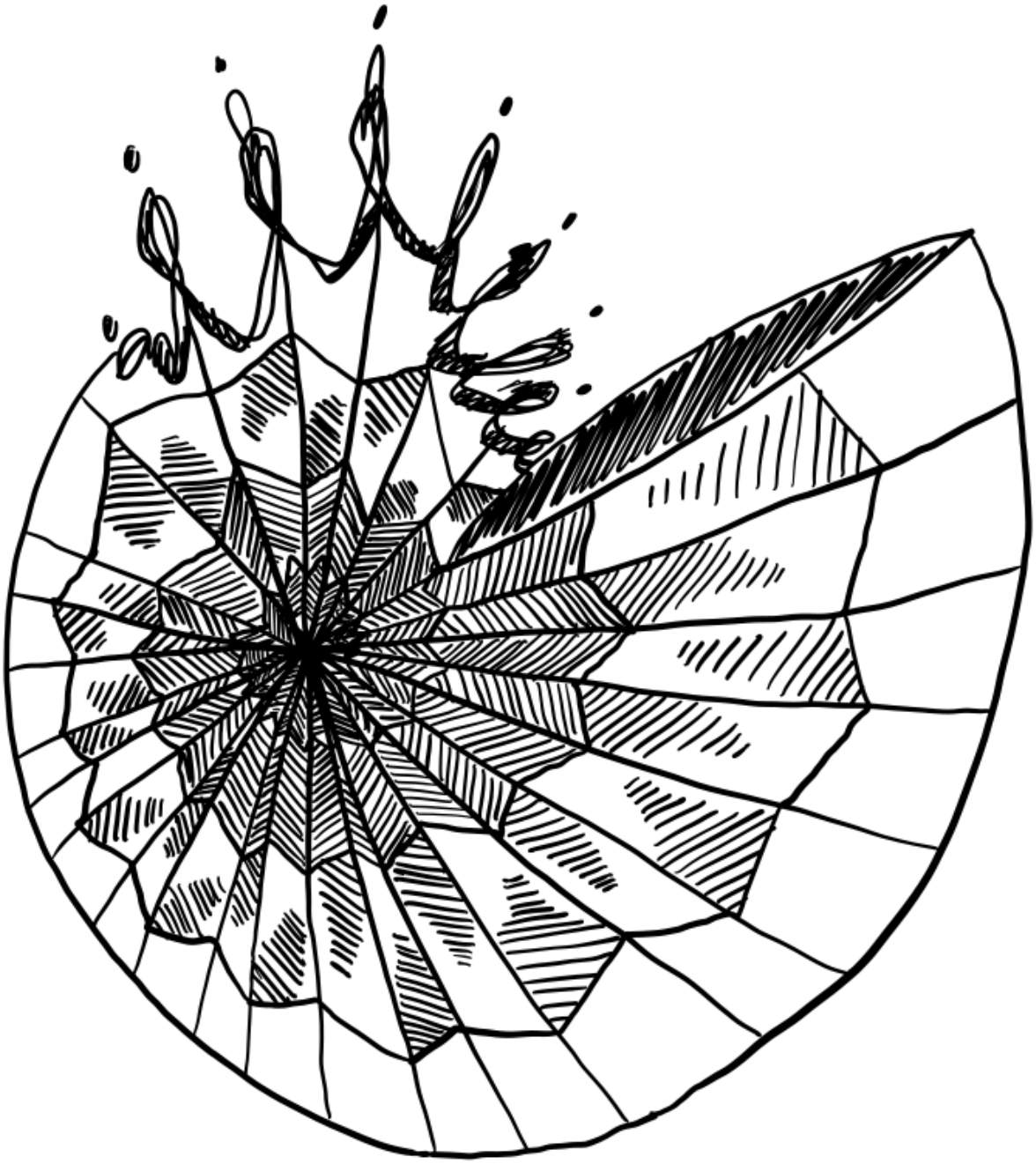
Aporia

XXV



Aporia

Undergraduate Journal of the St Andrews Philosophy Society

VOLUME XXV

Aporia is funded by the University of St Andrews Philosophy Society, which receives funds from the University of St Andrews Department of Philosophy, the University of St Andrews Students' Association, and independent benefactors.

Foreword

Congratulations to all those involved with *Aporia* for reaching the milestone of a twenty-fifth volume! I was a first-year undergraduate when the first issue was released. I look back at that day fondly. We celebrated with a wine reception in Edgecliffe (though, to be entirely honest, we didn't need much of an occasion to indulge in some wine!), and there were speeches from Vera Schoeller (the society president) and Professor Sarah Broadie, who was always a vocal supporter of PhilSoc.

The journal has also published some fantastic work over the years, including by authors who have gone on to be very successful professional philosophers. The very first issue included pieces by Marcus Rossberg and Philip Ebert (now professors at Connecticut and Stirling, respectively) and a special contribution from Duncan Pritchard (now UC Irvine). Later issues included Andreas Stokke (Uppsala), Fenner Tanswell (Technical University Berlin), Steffen Koch (Bielefeld) and Michael Hicks who is currently my colleague at the University of Glasgow. I was a co-editor (with Kyle Mitchell) for volumes II and III. I was also tasked with designing and formatting (which, unfortunately, was not my forte). Since my time at the helm, I am pleased see that the high quality of philosophy has continued, and the design quality has significantly improved. The journal is now an extremely professional-looking production, and something for all involved to be proud of.

The St Andrews Philosophy Society will always hold a very dear place in my heart. It was there that I got to know many people who are my friends until this day. It is wonderful that the society is still going strong, and that the journal continues to be a roaring success.

Dr Joe Slater
Lecturer in Moral and Political Philosophy, University of Glasgow
St Andrews Philosophy Society President (2011/12)

Letter from the Editor

Dear Reader,

It is with pleasure that I present the 25th edition of *Aporia*, the undergraduate philosophy journal of St Andrews. I hope it shall prove a stimulating and enjoyable read.

My particular thanks, firstly, to Joe Bradstreet, my deputy, who has been a well-informed and clever editor, a diligent administrator, and a pleasant balance to all my neuroticisms. My thanks also to Christina Landys Herre, who bravely shouldered much of the administrative burden. And I cannot fail to mention Mohit Agarwal, whose technical prowess and willingness to take on tasks far beyond my capabilities have been utterly invaluable. I am extremely grateful to the whole editorial team for their dedication, expertise, and good humour; I should have been quite at sea with many of the papers were it not for their excellent work.

I am often teasingly asked about our readership; I think I may honestly answer that the people who truly gain the most from *Aporia* are the editors and writers. From my editorship I have been exposed to areas of philosophy untouched by my courses, learned more about the rights (and perhaps more frequently wrongs) of academic writing than any lecturer has taught me, and acquired the skill of working with a relatively large and diverse team. For all the vexations of the position, I am pleased to have had the opportunity of dipping my toe into the murky waters of academic research and publishing.

It is a real triumph that *Aporia* has survived 18 years of disorganized philosophy students, incomprehensible technological advancements, interrupted teaching, and unreliable funding. Let us hope it shall long continue for the next generations of budding philosophers.

Yours faithfully,

Kirsty Graham
Editor-in-Chief, *Aporia*

Acknowledgements

Editors

Aliza Ashraf

Beth Cook

Maria de Feo

Ella Johnston

Hanaa Khan

Zack Ledesma

Christina Landys Herre

Nawal Mirza

Claire Mizrahi

Phoebe Ray

Nathalie Rogers

Rosa Velasco Saavedra

Jacob Walchuk

Hoochang Yi

Kirsty Graham
Editor-in-Chief

Joe Bradstreet
Deputy Editor-in-Chief

Mohit Agarwal
Technical Officer

Sofia Mona
Cover Artist

Contents

1 **Selfish Comparative Optimism: A Rejoinder to Nagasawa's *Problem of Evil for Atheists*** 1

Wilson Sugeng, *University of St Andrews* . . . . . .

2 **Supervaluationism, Dynamic Supervaluationism, and Higher-Order Vagueness** 11

Wiktor Przyborowski, *University of St Andrews* . . . . . .

3 **Defending Williamson's Explanatory Challenge to Contingentism** 27

Koda Li, *Brown University* . . . . . .

