kind of end".[61] This is more lucid if one considers that the 'lack' of the PUL need not be thought of as gaps (e.g., 'gaps in knowledge') but rather as *potentials* in which the relevant pleasure consists in its end (as an actualisation) occurring for the sake of itself. For example, the potential to see the beauty of a perfect geometrical shape when actualised is pleasurable from the moment that shape is perceived, and when left as a mere potential constitutes a lack in the sense that it remains as a mere *capacity*. What is advantageous about this account is that such pleasures occur for the sake of themselves; they are fulfilled from the instant they are actualised. Hence, this account is compatible with the good, though it makes a stark contrast between the nature of the pure and impure pleasures: "one cannot ignore fundamental distinctions between

[55] 51a7–8.
[56] 54b.
[57] 51b5–6.
[58] Fletcher 2014, p. 127.
[59] Ibid., p. 154.
[60] Although this is an alternative account of pleasure, it still subsumes under the overarching account of pleasure, i.e., that pleasure is a perceived process of change (kinesis) that occurs in sentient beings.
[61] Carone 2000, p. 268.

various classes of pleasures (such as pure and impure ones), against Frede's indiscriminate treatment of all pleasure as a process of replenishment".[62]

However, even if pleasure is regarded as the actualisation of a mere potentiality that occurs for the sake of itself, as opposed to the restoration of some 'lack', such a condition is still not strict enough. Although it now has the potential to be good since it can occur for the sake of itself (i.e., is not a generation despite being true), as mentioned at the start of this section, it still does not exclude certain pleasures that are intuitively pure from being actualised into the good life. For instance, if I have the potential to pop a pimple and I take pleasure in actualising that state of affairs, surely this cannot be a pure pleasure since it is ugly, and thus not compatible with beauty.[63]

3.3 The Objects of Pure Pleasure

It has been argued that the PUL condition cannot independently constitute an identity criterion for the pure pleasures since it fails to restrict the scope of pure pleasure to only those that are good. Since only a sub-group of the pleasures that result from unperceived lacks are pure, it could be argued that some emphasis must be placed on the 'object' of pleasure that bears the required relation to both the unperceived lack (potential) necessary for pleasure and its relation to the good. "Pure pleasures do not accompany the perception of every object, but only the perception of perfect shapes or pure colours or sounds."[64] If the pure pleasures arise from imperceptible painless lacks, the PUL condition is merely necessary; their lacks must also only be 'completed' by objects with specific properties, i.e., those that constitute the good: truth, beauty, and measure.[65] As such, mere 'grasping' or perceiving of an object or activity of pleasure with the relevant properties fills a lack in virtue of actualising a potential (that is compatible with the good, i.e., true, beautiful, and measured).[66]

One property that heightens the status of the pleasures it produces is non-relative beauty, those objects that are "forever beautiful by themselves" and "provide their own specific pleasures".[67] Plato explicitly dismisses the beauty of people or paintings as *wholes* (though he argues that it could be possible for beauty to be abstracted from such sensible objects). Instead, he alludes to the geometrical exactness of plane figures or solids (constructed out of a compass, ruler, and square, and to smooth and bright sounds.[68] Hence, pleasure derived from the perception of a particular object of pleasure need not imply that such properties are sensible: "the object of pleasures must […] be abstracted from the particular sensible object given that it is impossible

---

[62] Ibid., p. 261.

[63] Perhaps opposite properties can occur simultaneously is an object (when abstracted). However, it is certainly not obvious, even if this was possible, how squeezing a pimple can ever be beautiful, especially in the non-relative sense.

[64] Fletcher 2014, p. 124.

[65] Lang 2010, p. 154.

[66] The term 'grasp' is used when the properties of the object are not sensible, such that they cannot be perceived in the conventional sense.

[67] 51c6–7.

[68] 51c4–7, 51d6–8.

to have a sensible object which is white but no other predicate such as 'round' or 'chair'.[69] This non-relative conception of beauty is abstract since it is unintuitive to merely grasp a 'perfect' circle and to admire and appreciate its intrinsic beauty, yet to ignore other 'more obvious' beauty (e.g., that of your partner). However, this theory does cohere with concepts in modern science and philosophy about the nature of beauty as a mathematical golden ratio (*phi*) inherent within nature. Accordingly, depicting the properties of beauty requires rational powers since such properties are not *intuitively* beautiful. As such, the experience of pure pleasure is only accessible to the agent with rational expertise capable of actualising the unperceived rational potential to appreciate beauty.

Measure is also a property of the good which "turns[s] out to be everywhere in beauty and goodness"; this property appears to be of heightened significance for not only does it come first in the hierarchy of the good, but it *imposes limit* on the pure pleasures such that the pure pleasures are categorised as belonging to the mixed ontological kind – an ontological *combination* of the limited kind and the unlimited kind.[70] In virtue of being a *combination*, the pure pleasures do not have the tendency to be excessive or deficient: "any mixture that does not in some way or other possess measure or the nature of proportion will necessarily corrupt its ingredients and most of all itself".[71] Rather, when taking pleasure in an inherently measured object, such pleasures are entirely satiable and stable. However, again, identifying the measured properties in objects and taking pleasure in them requires the exercise of rational powers. For example, identifying which note is measured on the infinite scale of pitch requires expertise – and is immensely pleasurable to the agent who has the rational powers to appreciate it – whereas to the untrained ear, such a sound is merely generic.

Hence, it seems that pure pleasures are those necessarily preceded by rational, painless (unperceived) potentials ('lacks') whose objects cohere with the good: they are true (synonymous with pure) and their objects are inherently measured and beautiful. It is important to recognise that under these further constraints, purity bears a qualitative relation to pleasure (not quantitative), the pure pleasure is superior to impure pleasure not in its magnitude but in its very nature, implying that Plato was not seeking to promote a hedonistic maximisation model of pleasure. "Every small and insignificant pleasure that is unadulterated by pain will turn out to be pleasanter, truer and more beautiful than a greater quantity and amount of the impure kind."[72]

This paper has thus far outlined the nature of the pure pleasures. However, it will now present one potential objection concerned that this account has restricted the scope of pure pleasures too much, since it is not obvious how the pure pleasures of learning meet such conditions (perhaps in part because it is not obvious what Socrates means by the pleasures of learning in general).[73] Although the pleasures of learning seemingly conform to the PUL condition (there is, at least intuitively, no precedent

[69] Lang 2010, p. 157.
[70] 64e6–7, cf. 65a; Cooper 1977, p. 715.
[71] 64d8–e2.
[72] 53c.
[73] Lang 2010, p. 155.

perceived lack prior to learning), it is not obvious in what sense the object of the pleasure (learning) can have properties such as truth, beauty, and moderation.

On the one hand, it could be argued that the 'object' of pleasure refers to its content, such as "learning that the symbol of gold is Au"; as such, if the content of learning is true, moderate, and beautiful, that pleasure is pure.[74] However, since the content of a pleasure of learning can be of any variety (Plato does not discriminate) such objects could involve learning that, for example, one's dog has died a painful premature death, which is immoderate ('violent'), and certainly not beautiful in the non-relative sense. Hence, this reading is implausible.[75]

Contrastingly, one could appeal to Plato's characterisation of the soul "in which opinions or questions with propositional content are written down […] followed by the illustration of that judgement as a painting in the soul" to suggest that the pleasures of learning involve the process of remembering 'objects' in the form of 'scribes' or 'paintings'.[76] Here, the distinctive feature of the pleasures of learning lies in the *memory and judgement* of that pleasure's 'object' as an imprint on the soul: "I cannot properly classify the world through my senses if I do not have memory or judgement to rely upon, and so I cannot take pleasures in even those aspects of it that are true, measured , and sufficiently themselves."[77] Here, then, it appears the pleasures of learning (i.e., judging and remembering) act as a means to classify the world as it is and thereby take pleasure in perceiving and recalling that which is true, pleasant, and beautiful. However, our judgements and memories can be fallacious since it is not obvious that they do in fact depict or recall the world as it really is.

Alternatively, perhaps the most plausible account of the pleasure of learning (insofar as it makes such pleasures compatible with the good) is to depict is as the pleasure of acquiring (or 'recollecting'[78]) knowledge (*epistêmê*), which "in its most accurate sense and appropriate use" is "applied to insights into true reality".[79] Although this paper does not have space to defend a full account of knowledge as presented in *Philebus*, it is perhaps plausible to claim that by 'true reality' Plato was referencing the Forms:

> The world that appears to our senses is in some way defective and filled with error, but there is a more *real* and perfect realm, populated by entities (called "forms" or "ideas") that are eternal, *changeless*, and in some sense paradigmatic.[80]

Under this account, the objects of knowledge are the Forms and the process of understanding (learning/recollecting) the Forms is pleasurable. Since the Forms are 'exemplars', they are true (i.e., have a place in reality), moderate (stable, 'unchanging'),

---

[74] Ibid.
[75] 52c5.
[76] Lang 2010, p. 156.
[77] Ibid., p. 158.
[78] Plato's epistemological views presented in *Meno* suggest that "what we think of as discovery [of knowledge] is in fact the recovery of knowledge which the soul has previously possessed but which it has forgotten" (Taylor 2008, p. 4).
[79] 59d.
[80] Kraut 2017, §1.

and perhaps beautiful[81] – the objects of the pleasure of learning do in fact make such pleasures appropriate for the good life. This account requires a substantial amount of further justification, though it does help to elucidate at least one way in which the pleasures of learning can be seen to be compatible with the tripartite structure of the good presented in *Philebus*.