4 **A Defence of the Interpretational Account of Validity** 39

Audrey Hammer, *University of Cambridge* . . . . . .

Contributors 49

Selfish Comparative Optimism: A Rejoinder to Nagasawa's *Problem of Evil for Atheists*

Wilson Sugeng, *University of St Andrews*

Yujin Nagasawa's problem of systemic evil (POSE) argues that systemic evils like natural selection pose a greater challenge to atheism/non-theism than to theism, as they conflict with "modest optimism": the view that the world is fundamentally "not bad." Nagasawa suggests theism resolves this by appealing to a heavenly bliss, offsetting natural evils, a strategy unavailable to atheists/non-theists. However, I argue that atheists/non-theists are better equipped to address POSE because they are not constrained by the theistic commitment to a categorically good world.

In Section 1, I critique two theistic approaches to POSE. Extreme optimism defends the actual world as the best possible one, requiring problematic justifications such as free-will and "only-way" theodicies to explain systemic evils as necessary. Neutral optimism, while allowing for multiple good worlds, still struggles to reconcile systemic evils with a benevolent God, merely shifting the problem to other possible worlds.

In Section 2, I explore how atheists/non-theists can bypass POSE. They can adopt personal, rather than cosmic, optimism, valuing their own existence without affirming the world's overall goodness. Alternatively, they can embrace comparative optimism, viewing existence as better than non-existence without attributing intrinsic value to natural processes like evolution. These flexible approaches free non-theists from the philosophical burdens tied to systemic evils.

In Section 3, I argue that even if POSE persists, atheists/non-theists can "borrow" theists' theodicies without committing to their metaphysical assumptions. By adopting naturalistic or subjective frameworks, non-theists can justify their modest optimism without the theological constraints imposed by theism. This demonstrates that POSE ultimately challenges theistic frameworks more than atheistic ones.

Introduction

In *The Problem of Evil for Atheists,* Yujin Nagasawa develops a problem of systemic evil (POSE) that he claims challenges both atheists/non-theists and theists alike.[1] He identifies a tension between two widely held theses:

(1) Systemic evil: The process of natural selection necessitates significant suffering and pain for countless sentient animals.

(2) Modest optimism: Overall and fundamentally, the environment in which we exist is not bad.[2]

While theists naturally affirm modest optimism due to their belief in a benevolent creator God, Nagasawa observes that atheists/non-theists are also generally grateful for their existence.[3] For instance, popular atheist Richard Dawkins suggests that contemplation of the law-like evolutionary processes behind our existence

[1] When I say, "God" and "Theism" in this paper, I assume an omniscient, omnipotent, and omnibenevolent singular/simple creator.

[2] Yujin Nagasawa, *The Problem of Evil for Atheists* (Oxford University Press, 2024), 133, 140.

[3] Nagasawa, *The Problem of Evil for Atheists*, 161.

puts us "in a position to give thanks for our luck in being here"—not a gratitude directed towards any agent or being, but rather a "gratitude in a vacuum."[4] Nagasawa sees this as inconsistent: expressing existential gratitude without acknowledging the systemic evils underpinning it implies a tacit endorsement of these evils.

To illustrate this tension, Nagasawa adapts Janna Thompson's apology paradox, which holds that regretting an unjust historical event can be problematic if one's existence depends on that event. For example, a Jew whose grandparents met during the Holocaust faces a paradox: to regret the Holocaust may seem to imply regretting her own existence.[5] Thompson resolves this by distinguishing between regretting *how* one came to exist and *that* one exists—the Jew can regret *how* her grandparents met, without regretting *that* they met at all.[6] Applied to POSE, this seems to suggest that one can regret the mechanisms of natural selection without regretting the outcome of our existence.

However, Nagasawa argues that this resolution fails in the context of POSE. Unlike historical events, natural selection is not a contingent circumstance but a fundamental feature of the natural world.[7] To reject it is not to regret a particular pathway to existence, but to undermine the very conditions that make existence possible. That is, there is no possible world where natural selection does not govern nature and beings like us still exist.

Theists, Nagasawa argues, are better positioned to defend modest optimism, drawing on "heavenly bliss" theodicies that justify or outweigh earthly suffering with the promise of an afterlife. These come in two forms: (1) as a deferred justification, where evolution is acceptable because it leads to eternal reward, and (2) as a utilitarian offset, where infinite heavenly bliss outweighs finite worldly suffering. Because atheists cannot appeal to such concepts, POSE, he claims, presents a more serious problem for atheists.

Contrary to Nagasawa, I argue that atheists and non-theists are better positioned to address POSE because they are not constrained by the theistic requirement to see the world as overall categorically good. To support this claim, I first critique two theistic attempts at resolving systemic evil, namely extreme and neutral optimism, illustrating their shortcomings. Subsequently, I explore how atheists/non-theists might effectively sidestep POSE by adopting personal rather than cosmic optimism, or by embracing a comparative optimism which sees existence as preferable to non-existence without categorically endorsing the systems that facilitated it. Finally, I turn Nagasawa's borrowing argument around to propose that, even if POSE remains challenging, atheists/non-theists can strategically adopt theistic theodicies without their accompanying metaphysical assumptions, thereby reducing POSE's impact and revealing it to be ultimately a greater challenge for theistic frameworks than for atheistic or non-theistic ones.

Section 1: Two Theist Modest Optimists

1.1 Extreme optimism

The first theist modest optimists—extreme optimists—claim that because God actualised the best among all possible worlds, systemic evil must necessarily exist in all good worlds. Although Gottfried Wilhelm Leibniz does not himself discuss systemic evil and predates evolution, his *Theodicy* (1710) presents a system where given God's omnibenevolence and omniscience—if a possible world is better than the actual, then God would either not be good enough to desire the best for the world, or ignorant in not knowing which world is the best.[8]

As an implication, extreme optimists must affirm Nagasawa's claim that no possible world exists in which natural selection does not govern nature; for if God is necessary, then no other world is possible. Natural selection must therefore serve an instrumental role in the world's goodness. Building on this system, Austin Farrer argues that the removal of any such purported evil systems will undermine God's mechanism for bringing about the best world. The goodness of a physical system, for instance, inherently includes

[4] Richard Dawkins, "The Greatest Show on Earth Live" (lecture, University of Auckland, Auckland, New Zealand, 13 March 2010).

[5] Janna Thompson, "The Apology Paradox," *The Philosophical Quarterly* 50, No. 201 (2000): 471.

[6] Janna Thompson, "The Apology Paradox," *The Philosophical Quarterly* 50, No. 201 (2000): 475.

[7] Nagasawa, *The Problem of Evil for Atheists*, 167.

[8] G. W. Leibniz, *Theodicy,* edited by Austin Farrer, translated by E. M. Huggard (Open Court Publishing Company, 1985), 249.

the potential for mutual interference, leading to evils like predation. Without this interference—if this world were a "magically self-arranged garden" free of competition for space or resources—physicality itself ceases to exist.[9] Removing such systems would be akin to relieving an animal's pain "by the removal of its nervous system; that is to say, of its animality."[10] Regretting natural selection thus implicitly challenges God's rationality and goodness in creating us as physical beings rather than spiritual entities.[11]

An immediate difficulty with extreme optimism is that claiming this world to be the best possible one is hard to reconcile with the presence of seemingly avoidable evils observed throughout nature. This tension is captured ironically in the eponymous character of Voltaire's *Candide* (1759) who insists that this is the best possible world as he faces a world plagued with wars, earthquakes, and slavery.[12] Or when Darwin questions why God permitted the creation of the Ichneumonidae who brutally feeds inside the living bodies of caterpillars.[13] This presents a major challenge: extreme optimism struggles to align with observable, avoidable evils unless it denies these empirical observations—as some Creationists do—or reinterprets such systemic evils as necessary.[14]

Granting natural selection's empirical truth, theists generally present two kinds of theodicies for *why* God actualised natural selection. Firstly, theists have adapted the free-will theodicy to address some non-agential non-human suffering. In traditional free-will theodicies, God permits agents the capacity to choose evil over good as the goodness of human agency outweighs the risks of their choosing evil. In one adaptation, Richard Swinburne argues that animal pain and suffering exists as examples of evil actions humans can inflict on each other. Predation therefore exists as an educational tool for humans to observe and understand how to commit evil, thereby enabling their capacity for moral choice.[15]

Secondly, theists have adapted a variation of the soul-making theodicy known as the "only-way" theodicy, arguing that certain natural goods can only develop through natural selection. Holmes Rolston observes that the predator-prey cycle is instrumental to the beautiful diversity of animals, where "The cougar's fang has carved the limbs of the fleet-footed deer, and vice versa."[16] While Young-Earth Creationism may have created this diversity instantaneously, Christopher Southgate argues that natural selection is the only way creatures can develop into biological "selves" with their own interests and behaviours.[17] This offsets any evolutionary evils for it culminates into complex "selves" that conform to God's image.[18] This "selving" must come independently, for Peter van Inwagen argues that an irregular world is a defect: God who constantly intervenes and violates his own laws is either a irrational or evil.[19] So, common to both free-will and "only-way" theodicies is a notion that some ultimate good offsets the evils of natural selection as an instrument.