In sum, this section has outlined the general account of the good as presented in *Philebus* and defended the claim that the pure pleasures must be preceded by an unperceived lack (rational potentials), have objects that are true, moderate, and beautiful, and are *sufficient*.

4. Conclusion

This paper has argued for the claim that pleasure is a process of change (kinesis) that occurs in sentient beings either when the harmonious natural condition is genuinely or apparently restored (impure pleasures), or when certain potentials are actualised by the rational human (pure pleasures). It is noteworthy that despite their differences, pleasure is not necessarily ambiguous in *Philebus*: "any worthwhile discipline finds a unity in opposites, so there is nothing to be surprised about in pleasure's situation".[82] However, this is worthy of a paper in its own right.

A limitation of this paper is that it is unable, and has thus failed, to *prove* that the pure and impure pleasures are *exhaustive* of the varieties of pleasure presented in *Philebus*. One reason for this is the fact that Plato himself does not label all varieties of the pleasures depicted in this account. However, on a charitable reading of his text, the majority (if not all) of the pleasures presented in *Philebus* can either meet the identity criterion of pure pleasures or are characterised (somewhat) by being a restoration of some precedent perceived lack (impure). The task of demonstrating that *all* pleasures subsume under either model would be an incredible feat, though perhaps this is a challenge worthy of further research. Despite such a limitation, this charitable reading of *Philebus* is plausible. This account has pulled together a complex multiplicity of intricate concepts and interpretations to provide a coherent, in-depth analysis of the nature of pleasure as presented in Plato's *Philebus*.

[81] Since the Forms are abundant (e.g., in *Republic* X.596b, there is even reference to a Form of Bed), it is difficult at this stage to establish, without further tangential discussion, as to whether or not all of the Forms are beautiful – though it is certainly possible.

[82] Gosling & Taylor 1982, p. 131.

References

Carone, G. R. (2000). 'Hedonism and the Pleasureless Life in Plato's *Philebus*.' *Phronesis, 45*(4), pp. 257–83. https://www.jstor.org/stable/4182653.

Cooper, J. M. (1977). 'Plato's Theory of Human Good in the *Philebus*.' *The Journal of Philosophy, 74*(11), pp. 714–30. https://doi.org/10.2307/2025776.

Cooper, N. (1968). 'Pleasure and Goodness in Plato's *Philebus*.' *The Philosophical Quarterly, 18*(70), pp. 12–15. https://doi.org/10.2307/2218024.

Davidson, D. (1990). *Plato's* Philebus. New York: Garland Publishing.

Fletcher, E. (2014). 'Plato on Pleasure and the Best Life.' *Phronesis, 59*(2), pp. 113–42. https://doi.org/10.1163/15685284-12341263.

Fletcher, E. (2017). 'The Divine Method and the Disunity of Pleasure in the *Philebus*.' *Journal of the History of Philosophy, 55*(2), pp. 195–206. https://doi.org/10.1353/hph.2017.0026.

Fletcher, E. (2018). 'Two Platonic Criticisms of Pleasure.' *Pleasure: A History*, edited by L. Shapiro, pp. 15–41. New York: Oxford University Press.

Frede, D. (1985). 'Rumpelstiltskin's Pleasures: True and False Pleasures in Plato's Philebus.' *Phronesis, 30*(2), pp. 151–80. https://doi.org/10.1163/156852885X00039.

Gill, M. L. (2019). 'The Fourfold Division of Beings: *Philebus* 23b–27c.' In *Plato's* Philebus*: A Philosophical Discussion*, edited by P. Dimas, R. E. Jones, and G. R. Lear, pp. 71–89. Oxford Scholarship Online. https://doi.org/10.1093/oso/9780198803386.003.0005.

Gosling, J. C. B, and Taylor, C. C. W. (1982). *The Greeks on Pleasure*. Oxford: Clarendon Press.

Katz, L. D. (2016). 'Pleasure.' In *The Stanford Encyclopedia of Philosophy*, Winter 2016 ed., edited by E. N. Zalta. https://plato.stanford.edu/archives/win2016/entries/pleasure/.

Kraut, R. (2017). 'Plato.' In *The Stanford Encyclopedia of Philosophy*, Fall 2017 ed., edited by E. N. Zalta. https://plato.stanford.edu/archives/fall2017/entries/plato.

Lang, P. M. (2010). 'The Ranking of the Goods at *Philebus* 66a-67b.' *Phronesis, 55*(2), pp. 153–69. https://doi.org/10.1163/156852810X489049.

Nicklaus, S. (2016). 'The Role of Food Experiences During Early Childhood in Food Pleasure Learning.' *Appetite, 104*, pp. 3–9. https://doi.org/10.1016/j.appet.2015.08.022.

Plato. (1941). *Philebus*, translated by D. Frede. Indianapolis: Hackett Publishing Company.

Taylor, C. C. W. (2008). 'Plato's Epistemology.' In *The Oxford Handbook of Plato*, edited by G. Fine. Oxford Handbooks Online. https://doi.org/10.1093/oxfordhb/9780195182903.003.0007.

Tuozzo, T. M. (1996). 'The General Account of Pleasure in Plato's *Philebus*.' *Journal of the History of Philosophy, 34*(4), pp. 495–513. https://doi.org/10.1353/hph.1996.0084.

Tuozzo, T. (2018). 'Rethinking the Division of *Pleasure* in Plato's *Philebus*.' *Proceedings of the XXIII World Congress of Philosophy, 2*(2), pp. 325–29. https://doi.org/10.5840/wcp232018221328.

Undergraduate Philosophy Journal of Australasia www.upja.online
VOL. 2, NO. 2, DECEMBER 2020

Star Models and the Semantics of Infectiousness

MATTHEW W. G. MCCLURE*

UNIVERSITY OF EDINBURGH

Abstract

The first degree entailment (FDE) family is a group of logics, a many-valued semantics for each system of which is obtained from classical logic by adding to the classical truth-values *true* and *false* any subset of {*both*, *neither*, *indeterminate*}, where *indeterminate* is an *infectious* value (any formula containing a subformula with the value *indeterminate* itself has the value *indeterminate*). In this paper, we see how to extend a version of star semantics for the logics whose many-valued semantics lack *indeterminate* to star semantics for logics whose many-valued semantics include *indeterminate*. The equivalence of the many-valued semantics and star semantics is established by way of a soundness and completeness proof. The upshot of the novel semantics in terms of the applied semantics of these logics, and specifically infectiousness, is explored, settling on the idea that infectiousness concerns ineffability.

1. Introduction

The interesting relationships between *strong Kleene logic* (K3), the *logic of paradox* (LP), and *classical logic* are well known, as are relationships between these three logics and *first degree entailment* (FDE), the conditional-free fragment of relevant logics.[1] When ordered by strength, this quartet forms a lattice structure: the standard semantics for LP has a truth-value in between *true* and *false* interpreted as *both*, K3 an intermediate truth-value interpreted as *neither*, and FDE (in its four-valued form) has both of these.

More recently, Graham Priest has generalised this quartet to an octect he calls 'the FDE family' which contains, in addition to the four logics mentioned, *weak Kleene logic*, some other 'logics of nonsense', and a logic developed alongside Jay Garfield in studying Nāgārjuna's use of the *catuṣkoṭi*.[2] In the present paper, I extend existing star

---

* *Matthew W. G. McClure is in the second year of the undergraduate philosophy programme at the University of Edinburgh. Matthew isn't sure exactly what their philosophical interests are, but logic is probably among them.*

[1] Some classic references and contemporary discussions are Asenjo 1966; Beall 2017, 2018, 2019; Belnap 2019a, 2019b; Dunn 1976; Omori & Wansing 2017; and Priest 1979. Note that strong Kleene logic differs from the logic of paradox only by the intermediate value not being designated (as would befit *neither true nor false*).

[2] Priest 2014, 2010, 2019; Garfield & Priest 2009.

semantics (a form of semantics making use of an operator * rather than extra truth-values) to cover the whole family. This has implications for how we interpret the logics.

First, I introduce the FDE family and its many-valued semantics (§2). Then I give generalisations of existing star semantics for four of the logics in the family – first degree entailment, the logic of paradox, strong Kleene, and classical logic. The main contribution of the present paper (§3) is novel star semantics for the '*i*-variants', logics in the family whose many-valued semantics involves the truth-value i. After presenting and discussing the star semantics for these logics, I show that the many-valued semantics and the star semantics are equivalent (§4). I then consider how the star semantics affects interpretations of these logics, and specifically of the idea of infectiousness (§5). I suggest that three interpretations of infectiousness (the *nonsense* interpretation, the *off-topic* interpretation, and the *emptiness* interpretation) converge with one another, with infectiousness capturing something like ineffability.

2. The FDE Family and *i*

2.1 The FDE Family

The FDE family consists of two quartets, each with a lattice structure when ordered by strength (*L* is properly stronger than *K* iff everything which is *K*-valid is *L*-valid, but not the other way around). The FDE quartet has first degree entailment (**BN**) as its weakest logic and classical logic (**Ø**) as its strongest, with strong Kleene logic (**N**) and the logic of paradox (**B**) between them. The other four logics are what I will call the *i-variants* of each of these systems, logics obtained by adding the value *i* to a many-valued semantics, which are similarly arranged, with each *i*-variant being weaker than its *i*-free twin. Figure 1 shows the relationship between the logics.

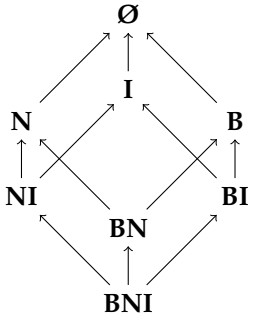

Figure 1: The FDE family ordered by strength.

Many of the logics in the FDE family are familiar to logicians. Table 1 records the details. *t* is the value true only, *f* false only, *b* both true and false, *n* neither true nor false, and *i* the infectious indeterminate value (some authors write '*e*' for 'empty' for this value). The names of the logics in this paper are taken from the values added to *t*, *f* to arrive at the many-valued semantics for the logic in question.

Often called	Truth-values	Name in this paper
Classical logic	t, f	**Ø**
Logic of paradox	t, f, b	**B**
Strong Kleene	t, f, n	**N**
Weak Kleene	t, f, i	**I**
First degree entailment	t, f, b, n	**BN**
S_{fde}	t, f, b, i	**BI**
——	t, f, n, i	**NI**
FDE_{ϕ}	t, f, b, n, i	**BNI**

Table 1: Logics in the FDE family.[3]

Throughout this paper, a many-valued semantics for one of these logics is marked with a '+' (e.g., **BN**+ is the many-valued semantics for first degree entailment) whereas a star semantics is marked with a '*' (e.g., **BN*** is the star semantics for first degree entailment). This paper is concerned only with the propositional systems.

Before providing a many-valued semantics, let us, for sake of explicitness, define our vocabulary (which is the same no matter the sort of semantics). The set of sentences or formulae Sent is defined inductively from the set of propositional parameters Prop = $\{p, q, \ldots, p_1, \ldots\}$, where $A, \ldots$ are metavariables standing for sentences:

- All propositional parameters are sentences.
- If $\ulcorner A \urcorner$ is a sentence, then $\ulcorner \neg A \urcorner$ is a sentence.
- If $\ulcorner A \urcorner$ and $\ulcorner B \urcorner$ are sentences, then $\ulcorner (A \vee B) \urcorner$ is a sentence.
- If $\ulcorner A \urcorner$ and $\ulcorner B \urcorner$ are sentences, then $\ulcorner (A \wedge B) \urcorner$ is a sentence.

The material conditional can be defined as an abbreviation: $\ulcorner (A \supset B) \urcorner := \ulcorner (\neg A \vee B) \urcorner$. It shall not be discussed in detail. For the most part, logics in the FDE family are extended by other conditionals, e.g., strict and relevant conditionals. In all but **Ø**, it fails to satisfy at least one of $\vDash A \supset A$ and $A, A \supset B \vDash B$.

[3] Priest 2019, p. 281.

2.2 Many-valued Semantics

We now turn to the many-valued semantics. A **BNI**+ model m:Sent $\rightarrow V$ is a mapping of sentences to the truth-values V constrained by the evaluation scheme, which is shown in Table 4. (V is the union of $\{t, f\}$ and some subset of $\{b, n, i\}$, as in Table 1.) The scheme for any logic with less than the full set of truth-values just omits those entries in the tables containing truth-values the logic lacks. In **BN**+ and stronger, the truth-values can be thought of as forming a lattice as in Figure 2,[4] with conjunction as the greatest lower bound, and disjunction as the least upper bound. (Negation is a De Morgan involution with b, n, and e fixed points.)

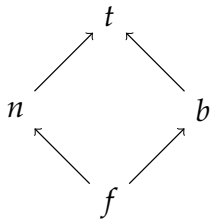

Figure 2: The *truth* partial order on $\{t, f, b, n\}$.[5]

$\neg$	
t	f
f	t
b	b
n	n
i	i

$\wedge$	t	f	b	n	i
t	t	f	b	n	i
f	f	f	f	f	i
b	b	f	b	f	i
n	n	f	f	n	i
i	i	i	i	i	i

$\vee$	t	f	b	n	i
t	t	t	t	t	i
f	t	f	b	n	i
b	t	b	b	t	i
n	t	n	t	n	i
i	i	i	i	i	i

Table 2: The connectives in **BNI**+.

In systems with $b \in V$, the set of designated – roughly, *at least true* – values $D = \{t, b\}$; otherwise $D = \{t\}$.

Definition. A model m:Sent $\rightarrow V$ *satisfies* a sentence A iff $m(A) \in D$.

Definition. A model m *satisfies* set of sentences Γ iff for all $A \in \Gamma$, m satisfies A.

[4] More generally, they form a bilattice (a set plus two lattice orderings) with a truth ordering (shown) and an information ordering in which b carries maximal information, n minimal information, and t, f incomparable with one another in between b and n.

[5] Belnap 2019a, p. 60.

Validity is then informally understood as *designation preservation over all models*.[6] We can define a multiple-premiss, single-conclusion consequence relation in the ordinary way: $\Gamma \models A$ iff any model satisfying Γ satisfies A. Formally:

$$\Gamma \underset{+}{\models} A \Leftrightarrow \forall m(\forall B_{\in\Gamma}(m(B) \in D) \Longrightarrow (m(A) \in D).$$

3. Star Semantics

*3.1 The Basic Picture: BN**

Star semantics are a form of Kripke-style semantics developed for FDE by Richard Sylvan and Val Plumwood.[7] The basic insight is that instead of adding truth-values, we add points (also called 'situations', 'set-ups', or 'worlds' – though this term might have metaphysical undertones we want to avoid) and make negation an intensional rather than extensional operator. This results in a simpler, two-valued, evaluation scheme.

The following is a simplified form of the variant of star semantics for **BN** I will call **BN***.[8] (The semantics in its most general form comes later.)

A model M is a triple $\langle W, ^*, \Vdash \rangle$ satisfying the following constraints:

- $W = \{@, \dots\}$ is a set of points (@ is a designated point);
- $^*: W \to W$ is a function on points satisfying
  - $w^{**} = w$ (*involution*); and
- $\Vdash \subseteq W \times \text{Sent}$ such that for $w \in W$:
  - $w \Vdash (A \vee B)$ iff $w \Vdash A$ or $w \Vdash B$;
  - $w \Vdash (A \wedge B)$ iff $w \Vdash A$ and $w \Vdash B$;
  - $w \Vdash \neg A$ iff $w^* \nVdash A$.

Figure 3 gives an example.

---

[6] A reviewer thought it a good idea for me to explain why I am not using the term 'truth preservation', which might be more familiar. *Designation preservation* concerns the designated values t, b. Since there is a truth-value t for *true only*, *truth preservation* might most naturally be thought of as preservation of the value t, but this is not what we want: t is merely one of the values we are interested in preserving (the *designated* ones). 'Designation preservation' avoids this confusion.

[7] Routley & Routley 1972.

[8] This differs in some ways from standard contemporary presentations of FDE's star semantics. For those, see, e.g., Omori & Wansing 2017, p. 1024; Priest 2008, pp. 151–52.

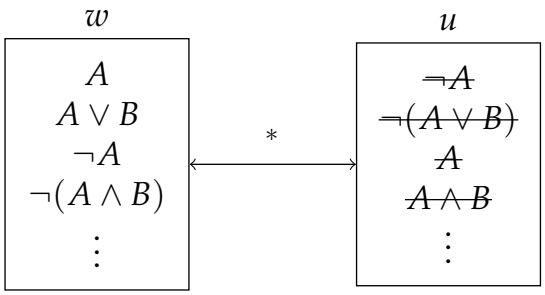

Figure 3: An example star model (*A* at a point means *A* fails there).

Definition. A point w (in some model) *satisfies* a sentence A iff $w \Vdash A$ (in that model).

Definition. A model $M = \langle W, N, ^*, \Vdash \rangle$ *satisfies* a sentence A iff $@ \Vdash A$ ($@ \in W$).

Definition. A model M *satisfies* set of sentences Γ iff for all $A \in \Gamma$, M satisfies A. (Same as before.)

The definition of multiple-premiss, single-conclusion consequence is broadly the same as before – $\Gamma \models A$ iff any model satisfying Γ satisfies A – but differs in terms of satisfaction of a sentence by a model:

$$\Gamma \underset{*}{\models} A \Leftrightarrow \forall M(\forall B_{\in \Gamma}(@ \Vdash B) \Longrightarrow (@ \Vdash A)).$$

*3.2 Extensions of BN**

The semantic systems **B***, **N***, and **Ø*** are obtained from **BN*** as just presented by adding to the model structure and imposing additional constraints. A model M is in this case a quintuple $\langle W, G^+, G^-, ^*, \Vdash \rangle$ constrained just as a **BN*** model, but with the further constraints that

- $G^+, G^- \subseteq W$ (G^+ and G^- are subsets of W);
- for $w \in G^+$:
  - $w \Vdash A$ or $w^* \nVdash A$ (*intensional exhaustion*);[9]
- for $w \in G^-$:
  - $w \nVdash A$ or $w^* \Vdash A$ (*intensional exclusion*).