However, these two theodicies only defer the problem of evil to another system underlying the challenged system. For instance, free-will theodicies must still address Pierre Bayle's objection: If God's omniscience foresees that giving humanity free will inevitably results in unrighteousness, then God is either reckless or cruel to "gift" humanity agency, knowing it would lead to their harm and judgment under his wrath.[20] Echoing Bayle, Robert John Russell questions, "Why did God choose to create *this* universe with *these* laws of physics knowing that they would not only make Darwinian evolution unavoidable, and with it the sweep of natural evil in the biological realm?".[21] It appears, then, that extreme optimism is burdened with regressive manifestations of the problem of evil.

[9] Austin Farrer, *Love Almighty and Ills Unlimited* (Collins, 1962), 53-54.

[10] Austin Farrer, *Love Almighty and Ills Unlimited* (Collins, 1962), 51.

[11] Austin Farrer, *Love Almighty and Ills Unlimited* (Collins, 1962), 67.

[12] Nagasawa, *The Problem of Evil for Atheists,* 129.

[13] Charles Darwin, "22 May 1860 Letter to Asa Gray," Darwin Correspondence Project, accessed on 5 December 2024, https://www.darwinproject.ac.uk/letter/DCP-LETT-2814.xml.

[14] Paul Prescott, "The Secular Problem of Evil: An Essay in Analytic Existentialism," *Religious Studies* 57 (2021): 102.

[15] Richard Swinburne, "Natural Evil," *American Philosophical Quarterly* 15, No. 4 (1978): 299.

[16] Holmes Rolston III, *Science and Religion: A Critical Survey* (London: Templeton Foundation press, 2006), 134.

[17] Southgate, *The Groaning of Creation*, 58.

[18] Southgate, *The Groaning of Creation*, 72.

[19] Peter van Inwagen, "The Problem of Evil, the Problem of Air, and the Problem of Silence," *Philosophical Perspectives* 5 (1991): 143-45.

[20] Pierre Bayle, *Historical and Critical Dictionary: Selections,* translated by Richard H. Popkin and Craig Brush (Hackett, 1991), 177.

[21] Robert John Russell, "Natural Theodicy in an Evolutionary Context," in *Cosmology: From Alpha to Omega* (Fortress Press, 2008), 259.

In sum, while extreme optimists attempt to reconcile systemic evil with the claim that this is the best possible world through the use of free-will and "only-way" theodicies, such strategies ultimately defer rather than resolve the problem. Faced with empirical evidence of seemingly gratuitous suffering, they must either deny these realities or accept increasingly speculative theological explanations. While extreme optimism may appeal to the heavenly bliss defence, it still does not explain *why* natural selection is the best possible means towards that end without returning to this regress or begging the question. As such, extreme optimism appears ill-equipped to resolve the tension Nagasawa identifies between systemic evil and modest optimism. So, theists must either concede that natural selection is not the best necessary instrument in the best possible world, or following Bayle and Russell accept the former's pessimism or latter's "agnostic cosmic theodicy" in accepting that POSE cannot be answered.[22]

1.2 **Neutral optimism**

The second theist modest optimists, the neutral optimists, reject that the actual world is necessarily the best, but rather affirms that God actualised one of many possible overall good worlds. For instance, Robert Merrihew Adams argues that extreme optimism inappropriately imposes a utilitarian standard of moral goodness to God's omnibenevolence. Instead, he argues that traditional Judeo-Christian ethics account for God's goodness in terms of his grace—an inclination to love that is not based on the merit of the one being loved.[23] Indeed, core to Abrahamic monotheism is an affirmation of God's aseity, his self-sufficiency and independence from any external cause or necessity. [24] If God were obligated to create the best possible world in order to express his power or love, then his omnipotence and omnibenevolence would become contingent on something external—namely, the existence of that world—thereby undermining his aseity. It follows, therefore, that a being who never exists is not wronged by not being created, since existence itself is not owed to any potential being.[25] Furthermore, beings in the actual but not best world have no right to complain, lest they express an unmerited claim for special treatment or violate modest optimism.[26] God's omnibenevolence, therefore, does not demand that he create the best world possible.

As an implication, neutral optimists can entertain that there is a possible world without natural selection where we exist. However, two considerations may constrain this possibility. Firstly, this possible world must be logically coherent. Thomas Morris argues that if God's omnipotence is committed to what is logically and semantically possible, God becomes a "delimiter of possibilities."[27] That is, as God's existence is necessary in all possible worlds, those worlds must reflect his omnipotence by being logically coherent and his omnibenevolence by being overall good. This means that if a world without natural selection either fails to be logically coherent or cannot sustain overall goodness without introducing other systemic evils, it may not be a genuine possibility after all. Secondly, this limitation implies that a possible world without natural selection where we exist is not necessarily better or worse than the actual world. It could very well be that following the "only-way" theodicies, the goodness of true biological selves must necessarily come through natural selection and that this outweighs the evil of natural selection. Regardless, the neutral optimist is distinct in that they can be grateful for their existence without necessarily implying that natural selection is instrumentally good.

One obvious challenge against neutral optimism is its shifting definition of God's omnibenevolence may not be intuitively satisfying. For instance, Adams's definition of God's "grace", which does not require universal benevolence to all creatures, may only be satisfactory to some Calvinists or those within certain theological traditions. While this conception asserts that natural selection does not need to be justified as instrumentally good, the reality and impact of systemic evil make it difficult for suffering beings to reconcile that God's omnibenevolence does not require him to show grace to them, in tension with their own intuitions about what it means to be loving. However, as this critique may hold less weight for those

[22] Robert John Russell, "Natural Theodicy in an Evolutionary Context," in *Cosmology: From Alpha to Omega* (Fortress Press, 2008), 255.

[23] Robert Merrihew Adams, "Must God Create the Best?", *Philosophical Review* 81 (1972): 324.

[24] Ian A. McFarland, *From Nothing: A Theology of Creation* (Westminster John Knox Press, 2014), 61.

[25] Adams, "Must God Create the Best? 319-20.

[26] Adams, "Must God Create the Best? 319-20.

[27] Thomas V. Morris, "The Necessity of God's Goodness," *New Scholasticism* 59 (1985): 425.

aligned with certain Calvinist doctrines, where such a conception of grace is more readily accepted, it will be set aside as a doctrinal matter.

A more universal challenge is that even if a neutral optimist can maintain modest optimism about their existence while affirming systemic evil through yearning for another possible world, logical constraints on such worlds mean that regretting the evils of the actual world may require relinquishing goods unique to its constitution. For example, recalling Swinburne's free-will theodicy, a possible world without natural selection might lead to it not having human agency. Similarly, recalling Southgate's "only-way" theodicy, a world without natural selection could lack independent selves. If the existence of goods like human agency or autonomous selves carry significant moral weight, then removing the conditions that produce them (i.e., natural selection) may render the alternative world no longer overall good—and thus not genuinely possible. At best, such possible worlds without natural selection might not involve a loss of goods significant enough to undermine modest optimism. At worst, the trade-offs could introduce greater problems of evil. A creationist world, for instance, implies that God played a direct role in designing cruel beings like the Ichneumonidae than if they developed independently through evolution.

Comparing extreme and neutral theistic optimism, both conceptions of modest optimism requires that the world is overall good. This is because evidence of systemic evils must be outweighed by some other goodness or burdened with a theodicy. This, however, is not a requirement for atheist/non-theist optimism.