BN* in its more general form has this quintuple model structure and satisfies these constraints, but places no further constraints on the models.

B* adds to **BN*** the constraint that

- $@ \in G^+$,

[9] This principle comes from Beall 2009, p. 9.

ensuring (by exhaustion) that @ satisfies at least one of A, $\neg A$ for all A, ruling out the equivalent of n.[10]

N* adds to **BN*** the constraint that

- $@ \in G^-$,

ensuring (by exclusion) that @ satisfies at most one of A, $\neg A$ for all A, ruling out the equivalent of b.

Ø* adds to **BN*** the constraint that

- $@ \in G^+ \cap G^-$,[11]

ensuring @ satisfies exactly one of A, $\neg A$ for all A.[12]

3.3 Generalisation to the i-*variants*

We are now in a position to present the semantics for **BNI***. It is similar to **BN***, though with different constraints. A model M is a sextuple $\langle W, N, G^+, G^-, ^*, \Vdash \rangle$ satisfying the following constraints:

- $W = \{@, \ldots\}$;
- $N, G^+, G^- \subseteq W$;
- $^*: W \to W$ is a function on points satisfying
  - $w^{**} = w$ (*involution*);
  - $w \in N \Longrightarrow w^* \in N$ (*closure*: N is closed under *);
- $\Vdash \subseteq W \times$ Sent such that
  - for $w \in G^+$:
    - for $w \in N$:
      - $w \Vdash A$ or $w^* \nVdash A$;
  - for $w \in G^-$:

---

[10] One will note that if w satisfies exhaustion, then w^* will satisfy exclusion, and that if w satisfies exclusion, w^* will satisfy exhaustion.

[11] One can also get **Ø*** from the simpler **BN*** semantics by imposing constraint that * satisfies $w^* = w$ (*identity*). This ensures that *every point* satisfies exactly one of A, $\neg A$ (ruling out the equivalents of b and n). Note that any identity function is an involutory function, so we need not explicitly impose the constraint that $w^{**} = w$.

[12] 'What about $G^+ \cup G^-$?', one might ask. The constraint that $@ \in G^+ \cup G^-$ should yield *symmetric three-valued logic* (Field 2008, pp. 78–81). Importantly, $A \wedge \neg A \vDash B \vee \neg B$, which is valid in this logic but not **BN**, will turn out valid: any model with $@ \in G^+$ will have $@ \Vdash B \vee \neg B$, and no model with $@ \in G^-$ will have $@ \Vdash A \wedge \neg A$, so every model with $@ \in G^+ \cup G^-$ will either fail to satisfy $A \wedge \neg A$ or satisfy $B \vee \neg B$.

        - $w \nVdash A$ or $w^* \Vdash A$;
    - for $w \in N$:
        - $w \Vdash (A \vee B)$ iff $w \Vdash A$ or $w \Vdash B$,
        - $w \Vdash (A \wedge B)$ iff $w \Vdash A$ and $w \Vdash B$,
        - $w \Vdash \neg A$ iff $w^* \nVdash A$; and
    - for $w \notin N$:
        - $w \Vdash (A \vee B)$ iff $w \Vdash A$ or $w \Vdash B$ and *w acknowledges A* and *B*,
        - $w \Vdash (A \wedge B)$ iff $w \Vdash A$ and $w \Vdash B$ and *w acknowledges A* and *B*,
        - $w \Vdash \neg A$ iff $w^* \nVdash A$, and
        - $w \Vdash A$ iff $w^* \nVdash A$.

Note that G^+'s exhaustion condition holds for points in N (that is, $G^+ \cap N$).

Definition. A point *w acknowledges* a sentence A iff w stands in $\Vdash$ to A or $\neg A$.

The definition of validity is unaffected, and business is as usual for *normal points* (those in N). Let us unpack those conditions on the points outside of N (*abnormal points*). The constraint that $w \Vdash A$ iff $w^* \nVdash A$ ensures that whenever A holds at w (w abnormal), A fails at w^*, and whenever A holds at w^*, A fails at w. By the truth-conditions of negation, this means that whenever A holds at an abnormal point, so does $\neg A$, and whenever $\neg A$ holds, so does A.

The acknowledgement condition requires that a point w stands in $\Vdash$ to A, $\neg A$. By the constraint on negation, the condition simplifies to:

Definition. A point *w acknowledges* a sentence A iff w stands in $\Vdash$ to A.

The above conditions for conjunction and disjunction are therefore equivalent to:

- $w \Vdash (A \vee B)$ iff $w \Vdash A$ and $w \Vdash B$,
- $w \Vdash (A \wedge B)$ iff $w \Vdash A$ and $w \Vdash B$.

The result is that for a formula B containing any of the connectives as a major connective, an abnormal point w satisfies B just in case it satisfies all of the subformulae which are arguments of the main connective, and each of those are satisfied just in case the same condition holds with respect to their main connective (or, if the subformula in question is a propositional parameter p, w satisfies p). Conversely, should any propositional parameter p occurring in a formula B be unsatisfied at an abnormal point w, the smallest subformula of B containing p as an argument of a connective will be unsatisfied by w (if there is one – if not, B fails at w trivially), and the smallest subformula of B containing this subformula as an argument of a connective shall likewise be unsatisfied by w (if there is one, else the buck stops here), …, and so B shall itself be unsatisfied by w.

All this is to say that nonsatisfaction by an abnormal point is *infectious* – the failure of any subformula of a formula at an abnormal point ensures the failure of that formula there.

Note that this more general model structure holds for **BN***, **B***, and **N*** too in their most general form, but is moot, since there are no abnormal points ($N = W$ so $W - N = \varnothing$) – hence, we can use the simpler semantics without any worries. From these more general models we obtain those of **BN*** and its extensions by imposing the constraint that $@ \in N$.

To obtain **BI***, **NI***, and **I*** from **BNI***, impose the same constraints that when imposed on **BN*** yield **B***, **N***, an **Ø***, respectively:

logic	constraint
BI*	$@ \in G^+$
NI*	$@ \in G^-$
I*	$@ \in G^+ \cap G^-$

4. Equivalence of Many-valued and Star Semantics

In this section, we see that the star semantics and the many-valued semantics are equivalent for the logics **BNI**, **BI**, **NI**, and **I**. (I do not discuss the other logics in detail for sake of brevity.)

4.1 Natural Deduction for the FDE Family

In recent work, Priest provides natural deduction systems for the FDE family.[13] Let us mark a natural deduction system of this kind with a subscript 'G' (e.g., $\mathbf{BN}_G$ is the natural deduction system for **BN**). Priest has proved these systems sound and complete relative to the many-valued semantics for all the logics in the FDE family. Soundness and completeness results relative to the star semantics then establish the equivalence of the star and many-valued semantics for the logic in question.

In the natural deduction systems, a *basic deduction* in the system is of the form A; *complex deductions* are formed by applying rules to basic deductions and other complex deductions. Then $\Gamma \vdash B$ iff B is at the end of a deduction whose undischarged assumptions (if there are any) are all in Γ. For example, $\{A\} \vdash A$ since A is a deduction whose undischarged assumptions are only in $\{A\}$.

The rules are given in Table 3, in which a double line means that a rule goes both ways, $\phi(A)$ may be any sentence containing all of the propositional parameters that occur

[13] Priest 2019.

within A (Priest's notation for this is 'A^{+}'), and $[A]^n$ is an assumption discharged by the rule labelled with 'n'. Systems for stronger i-variants add rules from Table 4:

logic	extra rules
$\mathbf{BI}_G$	wxm
$\mathbf{NI}_G$	efq
$\mathbf{I}_G$	wxm, efq

$$\text{dn: } \frac{A}{\neg\neg A} \qquad \text{dem: } \frac{\neg(A \wedge B)}{\neg A \vee \neg B} \quad \frac{\neg(A \vee B)}{\neg A \wedge \neg B}$$

$$\text{adj: } \frac{A \quad B}{A \wedge B} \qquad \text{s: } \frac{A \wedge B}{A} \quad \frac{A \wedge B}{B}$$

$$\text{wad: } \frac{A \quad \phi(B)}{A \vee B} \quad \frac{\phi(A) \quad B}{A \vee B}$$

$$\text{sc: } \frac{A \vee B \qquad \begin{array}{c}[A]^1 \\ \vdots \\ C\end{array} \qquad \begin{array}{c}[B]^2 \\ \vdots \\ C\end{array}}{C}\ {}^{1,2}$$

Table 3: Priest's rules for $\mathbf{BNI}_G$.

$$\text{efq: } \frac{A \wedge \neg A}{B} \qquad \text{wxm: } \frac{\phi(A)}{A \vee \neg A}$$

Table 4: The extra rules for stronger i-variants.

In what follows, I shall write the inverse of the star function thus: ${}^*w = u$ iff $u^* = w$ – sc. $({}^*w)^* = {}^*(w^*) = w$.

4.2 Soundness

Theorem. $\Gamma \vdash_{\mathrm{BNI}_G} A$ only if $\Gamma \vDash_{\mathrm{BNI}^*} A$. Soundness for **BNI**.

Proof. By recursion: the base case shows that the basic deduction ($A \vdash A$) is valid, and the step cases show that satisfaction carries over each rule. Base case: $A \vdash A$ only if $A \vDash A$. Take arbitrary model M; if M satisfies A then $@ \Vdash A$, which is the conclusion.

The general idea for the step cases is that we start with a deduction

$$\dfrac{\begin{array}{cccc}\Gamma_1 & \Gamma_2 & & \\ \vdots & \vdots & \dots & \\ input_1 & input_2 & & \end{array}}{rule\ output}$$

and for each rule show that any model M satisfies $\Gamma = \Gamma_1 \cup \Gamma_2 \cup \cdots$ (and the *rule inputs* thereby) only if M satisfies the *rule output*. I will (except in the case of sc) leave the deduction from Γ to the rule inputs implicit for sake of brevity and simplicity, and show that M satisfies the inputs only if it satisfies the outputs. (Another way to think of this is to restrict M to models satisfying Γ.)

- dn↑: $\neg\neg A \vDash A$. Suppose arbitrary M satisfies $\neg\neg A$. Then $@ \Vdash \neg\neg A$. By the truth-conditions of $\neg$, $@ \Vdash \neg\neg A$ iff $@^* \nVdash \neg A$ iff $@^{**} \Vdash A$. By involution, $@^{**} = @$, so $@ \Vdash A$.

- dn↓: $A \vDash \neg\neg A$. Suppose arbitrary M satisfies A. Then $@ \Vdash A$. By the truth-conditions of $\neg$, $@ \Vdash A$ iff ${}^*@ \nVdash \neg A$ iff ${}^{**}@ \Vdash \neg\neg A$. By involution, ${}^{**}@ = ({}^{**}@)^{**} = @$, so $@ \Vdash \neg\neg A$.

- dem_L ↓: $\neg(A \wedge B) \vDash \neg A \vee \neg B$. Take arbitrary M such that $@ \Vdash \neg(A \wedge B)$. If $@ \notin N$, $@ \Vdash \neg(A \wedge B)$ iff $@ \Vdash A \wedge B$ iff $@ \Vdash A$ and $@ \Vdash B$ iff $@ \Vdash \neg A$ and $@ \Vdash \neg B$ iff $@ \Vdash \neg A \vee \neg B$. In case $@ \in N$, we have $@^* \nVdash A \wedge B$; so, by the truth-conditions of $\wedge$, one of A, B must fail at $@^*$: $@ \nVdash A$ or $@ \nVdash B$. By the truth-conditions of $\neg$, in former case $@ \Vdash \neg A$, so by the truth-conditions of $\vee$, $@ \Vdash \neg A \vee \neg B$; in latter case, $@ \Vdash \neg B$ and thus $@ \Vdash \neg A \vee \neg B$.

- dem_L ↑: $\neg A \vee \neg B \vDash \neg(A \wedge B)$. Take arbitrary M such that $@ \Vdash \neg A \vee \neg B$. The abnormal case is trivial. If $@ \in N$, either $@ \Vdash \neg A$ or $@ \Vdash \neg B$. From these follow, by the truth-conditions of $\neg$, $@^* \nVdash A$ and $@^* \nVdash B$, respectively. In either case, the truth-conditions for $A \wedge B$ are not met at $@^*$, so $@^* \nVdash A \wedge B$. But $@ \Vdash \neg(A \wedge B)$ iff $@^* \nVdash A \wedge B$.

- dem_R ↓: $\neg(A \vee B) \vDash \neg A \wedge \neg B$. Take arbitrary M such that $@ \Vdash \neg(A \vee B)$. The abnormal case is trivial. If $@ \in N$, we have $@^* \nVdash A \vee B$, so, by the truth-conditions of $\vee$, $@^* \nVdash A$ and $@^* \nVdash B$. By the truth-conditions of $\neg$, $@ \Vdash \neg A$ and $@ \Vdash \neg B$, and consequently $@ \Vdash \neg A \wedge \neg B$.

- dem_R ↑: $\neg A \wedge \neg B \vDash \neg(A \vee B)$. Take arbitrary M such that $@ \Vdash \neg A \wedge \neg B$. The abnormal case is trivial. If $@ \in N$, truth-conditions of $\wedge$ yield that $@ \Vdash \neg A$ and $@ \Vdash \neg B$, so $@^* \nVdash A$ and $@^* \nVdash B$. Therefore, the truth-conditions of $A \vee B$ cannot be met at $@^*$, so $@^* \nVdash A \vee B$, and thus, by the truth-conditions of $\neg$, $@ \Vdash \neg(A \vee B)$.

- adj: $\{A, B\} \vDash A \wedge B$. Take arbitrary M satisfying $\{A, B\}$; M then satisfies A and satisfies B, so $@ \Vdash A$ and $@ \Vdash B$. By the truth-conditions of $\wedge$, $@ \Vdash A \wedge B$.

- s: $A \wedge B \vDash A$. Take arbitrary M such that $@ \Vdash A \wedge B$. By the truth-conditions of $\wedge$, $@ \Vdash A$. The B case is analogous.
- wad: $\{A, \phi(B)\} \vDash A \vee B$. Take arbitrary M such that $@ \Vdash A$ and $@ \Vdash \phi(B)$. If $@ \in N$, then, by the truth-conditions of $\vee$, $@ \Vdash A \vee B$. If $@ \notin N$, then, by definition of $\phi()$, $@$ stands in $\Vdash$ to all propositional parameters occurring in B (call this set of sentences Φ_0). Let Φ_{n+1} be defined inductively as the union of Φ_n with the set of sentences formed by negating, conjoining, or disjoining any of the sentences in Φ_n, and let Φ be the union of Φ_n for all n. By the truth-conditions of operators at abnormal points (given the acknowledgement condition and the constraint that $w \Vdash A$ iff $w^* \nVdash A$), if $@$ stands in $\Vdash$ to all the sentences in Φ_n, $@$ stands in $\Vdash$ to all the sentences in Φ_{n+1}. Since $@$ stands in $\Vdash$ to everything in Φ_0 and $B \in \Phi$, $@ \Vdash B$, and thus, by the truth-conditions of $\vee$, $@ \Vdash A \vee B$. The $\{B, \phi(A)\}$ case is analogous.
- sc: Priest has a proof that works just as well.[14] To summarise, suppose we have $\Gamma_1 \vdash A \vee B$, $\Gamma_2 \cup \{A\} \vdash C$, and $\Gamma_3 \cup \{B\} \vdash C$. Assume for recursion that $\Gamma_1 \cup \Gamma_2 \cup \Gamma_3 \subseteq \Delta$ and that $\Gamma_1 \vDash A \vee B$, $\Gamma_2 \cup \{A\} \vDash C$, and $\Gamma_3 \cup \{B\} \vDash C$. $\Delta \vDash A \vee B$ (since $\vDash$ is monotonic), so, by the truth-conditions of $\vee$, $\Delta \vDash A$ or $\Delta \vDash B$, from each of which follows $\Delta \vDash C$ (since $\Delta \cup \{A\} \vDash C$ and $\Delta \cup \{B\} \vDash C$). □

Theorem. $\Gamma \vdash_{\mathrm{BI}_G} A$ only if $\Gamma \vDash_{\mathrm{BI}^*} A$. Soundness for **BI**.

Proof. We extend $\mathbf{BNI}_G$ with wxm:

- wxm: $\phi(A) \vDash A \vee \neg A$. Suppose arbitrary M satisfies $\phi(A)$, so $@ \Vdash \phi(A)$. If $@ \notin N$, by definition of $\phi()$, $@$ stands in $\Vdash$ to all propositional parameters occurring in A (this set Φ_0). Construct Φ as in wad. Then since $@$ stands in $\Vdash$ to all the sentences in Φ_0, and $A \in \Phi$ (and thus $\neg A \in \Phi$), we have $@ \Vdash A$ and $@ \Vdash \neg A$ and, by the truth-conditions of $\vee$, $@ \Vdash A \vee \neg A$.

In the case of $@ \in N$, we start with the fact that, since $@ \in G^+$, either $@ \Vdash A$ or $@^* \nVdash A$. If $@ \Vdash A$, $@ \Vdash A \vee \neg A$ follows by the truth-conditions for $\vee$. If $@^* \nVdash A$, $@ \Vdash \neg A$ and so $@ \Vdash A \vee \neg A$. □

Theorem. $\Gamma \vdash_{\mathrm{NI}_G} A$ only if $\Gamma \vDash_{\mathrm{NI}^*} A$. Soundness for **NI**.

Proof. We extend $\mathbf{BNI}_G$ with efq:

- efq: $A \wedge \neg A \vDash B$. Suppose not: then there is some model M such that $@ \Vdash A \wedge \neg A$, but $@ \nVdash B$. By the truth-conditions of $\wedge$, $@ \Vdash A$ and $@ \Vdash \neg A$. But $@ \Vdash \neg A$ only if $@^* \nVdash A$. Since **NI*** models require $w \nVdash A$ or $w^* \Vdash A$ (since $@ \in G^-$), M is not a model. □

Theorem. $\Gamma \vdash_{\mathrm{I}_G} A$ only if $\Gamma \vDash_{\mathrm{I}^*} A$. Soundness for **I**.