Section 2: Two atheist/non-theist modest optimists

2.1 Personal optimism

The first atheist/non-theist modest optimist approach argues that the scope of existential gratitude can be limited to the personal level without axiologically considering the world as an aggregate. While Dawkins expresses his gratitude for existing despite unfavourable odds, he regrets that, "Nature is red in tooth and claw. But I don't want to live in that kind of a world. I want to change the world in which I live in such a way that natural selection no longer applies."[28] However, we can resolve Dawkins' apparent disjunct by affirming *personal* existential optimism directed at one's own existence while rejecting *cosmic* existential optimism that the world is overall good. This is not methodologically novel; Asha Lancaster-Thomas observes that even within individuals' lifetimes, we are grateful for some parts of our lives, but not parts characterised by pain and suffering such as a painful chronic illness.[29]

An implication of personal, but not cosmic, optimism is that their existential gratitude does not need to consider the axiology of natural selection. One could remain axiologically agnostic towards the instruments of their existence, while valuing the goodness of their personal existence. Guy Kahane emphasises this distinction by arguing that even if natural selection is a causally fundamental instrument to our existence, it is axiologically irrelevant as instrumental value alone does not add any overall value to the world.[30] Under this conception, one could even be cosmically pessimistic but still be optimistic about their personal life as they experience it. Modest optimism is thus reinterpreted to affirm attitudinal optimism, that we are grateful to exist in this world; but not axiological optimism, that the world is overall good.[31]

However, after disregarding pessimism, personal optimism appears empirically challenged as most personal optimists are often implicitly also cosmic optimists. Responding to Kahane, Nagasawa grants that personal optimism does not necessarily entail cosmic optimism. However, he argues that this reformulation of modest optimism changes the target of POSE, which defines modest optimism as affirming both attitudinal and axiological optimism.[32] For he argues that rational personal optimists who procreate implicitly believe that the world they are bringing their child into is overall a good place.[33] The personal, but not cosmic,

[28] Frank Miele, "Darwin's Dangerous Disciple: An Interview with Richard Dawkins," *The Skeptic*, 27 October 2010, https://www.skeptic.com/eskeptic/10-10-27/.

[29] Asha Lancaster-Thomas, "Can Heaven Justify Horrendous Moral Evils? A Postmortem Autopsy," *Religions* 14, No. 296 (2023): 6.

[30] Guy Kahane, "Optimism without theism? Nagasawa on Atheism, Evolution, and Evil," *Religious Studies* 58 (2022): 706.

[31] Guy Kahane, "Optimism without theism? Nagasawa on Atheism, Evolution, and Evil," *Religious Studies* 58 (2022): 702.

[32] Nagasawa, *The Problem of Evil for Atheists,* 184.

[33] Nagasawa, *The Problem of Evil for Atheists,* 184.

reformulation of modest optimism, therefore, seemingly misses the original target of POSE and is only applicable to a minority of anti-natalist pessimists like David Benatar.

Responding to this, Nagasawa's formulation of modest optimism is already limited to the scope of"the environment in which we exist." The specific environment of individual experiences does not necessarily include the predation experienced by other preyed beings. Indeed, this does not preclude the modest optimist from being selfish for bringing a child into the world. Or disregarding the pains of the world, a personally optimistic individual can choose to be ignorant of the world's plights by never contributing to charitable causes to use the money to instead maximise personal pleasures. It is not evident, therefore, that most personal optimists must also be cosmic optimists.

2.2 **Comparative optimism**

The second atheist/non-theist modest optimist approach argues that modest optimism only views the world as *comparatively* good, but not necessarily *categorically* good. That is, the world must only be *comparatively* better than non-existence, rather than positively good. This distinction is significant, as Nagasawa's comparative argument for theism seems to present the axiology of the world in binary categorical terms. Theism's appeal to a heavenly bliss allows for a world with more goodness rather than evil.[34] But because atheists/non-theists are not committed to affirming an omnibenevolent God, Kahane argues that they are not obliged to claim that their existence is categorically good, or that the world contains more goodness than evil. Indeed, even under Leibniz's extreme optimism, the world is not necessarily categorically good, just that it is comparatively the best of all possible worlds.[35]

An implication of a comparatively better, but not categorically good, optimism is that natural selection does not have to be categorically good. Assuming that existence in itself is a good greater than all kinds of non-existence, an actual world with systemic evil is better than any unactualised world. So, modest optimism's "not bad" is equated to being comparatively better than non-existence. Opposing theism's appeal to the supernatural, this essentially lowers the requirement for modest optimism.

One major challenge is that this comparative-goodness version of modest optimism closely borders on pessimism, and therefore demands an account of why existence, despite systemic evils, is fundamentally and overall better than non-existence. The pessimist Benatar, for instance, argues that the absence of pain is always good, even if no one benefits, whereas the absence of pleasure is only bad if someone is deprived by it. This asymmetry supports his claim that existence, with its inevitable suffering, may be worse than non-existence, which guarantees goodness with no badness.[36]

Responding to Benatar, the optimist can follow Thaddeus Metz's argument against Benatar's claim that the absence of pain is good, describing the absence of pain as *not bad* rather than *good*.[37] Otherwise, the atheist/non-theist modest optimist can simply appeal to the previously-discussed personal, rather than cosmic, optimism. All modest optimism demands is that according to myself, it is better for me to exist than for me not to exist. Indeed, Benetar seems to grant this notion, as he distinguishes a present-tense "life worth continuing" and future-tense "life worth starting."[38] Personal optimists often experience instances where the goods of actualised pleasure outweigh the evils of pain, resulting in a net utility that makes existence preferable to non-existence. So, unless one is personally pessimistic, there is nothing paradoxical about claiming one's personal life is better to exist than not exist.

Combining these two approaches, the atheist/non-theist, can commit to a personal and comparative form of modest optimism that still accounts for the categorically systemic evil of the cosmos. Unlike theistic extreme optimism's commitment to the instrumental value of natural selection as a part of God's providence, personal optimists can simply remain agnostic about natural systems' axiological value. But while theistic neutral optimists can adopt a similar approach to the atheism/non-theism's comparative (not categorical) goodness, they remain committed to both that possible worlds must overall be good, and that God's creative ability is bound to logical laws, so that the possible worlds they yearn for must necessarily contain some

[34] Nagasawa, *The Problem of Evil for Atheists,* 171.
[35] Kahane, "Optimism Without Theism," 713.
[36] David Benatar, *Better Never to Have Been: The Harm of Coming into Existence* (Oxford University Press, 2006), 30.
[37] Thaddeus Metz, "Are Lives Worth Creating?", *Philosophical Papers* 40, No. 2 (2011): 241-45
[38] Benatar, *Better Never to Have Been,* 22-23.

other kind of systemic evil that requires a theodicy . The personal optimist on the other hand need not make this consideration of the overall goodness of other possible worlds. So, whilst theism can appeal to the heavenly bliss, the non-theist can simply bypass POSE without needing to address it.

Section 3: **Borrowing Theism's Optimism Without its Metaphysics**

But even if atheists/non-theists remain burdened by POSE due to perhaps their cosmic or even categorical optimism, I propose that they can "borrow" the theodicies used by theists to justify their modest optimism. This reverses Nagasawa's theistic strategy, which claims that theism's supernaturalist ontology (encompassing both natural and supernatural realms) subsumes the atheist/non-theist's naturalist ontology (limited to the natural world), thus allowing theists to "borrow" atheist/non-theist responses to POSE.[39] However, Nagasawa does not address the fact that supernaturalist ontologies bring additional axiological presuppositions—namely, that an omnibenevolent God exists and that his creation must necessarily be overall and categorically good. Non-theists, by contrast, can adopt the theist's belief that the world is overall good using the theist's rationalisations, without committing to these broader metaphysical claims about God. In essence, atheists/non-theists can justify their optimism in the face of POSE without having to commit to the theist's wider ontological framework.