Proof. We extend $\mathbf{BNI}_G$ with efq and wxm. □

---

[14] Ibid., pp. 282–83.

For soundness results for **BN**, **B**, **N**, and **Ø**, we would need to check the soundness of the extra rules relative to the relevant models.[15]

4.3 Completeness

Lemma. Henkin construction: $\Gamma \nvdash A$ only if there is some $\Pi \supseteq \Gamma$ such that $\Pi \nvdash A$; $\Pi \vdash B$ only if $B \in \Pi$ (closure); and $\Pi \vdash B \vee C$ only if $\Pi \vdash B$ or $\Pi \vdash C$ (primeness).

Proof. Here I will summarise Priest's proof.[16] The formulae B_i are enumerated and Π (a Henkin theory) is constructed from $\Gamma = \Pi_0$ by taking the union of each Π_i defined as $\Pi_{n+1} = \Pi_n \cup \{B_n\}$ iff $\Pi_n \cup \{B_n\} \nvdash A$, and $\Pi_{n+1} = \Pi_n$ otherwise. $\Pi \nvdash A$ follows from the fact that Π is compact. For closure, suppose for reductio that $\Pi \vdash B_n$ but $B_n \notin \Pi$; but then, by construction, $\Pi_n \cup \{B_n\} \vdash A$, so $\Pi \vdash A$. For primeness, suppose $\Pi \vdash B_n \vee B_m$ but $B_n \notin \Pi$ and $B_m \notin \Pi$. Then $\Pi_n \cup \{B_n\} \vdash A$ and $\Pi_m \cup \{B_m\} \vdash A$, so $\Pi \vdash A$. □

Lemma. Antitheory construction: Let Π be a Henkin theory. There is some Σ extending $\{\neg B \mid B \notin \Pi\}$ such that $\Sigma \nvdash C$ for all C in $\{C \mid \neg C \in \Pi\}$; $\Sigma \vdash D$ only if $D \in \Sigma$ (closure); and $\Sigma \vdash D \vee E$ only if $\Sigma \vdash D$ or $\Sigma \vdash E$ (primeness). Call Σ the *antitheory twin* of Π.

Proof. Enumerate the formulae B_i and construct Σ from $\{\neg C \mid C \notin \Pi\} = \Sigma_0$ as in the Henkin construction but the with inductive definition of Σ_n changed to $\Sigma_{n+1} = \Sigma_n \cup \{B_n\}$ iff for all $D \in \{D \mid \neg D \in \Pi\}$, $\Sigma_n \cup \{B_n\} \nvdash D$, and $\Sigma_{n+1} = \Sigma_n$ otherwise. Closure, primeness, and that $\Sigma \nvdash D$ for all $D \in \{D \mid \neg D \in \Pi\}$ are analogous to the Henkin case. □

Lemma. Let Π be a Henkin theory and Σ its antitheory. Then Π is the antitheory of Σ.

Proof. Let Φ be the antitheory of Σ, so we need to show that $\Phi = \Pi$: $A \in \Pi$ iff $A \in \Phi$. For the left-to-right conditional, suppose $A \in \Pi$: then, by closure via dn, $\neg\neg A \in \Pi$. By the construction of Σ, $\neg A \notin \Sigma$ and hence, by the construction of Φ, $\neg\neg A \in \Phi$, from which we get $A \in \Phi$ by closure via dn. The right-to-left conditional is similar. □

Theorem. $\Gamma \vDash_{\text{BNI}^*} A$ only if $\Gamma \vdash_{\text{BNI}_G} A$. Completeness for **BNI**.

Proof. By contraposition: $\Gamma \nvdash A$ only if $\Gamma \not\vDash A$. Let Π be a Henkin theory extending Γ, and let Σ be the antitheory twin of Π. We then construct a model $M = \langle W, N, G^+, G^-, ^*, \Vdash \rangle$ with $W = \{@, @^*\}$ constrained in the following way:[17]

- $@ \Vdash B$ iff $B \in \Pi$;
- $@^* \Vdash B$ iff $B \in \Sigma$;
- $@ \in N$ iff there is some C such that $\phi(C) \in \Pi$ but $C \notin \Pi$;
- $@^* \in N$ iff there is some C such that $\phi(C) \in \Pi$ but $C \notin \Sigma$;

[15] See ibid., pp. 286–89.
[16] Ibid., p. 283.
[17] For discussion on the size of models, I am grateful to an anonymous reviewer.

We now need to check that M really is a model. If so, M is a model satisfying Γ but not A, so $\Gamma \nvDash A$ (as wanted). Specifically, we need to show that that $\Vdash$ satisfies the truth-conditions of the connectives. We do so by recursion on the truth-conditions of $\wedge$ and $\neg$, with $A \vee B$ equivalent to $\neg(\neg A \wedge \neg B)$.

The base case is trivial, since $\Vdash$ imposes no constraints on the assignment of propositional parameters. The connectives need to satisfy:

- $w \Vdash A \wedge B$ iff $w \Vdash A$ and $w \Vdash B$
- $w \Vdash \neg A$ iff $w^* \nVdash A$

In the following I use @ and Π (matters are similar for @* and Σ).

- $\wedge$: $@ \Vdash A \wedge B$ iff $@ \Vdash A$ and $@ \Vdash B$ – viz. $A \wedge B \in \Pi$ iff $A \in \Pi$ and $B \in \Pi$. Left-to-right is by closure via s. Right-to-left is by closure via adj.
- $\neg$: $@ \Vdash \neg A$ iff $@^* \nVdash A$ – viz. $\neg A \in \Pi$ iff $A \notin \Sigma$. Left-to-right is from the construction of Σ by closure: since $\Sigma \nvdash A$ for $\{A \mid \neg A \in \Pi\}$, each such $A \notin \Sigma$. Right-to-left is by the fact that Π and Σ are antitheory twins.

Since $A \vee B$ is equivalent to $\neg(\neg A \wedge \neg B)$, we can define $\vee$ in terms of $\wedge$ and $\neg$, meaning there is no need for a separate case for $\vee$. The deductions are simple.[18] Since the rules are sound, they show semantic equivalence too. □

Theorem. $\Gamma \vDash_{\mathrm{BI}^*} A$ only if $\Gamma \vdash_{\mathrm{BI}_G} A$. Completeness for **BI**.

Proof. For wxm, we add the constraint to M that $@ \in G^+$. The truth conditions are unaffected. □

Theorem. $\Gamma \vDash_{\mathrm{NI}^*} A$ only if $\Gamma \vdash_{\mathrm{NI}_G} A$. Completeness for **NI**.

Proof. For efq, we add the constraint to M that $@ \in G^-$. The truth conditions are unaffected. □

Theorem. $\Gamma \vDash_{\mathrm{I}^*} A$ only if $\Gamma \vdash_{\mathrm{I}_G} A$. Completeness for **I**.

Proof. For both wxm and efq, we add both constraints, so $@ \in G^+ \cap G^-$. □

---

[18]

$$\dfrac{\dfrac{\neg(\neg A \wedge \neg B) \qquad \dfrac{[\neg\neg A]^1}{A} \qquad \dfrac{[\neg\neg B]^2}{B}}{A \vee B}\,{}_{1,2}}{\neg\neg A \vee \neg\neg B}$$

$$\dfrac{\dfrac{A \vee B \qquad \dfrac{\dfrac{[A]^1}{\neg\neg A}}{\neg\neg A \vee \neg\neg B} \qquad \dfrac{\dfrac{[B]^2}{\neg\neg B}}{\neg\neg A \vee \neg\neg B}}{\neg\neg A \vee \neg\neg B}\,{}_{1,2}}{\neg(\neg A \wedge \neg B)}$$

5. From Pure to Applied Semantics

5.1 What Does it all Mean?

Up until now, we have been broadly concerned with the 'pure' semantics of the FDE family and of infectiousness. Let us now turn to the 'applied' semantics. A pertinent motivation for considering this is given by reflecting on what the semantics mean – *really mean*. Johan van Benthem puts this concern bluntly for star semantics when he says: "Pending further explanation of the nature of [the * operator], one cannot even begin to say if [star semantics] is more than just a formal trick".[19]

What is at stake in the choice of semantics? Briefly, one important difference between the many-valued and star semantics concerns what is the most natural interpretation of negation. On the star semantics, it is natural to think of negation as an intensional *exclusion operator*: the fact that $\neg A$ holds at w is grounded in the fact that A fails at w^*, w^* being the point ('world', with all its metaphysical import, is perhaps appropriate here) recording what is *compatible* with w.[20] The many-valued semantics would seem to retain the classical interpretation of negation as the operator that makes the (at least) true (at least) false and the (at least) false (at least) true, but admits as logical possibilities those cases where sentences are both true and false and those where sentences are neither true nor false (and the strange cases, too).

Detailed discussion of each of these interpretations is outside the scope of this paper, but I would like to take the opportunity to briefly gesture in the direction of negation as an exclusion operator, which I hope will frame discussion in the rest of this section. We first note that the two interpretations are equivalent in terms of when something is true/false, at least when we think of falsity as truth of negation (which is widely held).[21] The consideration in favour of negation as exclusion is then the fact that this interpretation *explains why* falsity is truth of negation, whereas the classical interpretation says nothing beyond that falsity is truth of negation, and is thus open to the charge of ad hocery. A is false at w when $\neg A$ is true at w because $w \Vdash \neg A$ means $w^* \nVdash A$ – that is, A fails to be compatible with the way things are at w, so it is false.