Borrowing from extreme theistic optimism, the atheist/non-theist can still view natural selection as categorically good by appealing to the same free-will and "only-way" theodicies—without relying on theological assumptions. For instance, they may regard natural selection as instrumentally necessary for the emergence of goods like human free-will or biological selves and affirm these outcomes as categorically valuable in themselves. There is nothing inherently theological in valuing such features of natural history. While theists might argue that moral value requires an objective grounding in God, the atheist can respond in two ways: either by offering a naturalistic foundation for moral value, or by treating such value judgements—and the modest optimism they support—as subjective, grounded in personal or shared human perspectives. On this view, modest optimism need not depend on the objective truth of its content but rather functions as an attitudinal stance. Accordingly, theist theodicies can be borrowed by non-theists as explanatory tools, enabling them to affirm the world's overall goodness without committing to metaphysical claims that theists traditionally used to justify them.

Borrowing from neutral theistic optimism, the atheist/non-theist can still affirm that the actual world is not necessarily the best possible world, but still trust that it is better to exist than not to exist. The lack of a requirement for atheists/non-theists to commit to the idea that the world is categorically good allows for a more flexible position. Even if systemic evils suggest that the world is not fundamentally good, the personal optimist can still maintain a stance of cosmic neutrality. They can accept the world as it is—flawed, but not necessarily bad in a way that undermines their gratitude for existing. Indeed, without a commitment to an omnibenevolent God who governs over all creation's actions, the non-theist can simply adopt a position of gratitude for the outcomes of those processes without ascribing moral or intrinsic value to these violent/harsh (but not immoral) systemic processes themselves.

This strategic borrowing highlights a key asymmetry: while theists must reconcile systemic evil with a metaphysical commitment to a categorically good creation, non-theists can adopt similar explanatory frameworks without such constraints. In doing so, they preserve the practical benefits of modest optimism without incurring the theological debts that weigh down the theistic response to POSE.

Conclusion

POSE, therefore, remains a problem only for theists as their conception of modest theism must commit to the belief that a good God would create a categorically good world. This commitment imposes significant burdens ontheist extreme optimists, whose belief that the actual world is the best possible world obliges them either to embrace pessimism, appeal to mystery, or present a theodicy for systemic evils. And while responses like the free-will and "only-way" theodicies may present *prima facie* defences to POSE, they only

[39] Nagasawa, *The Problem of Evil for Atheists,* 173.

regress into deeper manifestations of the problem of evil unless the theist begs the question or makes an appeal to mystery. Likewise, theist neutral optimists, who holds that the actual world is only one of many possible worlds that are not necessarily the best ones, remain committed to asserting that world is overall good—which is still difficult to reconcile with or even amplifies the existence of systemic evils.

In contrast, the atheist/non-theist can either borrow the theist's theodicies, or maintain a personal comparative optimistic stance that disregards POSE overall. By selfishly narrowing modest optimism to the personal level, the atheist/non-theist can disregard systemic evils while remaining grateful for their own lives as they experience it. Furthermore, their non-commitment to categorical goodness allows them to value comparatively their personal lives as better than non-existence, even if by borrowing neutral optimism, they accept the world as it is and appreciate the outcomes of systemic processes like natural selection without assigning moral or intrinsic value to them.

Bibliography

Adams, Robert Merrihew. "Must God Create the Best?" *Philosophical Review* 81 (1972): 317-332.

Bayle, Pierre. *Historical and Critical Dictionary: Selections.* Translated by Richard H. Popkin and Craig Brush. Indianapolis, IN: Hackett, 1991.

Benatar, David. *Better Never to Have Been: The Harm of Coming into Existence*. Oxford University Press, 2006.

Darwin, Charles. "22 May 1860 Letter to Asa Gray." Darwin Correspondence Project. Accessed on 5 December 2024. https://www.darwinproject.ac.uk/letter/DCP-LETT-2814.xml.

Dawkins, Richard. "The Greatest Show on Earth." Lecture, University of Auckland, Auckland, New Zealand, 13 March 2010. Accessed on 3 May 2025. https://www.auckland.ac.nz/en/alumni/whats-happening/alumni-video-and-audio/alumni-videos-richard-dawkins.html.

Rolston III, Holmes. *Science and Religion: A Critical Survey.* London: Templeton Foundation Press, 2006.

Farrer, Austin. *Love Almighty and Ills Unlimited.* Collins, 1962.

Guy, Kahane. "Optimism without theism? Nagasawa on atheism, evolution, and evil." *Religious Studies* 58 (2022): 701-714.

Lancaster-Thomas, Asha. "Can Heaven Justify Horrendous Moral Evils? A Postmortem Autopsy." *Religions* 14, No. 296 (2023).

Leibniz, G. W. *Theodicy.* Edited by Austin Farrer. Translated by E. M. Huggard. Open Court Publishing Company, 1985.

McFarland, Ian A. *From Nothing: A Theology of Creation.* Westminster John Knox Press, 2014.

Metz, Thaddeus. "Are Lives Worth Creating?" Philosophical Papers 40, No. 2 (2011): 233-255.

Miele, Frank. "Darwin's Dangerous Disciple: An Interview with Richard Dawkins," The Skeptic. 27 October 2010. https://www.skeptic.com/eskeptic/10-10-27/.

Morris, Thomas V. "The Necessity of God's Goodness." *New Scholasticism* 59 (1985): 418-448.

Nagasawa, Yujin. *The Problem of Evil for Atheists.* Oxford University Press, 2024.

Prescott, Paul. "The Secular Problem of Evil: An Essay in Analytic Existentialism." *Religious Studies* 57 (2021): 101-119.

Russell, Robert John. "Natural Theodicy in an Evolutionary Context." In *Cosmology: From Alpha to Omega*. Fortress Press, 2008.

Southgate, Christopher. *The Groaning of Creation: God, Evolution, and the Problem of Evil.* Westminster John Knox Press, 2008.

Swinburne, Richard. "Natural Evil." *American Philosophical Quarterly* 15, No. 4 (1978): 295-301.

Thompson, Janna. "The Apology Paradox." *The Philosophical Quarterly* 50, No. 201 (2000): 470-475.

Van Inwagen, Peter. "The Problem of Evil, The Problem of Air, and the Problem of Silence," *Philosophical Perspectives* 5 (1991): 135-165.

Supervaluationism, Dynamic Supervaluationism, and Higher-Order Vagueness

Wiktor Przyborowski, *University of St Andrews*

The fact that the phenomenon of vagueness can itself be vague - and its vagueness be vague as well - seems impossible to make sense of without getting a headache. This so-called higher-order vagueness makes theorising about vagueness a notoriously difficult task for philosophers of logic and language. This difficulty manifests itself in that, even if a theory can convincingly explain what vagueness is and how we can reason about it, when faced with the vagueness of the just-tamed vagueness, it gets flooded with paradoxes and makes the initial theory seem implausible. In this paper, I argue that Rosanna Keefe's supervaluationism is one such theory. Even though it elegantly accounts for the first order of vagueness, it becomes less elegant when questioned about the higher orders. To demonstrate this, I show that Keefe's system fails to resolve various paradoxes of higher-order vagueness such as the finite series paradox or the D^* paradox. Furthermore, I argue that in her attempts to accommodate the paradoxes by adopting a rigid hierarchy of metalanguages, Keefe invites new worries. Given these criticisms, it is unlikely that Keefe's theory can be 'argued out' of these paradoxes - 'finite series' in particular. Instead, I argue that the theory must be substantially modified if it is to be salvaged, and one way to do so is by making the proposed structure more dynamic. I attempt to do so by sketching an outline of dynamic supervaluationism that can tackle the problems that Keefe's supervaluationism cannot. I close my essay by teasing out some challenges that the proposed theory could face and offering possible solutions. I believe that supervaluationism is a very attractive approach to vagueness and therefore, it is worth developing further into a more robust theory that could tackle its higher orders.

1. Introduction

Vagueness in language refers to an indeterminate relationship between its terms and the world they describe.[1] Minimally, a predicate is vague if it has three features: **admission of borderline cases** (objects to which its application is unclear), **a lack of known, sharp boundaries** (no clear case separating the positive and negative cases), and (apparent) **susceptibility to the Sorites paradox**.[2]

Vagueness is philosophically relevant because it raises two problems. First, the **semantic problem**: since the vague extension is unclear, classical semantics (where meaning is derived from extension), and hence classical logic, may not apply. Second, the **Soritical problem**. Consider a series of people of descending heights by 1cm. The first is clearly tall (200cm) and the last is clearly not (120cm). Since no known boundaries exist, vague predicates are tolerant - a small change will not alter the application. Thus, by inductive step, for any case **n**, 'if **n** is tall then **n** + 1 is tall'. Starting at 200cm is tall, via a series of conditionals, you validly conclude that 120cm is tall. However, this is a contradiction since 120cm is clearly not tall.[3] This argument exemplifies the classical form of the Sorites paradox.