So, the star semantics has something going for it. But how much has it going for it? Let us consider some interpretations of the i value, and try to make sense of them in terms of the semantics presented in this paper.

Three of the candidates for an interpretation of logics containing i are:

- the *nonsense* interpretation,
- the *off-topic* interpretation, and
- the *emptiness* interpretation.

---

[19] Van Benthem 1979, p. 341.
[20] Meyer & Martin 1986, pp. 306–10.
[21] Ibid., p. 308, do not hold this, instead thinking of falsity as failure at a point.

5.2 Nonsense

Perhaps the most prominent interpretation of *i* is due to D. A. Bochvar.[22] On the *nonsense* interpretation, a sentence assigned *i* is meaningless or senseless. This senselessness, it is thought, is inherited by any sentence in which it occurs. Let us take the liar sentence ('this sentence is false') as our candidate for the bearer of our infectious value. Infectiousness means the following are meaningless:

- 'It is not the case that this sentence is false.'
- 'Hillary climbed Everest and this sentence is false.'
- 'Hillary climbed Everest or this sentence is false.'

How is this thought extended to the star semantics? A natural thought is as follows. To fail at one of the abnormal points is to be senseless – it is just like being *i*. When we are in a context in which the meaningfulness of our expressions is guaranteed, then, we can constrain our models to those in which @ is normal. When our expressions may lack meaning/sense, @ may be abnormal.

But what grounds have we to say that this nonsense value is infectious? In the case of paradoxes like the liar, this is far from obvious. Bochvar takes these paradoxes as paradigm examples of meaningless sentences (the logic of his concern being **I**),[23] but in **N** or weaker or **B** or weaker, *indeterminate* does not seem the most obvious assignment. The liar sentence looks like it should be both true and false, and its twin ('this sentence is true') looks like it should be neither true nor false, and the thought that those things which walk and quack like ducks are probably ducks is a compelling one.[24]

Even if the liar and related puzzles are not genuinely antinomous in the sense *indeterminate* requires, there may well be other candidates, but it is not altogether obvious what these would be. This would seem to leave *i*-variant infectious logics insufficiently motivated with respect to their *i*-free uninfectious twins, offering little in the way of explanatory resources we did not already have – that pigs *could*, for all we know, fly is not a strong reason to revise our folk theory of porcine aviation. (But a flying pig is.)

The more underlying worry here is that if we cannot sensibly talk about something (in the way required for infectiousness to apply), why are there sentences in our language about it? Why can we conjoin, disjoin, and negate it? We leave this thought for now, and shall return to it later.

[22] Bochvar & Bergmann 1981.
[23] Ibid., pp. 105–07.
[24] Beall 2018, pp. 48–49.

5.3 Incongruity

The *off-topic* interpretation is a more recent suggestion of Jc Beall's, formulated in response to weaknesses in the *nonsense* interpretation.[25] On this account, the truth-values go:

- *t*: true (and not false), and on-topic;
- *f*: false (and not true), and on-topic;
- *b*: both true and false, and on-topic;
- *n*: neither true nor false, and on-topic;
- *i*: off-topic.

Suppose we have it that Marmite is tasty. Does it follow that Marmite is tasty or Wellington is in New Zealand? In the *i*-variants, no. But why not? On the *off-topic* interpretation, while *truth* is preserved over the inference, *topic* is not. The topic of 'Marmite is tasty' is *Marmite* or *tasty food* or similar, whereas the topic of 'Marmite is tasty or Wellington is in New Zealand' is some compound or product of the topics of the disjuncts.

The application to the star semantics is analogous to the *nonsense* interpretation case. The normal points are ones where everything is on-topic, and the abnormal points ones where things may be off-topic. When @ is constrained to the normal points, we are guaranteed to be on-topic; when it is not, we are not.

On such an interpretation, it is natural to think of points (normal points, at any rate) as *theories* concerning some topic, and classes of points as collections of theories concerning that topic. We might add the qualification that a theory may not be exhaustive with respect to its topic, so let us concern ourselves only with exhaustive theories.[26]

It is then natural to think of a model as an *exhaustive theoretical position* with respect to some topic: @ is the correct theory of the topic according to that position, other normal points are rival theories concerning the topic, regarded by that position as incorrect, and abnormal points are theories which are off-topic by the lights of that position.

Let us consider the topic that encompasses everything. There will be a class of exhaustive theoretical positions (models) concerned with such a topic, and among such a class of exhaustive theoretical positions, there will be a correct one – the *theory of everything*. @ will be the way that everything is, and other normal points will be ways that everything is not. But what are we to say about abnormal points?

---

[25] Beall 2016.

[26] What to say about inexhaustive theories on this sort of account is not totally clear. We could have it such that claims w does not decide A are such that $w \nvDash A$ and $w^* \nvDash A$ (neither true nor false), but then it is unclear how to draw the distinction between matters a theory makes no decision on, and matters a theory holds are underdetermined or otherwise neither true nor false.

When faced with such a question, two options seem salient. The first is to say that such a model will have no abnormal points, since it concerns everything, and everything means *everything*. Such an answer would yield that the correct logic, insofar as logic is concerned with 'absolute generality' or 'universal closure', is **BN** (FDE) or one of its extensions – at any rate, not an infectious logic. Indeed, Beall has made an argument along such lines for **BN**.[27] (Note that this would not make infectious logics useless, since we are nearly always concerned with less than everything, so things might still be off-topic with respect to what we are interested in.)

The second is to countenance the idea that there are certain matters which are off-topic with respect to everything. This is difficult to get one's head around, but it is to be expected that the theory of everything seems strange in certain respects. But even stranger is the idea that @ could be abnormal (as the *i*-variant models allow). To these thoughts we shall return.

5.4 Emptiness

It has been suggested that the *catuṣkoṭi* ('four corners' or tetralemma) of classical Buddhist philosophy corresponds to **BN**.[28] The *catuṣkoṭi*'s exclusive corners (and their **BN** counterparts) are:

Corner	**BN**+	**BN***
being A	$m(A) = t$	$@ \Vdash A$, $@^* \Vdash A$;
not being A	$m(A) = f$	$@ \nVdash A$, $@^* \nVdash A$;
both being and not being A	$m(A) = b$	$@ \Vdash A$, $@^* \nVdash A$;
neither being nor not being A	$m(A) = n$	$@ \nVdash A$, $@^* \Vdash A$

But Nāgārjuna, founder of the Madhyamaka school, sometimes rejects all the corners – the *fourfold negation*:

> Having passed into nirvana, the Victorious Conqueror
> Is neither said to be existent
> Nor said to be nonexistent.
> Neither both nor neither are said.[29]

Garfield and Priest analyse this as demanding another truth-value – our *i*, yielding **BNI**+ – to formally capture the Madhyamaka concept of *śūnyatā* (emptiness, the absence of *svabhāva* or essence).[30] Madhyamaka metaphysics holds that *ultimate reality* (linked to *ultimate truth*) exhibits emptiness in this sense – everything is grounded (in a certain sense) in other things. *Conventional reality* (linked to *conventional truth*) is not empty, since we speak and think of things as having essence. The picture is nihilistic

[27] Beall 2018, 2019.
[28] Priest 2010.
[29] Nāgārjuna & Garfield 1995, §21.17.
[30] Garfield & Priest 2009; Priest 2010, §4.

with respect to the ultimate truth of our views while allowing them some sort of (conventional) truth, forming a *middle way* (three guesses what 'Madhyamaka' translates to). So the trick, roughly, on the many-valued semantics is to assign *i* when we are speaking of ultimate reality, which defies theorisation, and assign the normal four truth values when we are talking about conventional matters.[31]

The story here for the star semantics is similar, too. Normal points are linked to effable conventional reality, and abnormal points to ineffable ultimate reality. A sentence evaluated at a normal point concerns conventional reality, and, at an abnormal point, ultimate reality. Since ultimate reality is ineffable, abnormal points will not satisfy sentences said of it. Hence infectiousness is motivated. So, when we allow ourselves to speak of ultimate reality, @ may be abnormal.

5.5 Disjunction and its Simulacra

Hitoshi Omori and Damian Szmuc have argued that one interesting feature of infectious logics (when they are given a plurivalent semantics) is this: while their conjunction operator does capture genuine conjunction, their disjunction operator *does not* capture genuine disjunction.[32] To see the force of this claim, consider the fact that $A \wedge B$ is at least true iff A is at least true and B is at least true, but it is not necessarily the case that $A \vee B$ is at least true iff A is at least true or B is at least true, since one of A, B could be at least true, while the other is infectiously untrue – in which case $A \vee B$ would not be at least true.