Theorizing about vagueness involves accounting for the nature, source and meaning of vagueness, providing vague semantics and resolving the Sorites. Furthermore, since it is unknowable where the positive extension changes to negative, it is equally unknowable where the positive changes to borderline. Thus, borderline cases themselves should be unbounded; hence there should be borderlines to borderlines. The process could be iterated to establish a possibly infinite hierarchy of borderline cases: the higher-order

[1] Kit Fine, *Vagueness: A Global Approach* (Oxford Academic, 2020), 2-3.
[2] Rosanna Keefe, *Theories of Vagueness*, (Cambridge University Press, 2000), 6-7.
[3] Fine, *Vagueness*, 3-7.

vagueness (HOV).[4]

Throughout this paper, I will follow Rosanna Keefe and other major supervaluationists in assuming that HOV is a genuine problem, that needs to be accounted for. However, it is worth pointing out that this is a debated matter in the field.[5] Nevertheless, under this assumption a successful theory of vagueness, given its commitments, must also account for HOV.

In this essay, I explore how one theory of vagueness – supervaluationism, advocated by Rosanna Keefe – does so. First, I outline her account of first-order vagueness (FOV). Then, I explain the problems posed by HOV, examining Tim Williamson's criticisms of the theory and how Keefe accommodates them. I will argue that although the Williamson problems are solved, the resulting view does not reflect how language actually functions and is paradoxical, making the HOV account unsatisfactory. I then attempt to modify the view by dynamizing it, developing the ideas of Hao-Cheng Fu. I defend the model by showing how it solves some of the critical issues faced by Keefe. Lastly, I raise a few possible issues endemic to the dynamic view and sketch responses to defend it.

2. Supervaluationism, a theory of vagueness

Supervaluationists claim that vagueness is a problem of language, not our epistemic capacities. They argue that vague predicates fail to draw sharp boundaries, not that these boundaries are unknowable, and that they admit borderline cases. The source is semantic indecision. A vague predicate admits a range of possible extensions, but it is semantically unsettled which one is correct. This is captured through the notion of precisification, a way to make a vague term precise.[6] A precisification must be admissible, reasonable in not licensing a misuse of language.[7] It also must be complete, it categorizes objects into positive and negative extensions, leaving nothing in-between. For illustration, consider the vague predicate 'tall'. We could (reasonably) use precisifications: 'tall' is true if '>175cm', '>180cm' and '>190cm', each of which would precisely divide objects into positive and negative extensions. Vague terms do not 'choose' between these; instead, all precisifications are equally good.[8]

Supervaluationists provide semantics for vague predicates, identifying truth with super-truth by considering all possible precisifications. $F\mathbf{a}$ is super-true (-false) iff F is true (false) of $\mathbf{a}$ under all complete and admissible precisifications. $F\mathbf{a}$ is neither true nor false iff F is true of $\mathbf{a}$ under some precisifications and false of $\mathbf{a}$ under others.[9]

Thus, vague predicates divide objects in a three-fold manner, where borderline cases are not assigned a definite truth value. Hence, supervaluationists give up bivalence, departing from classical semantics, by admitting truth value gaps. On the other hand, classical logic is mostly preserved because if a sentence is classically true, then it is true on all complete and admissible precisifications. Consider the law of excluded middle. Using any precisification of tall – every object will be either tall or not-tall, since every precisification divides objects into two sharp sets. Similarly, all classical theorems are retained, thus we can use classical logic to reason about vague predicates.[10]

This idea provides a straightforward solution to the Sorites. Namely, the inductive premise 'if $F\mathbf{n}$ then $F(\mathbf{n}+1)$' is super-false, since the antecedent will be true and the consequent false for some $\mathbf{n}$ under any complete and admissible precisification. This is because each precisification, being complete, provides a sharp cut-off between the true and false – a bordering pair where the first entry is true and second one

[4] Keefe, *Theories of Vagueness*, 31-32.

[5] Some philosophers, such as Dominic Hyde, claim that higher-order vagueness (HOV) is a pseudo-problem, arguing that the vagueness of vague is a real, but unproblematic, phenomenon. Others, including Hao-Cheng Fu and Susanne Bobzien counter that this stance fails to adequately address the complexity of the issue, maintaining that HOV is indeed a genuine problem. While an extensive discussion is beyond the scope of this essay, see Hyde, "Why Higher-Order Vagueness Is a Pseudo-Problem"; Fu, "Saving Supervaluationism from the Challenge of Higher-Order Vagueness Argument"; and Bobzien, "In Defense of True Higher-Order Vagueness" for further details.

[6] Keefe, *Theories of Vagueness*, 154-156.

[7] Timothy Williamson, *Vagueness*, (Routledge, 1994), 158.

[8] Keefe, *Theories of Vagueness*, 154-156.

[9] Keefe, *Theories of Vagueness*, 154.

[10] Rosanna Keefe, "Vagueness: Supervaluationism," *Philosophy Compass* 3, no. 2 (2008): 162-164.

is false.[11] Thus, the supervaluationist account fulfils the initial demands of theorizing about vagueness. Consult the footnote[12] for further clarification.

3. Supervaluationism and higher-order vagueness

The above metalanguage (talk of truth conditions) expresses the vagueness of the object language by dividing cases into three sharply bounded sets (true, false, borderline). This can be captured by adding a 'definitely' D operator to the object language, which functions akin to modal necessity.

The FOV of F is expressed as:

(1) DFx for definite positive cases (true under all complete and admissible precisifications)

(2) $\sim DFx \,\&\, \sim D\sim Fx$ for borderline cases (true/false under some)

(3) $D\sim Fx$ for negative cases (false under all)

This division is problematic since all cases are sharply categorized, allowing no borderlines between the definite and borderline cases, leaving no scope for HOV. Supervaluationists argue that this can be resolved by allowing the concept of 'admissibility' itself to be vague, thus making the metalanguage vague.[13]

Hence, the second-order vagueness of F is captured in the meta-metalanguage by expressing vagueness of DF (the metalanguage). This yields the following five-fold classification:

(1) $DDFx$, i.e., definitely definitely positive cases

(2) $\sim DDFx \& \sim D\sim DFx$, i.e., borderline between positive and borderline

(3) $D\sim DFx \& D\sim D\sim Fx$, i.e., definitely borderline cases

(4) $\sim DD\sim Fx \& \sim D\sim D\sim Fx$, i.e., borderline between negative and borderline

(5) $DD\sim Fx$, i.e., definitely definitely negative cases

The general idea is that for level vagueness of F, we need to show that **n** categories are vague. Thus, we need borderlines between those, in effect, drawing $2^n + 1$ categories.[14]

3.1. Williamson's challenge

Williamson argues that for this formalization to work, the D operator should not obey these two schemas:

(1) The S5 principle: If $\sim DF$, then $D\sim DF$.

(2) The S4 principle: If DF, then DDF.

If (1) and (2) hold, then whether a category is definite or indefinite, it will also be definitely so at higher levels. The supervaluationist cannot accept this since each category must be vague, otherwise it would draw sharp boundaries. Thus, Williamson recommends adopting a weaker modal logic, like T, with relative admissibility and no transitivity so that both S4 and S5 principles fail.[15] See the appendix for a more formal explanation.

[11] Keefe, "Vagueness: Supervaluationism," 315-316.

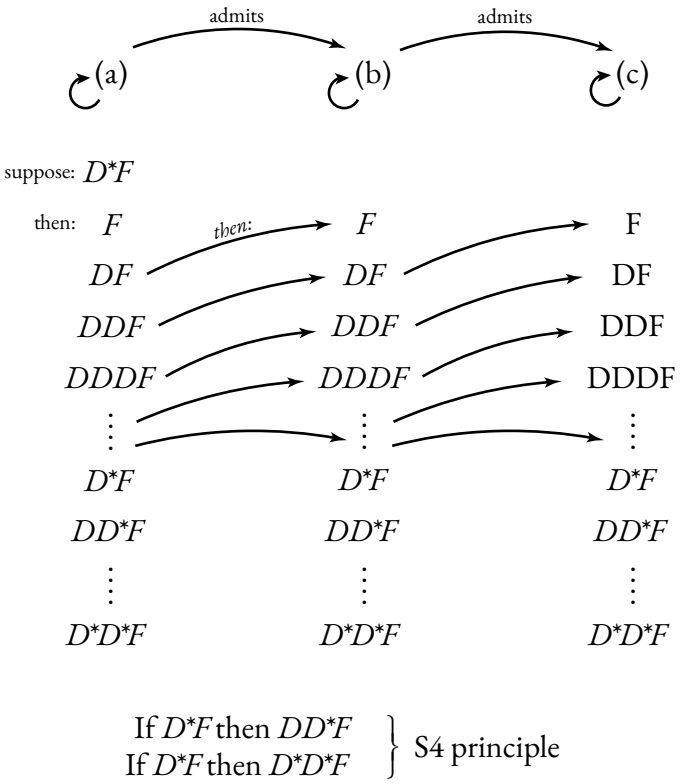
[12] Consider the series of people of varying heights again and suppose some examples of complete precisifications: **x** is short if (1) '< 160cm' or (2) '< 165cm' or (3) '< 170cm'. They are complete since they divide objects into positive (short) and negative (not-short) extensions with nothing in-between. It is easy to see how the inductive premise turns out false on each of these precisifications: (1) 'If 159cm is short, then 160 is short'; (2) 'If 164cm is short, then 165cm is short'; (3) 'If 169cm is short, then 170cm is short'. In each case, the antecedent is true and the consequent false (relative to precisification). Since the inductive premise turns out false for some pair under each complete precisification, it is super-false.

[13] Keefe, *Theories of Vagueness*, 202-204.

[14] Mark Sainsbury, "Concepts without Boundaries," in *Departing From Frege* (Routledge, 1990), 74.

[15] Williamson, *Vagueness*, 156-159.

However, Williamson argues that this is not sufficient to solve the problem via the D^* argument. He defines D^*F as an infinite conjunction F & DF & DDF & ... & D_nF. Suppose precisifications (a), (b), and (c), where (a) admits (b), and (b) admits (c), but (a) does not admit (c), since admissibility is non-transitive. Suppose D^*F at (a). This means that F, DF, DDF, ..., D_nF are true at (a). If DF is true at (a), then F is true at (b); if DDF is true at (a), then DF is true at (b); and so on. Thus, F, DF, DDF, ..., D_nF are all true at (b), and hence D^*F is true at (b). The same reasoning applies to (c). Thus, if D^*F is true at some precisification, then D^*F is true at all precisifications. Hence, DD^*F is true at all precisifications - and by the same reasoning, so is D^*D^*F. Therefore, the S4 principle effectively applies to D^* (see diagram below).



Consequently, Williamson concludes that higher-order vagueness disappears.[16] This is because, for supervaluationism to succeed, each metalanguage must be vague. Thus, supervaluationists need a borderline case between D^*F and $D^*{\sim}F$, namely ${\sim}DD^*F$ & ${\sim}D{\sim}D^*F$. However, ${\sim}DD^*F$ collapses to ${\sim}D^*F$ by modus tollens on the S4 principle. ${\sim}D^*F$ then collapses to $D{\sim}D^*F$, given closure of D.[17] In effect, ${\sim}DD^*F$ & ${\sim}D{\sim}D^*F$ reduces to $D{\sim}D^*F$ & ${\sim}D{\sim}D^*F$ which is a contradiction. Since there are no borderlines to D^*F, it is not vague.

Williamson offers supervaluationists a way out: to give up semantic closure. D^* can be vague but its vagueness cannot be expressed using D or D^*. Instead, we need a meta-language for D^*, enriched with a distinct operator, $D!$. Then, to express vagueness of $D!$, we need a meta-metalanguage with $D!!$. Williamson remarks that the process could continue infinitely.[18]

Keefe takes up this proposal and advocates adopting an infinite, hierarchical series of metalanguages. In this model, the vagueness of the n^{th}-level metalanguage can only be expressed in the $(n+1)^{\text{th}}$ metalanguage,

[16] Williamson, *Vagueness*, 160.
[17] Patrick Greenough, "Higher-Order Vagueness," *Proceedings of the Aristotelian Society, Supplementary Volumes* 79 (2005): 183.
[18] Williamson, *Vagueness*, 160-161.

which is essentially richer than the nth language. She argues that, since there is no reason not to adopt such an infinite sequence, she can just stipulate that all the languages in the series are vague.[19] Greenough sketches a formalization where the object language is enriched with indexed D operators where each D_{n+1} is used to express the vagueness of D_n. Such formalization stops the D^* paradox and ensures that a non-vague metalanguage cannot be generated.[20]

4. Evaluation

Even though the above account might seem abstract, its strength lies in its simplicity - Keefe only iterates her account of the first order to higher orders of vagueness. In effect, the initial solutions to vagueness problems equally apply to HOV. Vagueness at higher orders remains a matter of semantic indecision: we are undecided over whether a precisification counts as admissible. Furthermore, each level n admits borderline cases and lacks sharp boundaries – a fact that can be expressed in the $n + 1$ metalanguage using appropriate D operators.

Moreover, each higher order metalanguage is still Sorites susceptible. I will explain this by running the paradox for the metalanguage (second order vagueness) in natural language terms for clarity - though the same could be done using D operators. The inductive premise for the metalanguage can be restated, in natural language, as: 'if there are admissible precisifications that draw the boundary to 'tall' at height h, then there are ones that draw it at one-hundredth of an inch lower'.[21] The second order series could start with a clearly admissible precisification (e.g., taller than 190cm) and end with a clearly inadmissible one (e.g., taller than 110cm). Since one-hundredth of an inch does not make a difference in admissibility, you could run a series of conditionals, starting with 'taller than 190cm is admissible' to reach a conclusion that 'taller than 110cm is admissible'. This is a contradiction. To resolve the second-order paradox, Keefe reuses her earlier strategy: for any complete way of making 'admissible' precise (or making 'definitely' definite), there will be a pair such that the first precisification is admissible and the second is not. This could be run for any level of metalanguage.

Thus, Keefe's account of HOV fulfils all the demands of a theory of vagueness. Each metalanguage is vague since it (1) admits borderline cases, (2) draws no sharp boundaries and (3) is Sorites susceptible. The fact that she achieves this for each order while maintaining her initial commitments (using the same technique at each order, characterising all levels of vagueness as semantic indecision, and so on) makes her strategy simple and elegant.

Even though this iteration neatly maintains the supervaluationist method, iterating to infinity is problematic. Keefe boldly claims that 'if there is no general objection to the claim that the sequence of metalanguages for metalanguages is infinite, then what is the difficulty with adding 'and each of those languages is vague' '.[22] However, there is a fundamental difficulty in this addition. In Keefe's system, the vagueness of an n-level metalanguage can only be expressed via an n+1 level metalanguage. If all metalanguages are vague, then the infinite metalanguage would have to be vague. To express the vagueness of the infinite metalanguage, we would need to use the infinity + 1 metalanguage. However, adding another element to an infinite set would not alter the size of this set.[23] Thus, the infinite + 1 metalanguage would be on the same meta-level as the infinite metalanguage. Hence, the vagueness of the infinite metalanguage cannot be expressed and the statement 'each of those languages is vague' seems meaningless.

This objection points towards a more general issue with such Tarskian metalanguage hierarchies. Namely, that languages in such hierarchies cannot be globally quantified over.[24] Keefe could respond that even though the infinite metalanguage might not be definable in her structure, it does not mean that it does not exist. Her structure ensures that vagueness for any finite level can be expressed. Even though we cannot say that 'all metalanguages are vague', we also cannot identify any non-vague metalanguage within the structure.

[19] Keefe, *Theories of Vagueness*, 202-208.

[20] Greenough, "Higher-Order Vagueness," 184-186.

[21] Keefe, *Theories of Vagueness*, 207-208.

[22] Keefe, *Theories of Vagueness*, 208.