That the disjunction operator does not capture genuine disjunction – does not respect the truth-conditions of disjunction – seems even more stark on the star semantics, since the truth-conditions for $A \vee B$ at abnormal points are

- $w \Vdash A \vee B$ iff $w \Vdash A$ and $w \Vdash B$,

equivalent to to those of $A \wedge B$. To put it baldly, disjunctions at abnormal points are effectively conjunctions:

- $w \Vdash A \vee B$ iff $w \Vdash A \wedge B$.[33]

5.6 Putting These Thoughts Together

We saw that infectiousness (and that to which it applies), on the pure semantics given in this paper, can be interpreted as capturing some sort of meaninglessness or senselessness (on the nonsense interpretation), falling outside the scope of the most general topic (on the off-topic interpretation), and taking a view on ineffable ultimate

---

[31] Priest thinks the picture is ultimately more complicated, and we end up with a five-valued plurivalent logic (2010, §5).

[32] Omori & Szmuc 2017, pp. 279–81. For discussion of plurivalence, see Priest 2014.

[33] One might wonder whether something funny is going on with negation, too. Omori and Szmuc 2017 point out that it satisfies the $\neg A$ is at least true/false iff A is at least false/true condition, and it satisfies $w \Vdash \neg A$ iff $w^* \nVdash A$ too. So all seems fine here.

reality (on the emptiness interpretation). But we were left with two headscratchers: what are the candidates for senseless sentences, if not the traditional paradoxes, and how are we to understand abnormal points in our theory of everything?

These thoughts, which seemed puzzling in isolation, seem to fit together now. This is to say that the three interpretations considered, on the star semantics, seem to converge: What is senseless? Answer: that which is off-topic with respect to the absolutely general topic according to the correct absolutely general exhaustive theoretical position. But what could *that* be? Answer: ineffable ultimate reality. (According to the real world, on such a picture, @ would be abnormal.)

What this would seem to suggest is that there is something in this picture, and infectious logics are modelling something interesting. Three broken clocks do not often agree on the time.

However, interesting as that which is infectious seems, so too does there just seem to be something wrong with it – if something is infectious, you probably do not want it. This intuition would seem vindicated by the first option we considered about what to say about abnormal points with respect to the theory of everything, and by Omori and Szmuc's worries about disjunction. These thoughts fit together too: disjunction in infectious logic is not the right account of disjunction because infectious logic is not the right account of logic. Logic is interested in relations between sentences in true theories about some topic or other (or perhaps about the absolutely general topic in particular, depending on how we think of topic-neutrality).[34]

Here we find ourselves back with the distinction between many-valued and star semantics, for an interesting difference here emerges in terms of what it might be appropriate to call their *quarantine strategies*. How do we keep this abnormal infectious stuff from tearing down the logical edifice (and everything else with it)? In less dramatic language, how do we *quarantine* such a pathosis?

The star semantics, in a sense, handles quarantine all by itself. As one will recall, points in a star model for an infectious logic are split into the normal points, where disjunction is disjunction and all is well with the world, and the abnormal points, where things get quite strange. Mathematicians like to describe this sort of distinction as that between the *well-behaved* and the *pathological*, the latter of which seems particularly appropriate terminology to describe infectious logics.

In the star semantics, a sharp line is drawn between well-behaved points and pathological points – the abnormality is confined to $W - N$. In the many-valued semantics, however, models lack such a structural difference, and sentences assigned i are treated like everything else – they pathologise the whole model, so to speak. Quarantining, then, must be done at the level of models. (Maybe this difference counts as some sort of reason in favour of a star semantical treatment of infectiousness, and,

[34] For an account of logic along these lines, see Beall 2018, 2019.

by extension, negation as an exclusion operator, or perhaps it is just an interesting observation.)

6. Conclusion

In this paper, I have extended star semantics to the infectious logics in the FDE family, mirroring the existing many-valued semantics. Discussion of the interpretation of infectiousness started with the idea that both what infectiousness is and what is infectious seem of significant philosophical interest but also very difficult to pin down, and came to rest on the idea that three prominent interpretations of infectious logics (when adapted to match the star semantics) seem to converge on taking infectiousness to concern something like ineffability. There's gold in them thar hills, but there aren't really any hills.[35]

[35] My thanks are due to three anonymous *UPJA* reviewers, whose comments have proved invaluable.

References

Asenjo, F. G. (1966). 'A Calculus for Antinomies.' *Notre Dame Journal of Formal Logic*, *7*(1), pp. 103–05. https://doi.org/10.1305/ndjfl/1093958482.

Beall, J. (2009). *Spandrels of Truth*. Oxford: Oxford University Press.

Beall, J. (2016). 'Off-Topic: A New Interpretation of Weak-Kleene Logic.' *The Australasian Journal of Logic*, *13*(6), pp. 136–42. https://doi.org/10.26686/ajl.v13i6.3976.

Beall, J. (2017). 'There is no Logical Negation: True, False, Both, and Neither.' *The Australasian Journal of Logic*, *14*(1), pp. 1–29. https://doi.org/10.26686/ajl.v14i1.4025.

Beall, J. (2018). 'The Simple Argument for Subclassical Logic.' *Philosophical Issues*, *28*(1), pp. 30–54. https://doi.org/10.1111/phis.12133.

Beall, J. (2019). 'FDE as the One True Logic.' In *New Essays on Belnap-Dunn Logic*, edited by H. Omori and H. Wansing, pp. 115–25. Cham: Springer International Publishing.

Belnap, N. D. (2019a). 'A Useful Four-Valued Logic.' In *New Essays on Belnap-Dunn Logic*, edited by H. Omori and H. Wansing, pp. 55–76. Cham: Springer International Publishing.

Belnap, N. D. (2019b). 'How a Computer Should Think.' In *New Essays on Belnap-Dunn Logic*, edited by H. Omori and H. Wansing, pp. 35–53. Cham: Springer International Publishing.

Bochvar, D. A., and Bergmann, M. (1981). 'On a Three-valued Logical Calculus and Its Application to the Analysis of the Paradoxes of the Classical Extended Functional Calculus.' *History and Philosophy of Logic*, *2*(1–2), pp. 87–112. https://doi.org/10.1080/01445348108837023.

Dunn, J. M. (1976). 'Intuitive Semantics for First-Degree Entailments and "Coupled Trees".' *Philosophical Studies*, *29*(3), pp. 149–68. https://www.jstor.org/stable/4319017.

Field, H. (2008). *Saving Truth from Paradox*. Oxford: Oxford University Press.

Garfield, J. L., and Priest, G. (2009). 'Mountains are Just Mountains.' In *Pointing at the Moon: Buddhism, Logic, Analytic Philosophy*, edited by Jay L. Garfield, T. J. F. Tillemans, and M. D'Amato, pp. 71–82. Oxford: Oxford University Press.

Meyer, R. K., and Martin, E. P. (1986). 'Logic on the Australian Plan.' *Journal of Philosophical Logic*, *15*(3), pp. 305–32. https://www.jstor.org/stable/30226357.

Nāgārjuna, and Garfield, J. L. (1995). *The Fundamental Wisdom of the Middle Way: Nāgārjuna's Mūlamadhyamakakārikā*. Oxford: Oxford University Press.

Omori, H., and Szmuc, D. (2017). 'Conjunction and Disjunction in Infectious Logics.' In *Logic, Rationality, and Interaction*, edited by A. Batlag, J. Seligman, and T. Yamada, pp. 268–83. Berlin/Heidelberg: Springer Berlin Heidelberg.

Omori, H., and Wansing, H. (2017). '40 years of FDE: An Introductory Overview.' S*tudia Logica*, *105*(6), pp. 1021–49. https://doi.org/10.1007/s11225-017-9748-6.

Priest, G. (1979). 'The Logic of Paradox.' *Journal of Philosophical Logic*, *8*(1), pp. 219–41. https://doi.org/10.1007/BF00258428.

Priest, G. (2008). *An Introduction to Non-classical Logic: From If to Is* (2nd ed.). Cambridge: Cambridge University Press.

Priest, G. (2010). 'The Logic of the *Catuskoti*.' *Comparative Philosophy*, *1*(2), pp. 24–54. https://scholarworks.sjsu.edu/cgi/viewcontent.cgi?article=1032&context=comparativephilosophy.

Priest, G. (2014). 'Plurivalent Logics.' *The Australasian Journal of Logic*, *11*(1), pp. 2–13. https://doi.org/10.26686/ajl.v11i1.1830.

Priest, G. (2019). 'Natural Deduction Systems for Logics in the FDE Family.' In *New Essays on Belnap-Dunn Logic*, edited by H. Omori and H. Wansing, pp. 279–92. Cham: Springer International Publishing.

Routley, R., and Routley, V. (1972). 'The Semantics of First Degree Entailment.' *Noûs*, *6*(4), pp. 335–59. https://doi.org/10.2307/2214309.

van Benthem, J. F. A. K. (1979). 'What is Dialectical Logic?' *Erkenntnis*, *14*(3), pp. 333–47. https://doi.org/10.1007/BF00174898.

Founded in 2019, the *Undergraduate Philosophy Journal of Australasia* (UPJA) is the first undergraduate philosophy journal run by students from the Australasian region. We publish two issues and host two virtual conferences annually: one mid-year and another at the end of the year. Our calls for papers for each issue open roughly in February and August, respectively. We aim to be an inclusive and diverse journal that welcomes submissions on any philosophical topic attempting to make a substantive contribution to contemporary debate. Submissions from women and other members of underrepresented groups in philosophy, including those for whom English is not their first language, are particularly encouraged.