[23] MIT OpenCourseWare, *Session 11: Mathematics for Computer Science*, *6.042J: Mathematics for Computer Science, Spring 2015* (Massachusetts Institute of Technology, 2015).

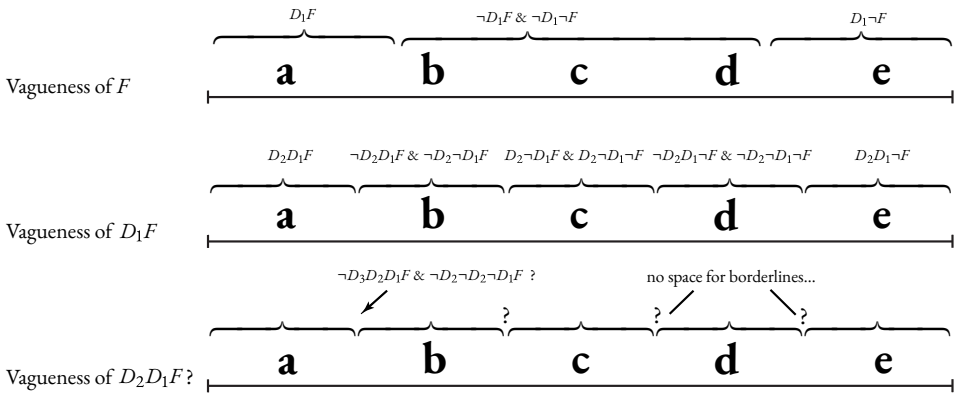
[24] Greenough, "Higher-Order Vagueness," 187.

Thus, even though the concept of infinity proves problematic for Keefe at the outset, I will assume that this problem does not threaten the explanatory power of her structure.

A further problem with the structure is that it is highly detached from how language functions. Competent speakers would find making sense of iterated uses of 'definitely' difficult, whether it is indexed or not. For example, saying someone is 'definitely definitely definitely tall' has little meaning apart from emphasis. Keefe might respond by pointing out that we do not use expressions like 'a googol of a googol of a googol' in ordinary conversation either, yet this does not mean the concept of 'googol' is not a meaningful mathematical concept. However, the issue goes deeper. As Saul Kripke pointed out, we cannot consistently assign levels to truth. Thus, even if we index the levels of 'definitely', it is difficult to assign them consistently. Consider the following statements: Jan says, 'Everything Alfred said is definitely false', and Saul says, 'Everything Jan said is definitely false'. To make sense of these, we would need to place one at a higher level in the hierarchy. However, this does not happen in natural language.[25]

Keefe might counter these natural language intuitions by arguing that her model is only an idealization which is not meant to exactly replicate how ordinary language works. While iterating 'definitely' (e.g., $D_3D_2D_1F$) may make little sense in casual conversation, the model is primarily defended by its explanatory power regarding HOV. She could further argue that even though different levels of metalanguages, when expressed in natural language, might not be clearly marked and distinguishable (such as in the Jan-Alfred example above), they can still function as distinct metalanguages in a formal framework. A further worry is that such an approach might over-idealise HOV making her account arbitrary. It raises the question over whether speakers genuinely use implicitly distinct levels of metalanguages to assign levels to truth. Thus, Keefe would need to give a more robust explanation of the relationship between her model and natural language.[26]

Lastly, even though Keefe's iteration method allows her to respond to Williamson's D^* paradox and establish that there cannot be a non-vague metalanguage, the non-vagueness of each metalanguage requires further borderline cases. We need $2^n + 1$ categories to express the vagueness of the nth metalanguage. However, there is a tension between an infinite number of categories and a finite number of objects in the series: the finite series paradox. Consider a simple series with 5 objects. To account for 1st level, we divide them into 3 categories. To account for 2nd level, we divide them into 5 categories. At 3rd level there are 9 categories to be filled but only 5 objects. This means that at some level we will run out of objects with which to fill the categories. As a result, there will be no borderline cases between the categories - providing a sharp boundary, as pictured below.[27] Whether or not Keefe indexes her D operators makes no difference, there will always be an insufficient number of objects in the series to fill all categories.



In conclusion, even though the rigid hierarchy in Keefe's structure might be defended to some extent, her

[25] Saul Kripke, "Outline of a Theory of Truth," *The Journal of Philosophy* 72, no. 19 (1975): 694-697.

[26] A full discussion of this issue is beyond the scope of this essay, though the problem would require further explanation to defend the account effectively.

[27] Greenough, "Higher-Order Vagueness," 180; 185-186.

appeal to an infinite hierarchy is fundamentally in conflict with the finite Sorites. There seems to be no way to accommodate the problem without making strong alterations to the model.

5. Positive proposal — dynamizing supervaluationism

5.1. Introducing dynamic supervaluationism

I believe that Keefe's problems can be addressed by making the structure's categories dynamic. My proposal is loosely based on Hao-Cheng Fu's model.[28] Fu rejects Keefe's claim that admissibility is vague and instead claims that, when considering a vague predicate, we are using a well-defined set of precisifications (p-sets). Keefe might argue this counterintuitive since we do not know what is admissible. However, this knowledge is unnecessary: the p-set is created when cases are categorized as true, false, or borderline at time t_1. For example, if 195cm and 190cm are tall, 170cm is not, and 180cm is borderline, the p-set is implicitly formed dividing cases into three groups, on my reading of Fu. Crucially, we judge first; the p-set is constructed afterward. What follows in the next paragraphs is my own development of the idea.

Fu applies the AGM theory[29] to give a complex account of the dynamics of p-sets; however, offers little formalisation and does not explain how this idea could be applied to the challenges of HOV[30]. Moreover, Fu does not address the paradoxes of HOV, and it is difficult to see how his account could solve them. In my view, we do not need such an elaborate account. I propose that a p-set is dynamic solely in virtue of changing when a case is judged inconsistently with it. For the sake of clarity, consider the above example again. Imagine another person, **x**, who is 168cm. You judge **x** as tall. This is clearly inconsistent with your p-set at t_1, since you judged 170cm as not tall. Thus, adding **x** to the tall category updates the t_1 set to the t_2 set with revised precisifications. This change occurs by either (1) expanding (adding a precisification), (2) contracting (removing one), or (3) both. Therefore, I retain the core idea of dynamic p-sets and Fu's terminology but limit the scope of the mechanism to a minimal principle: a p-set updates only when a judgment is made that conflicts with it.

I will now attempt to formalise the above proposed working of p-sets, which I will later apply to the challenges haunting supervaluationism. Vagueness, on the dynamic view, remains semantic indecision. At the first level, we follow Keefe's supervaluationism with a slight addition of the temporal component. While Fu does not offer a formalisation of his view in the spirit of Keefe's system with D operators, the following temporal framework develops my own way of modelling dynamic p-sets using temporally indexed D operators.

More precisely, at any time, t, cases divide into D_tF, $D_t{\sim}F$, and ${\sim}D_tF \,\&\, {\sim}D_t{\sim}F$: that is true, false, and borderline. However, unlike in Keefe's view, HOV arises not from undecided admissibility of a precisification but from the instability of precisifications. Suppose that you make some categorisations at t_1. According to the p-set that you just formed; some arbitrary case is classified as D_1F. Now suppose that you consider the series again, but you are no longer sure about the definiteness of your classification. Thus, your p-set is adjusted at t_2, and according to it, the case is borderline. Therefore, from t_2's perspective it was a borderline definite case at t_1 (${\sim}D_2D_1F$).

In general, when considering a borderline case after categorisation at t, tolerance ensures a mis-categorisation. To see this, remember that the supervaluation technique divides cases sharply into true, false, and borderline. However, tolerance guarantees that when viewing two neighbouring cases, we will not be able to tell the difference. Therefore, there is a clear tension; we divided sharply, enabling a border pair where, for instance, one member is true and another borderline. However, since we cannot distinguish between neighbouring cases, they must be categorised equally. That means that one of the cases had to be categorised mistakenly

[28] Hao-Cheng Fu, "Saving Supervaluationism from the Challenge of Higher-Order Vagueness Argument," in *Philosophical Logic: Current Trends in Asia* (2017), 147-152.

[29] AGM refers to the Alchourrón–Gärdenfors–Makinson model of belief revision, which accounts for rational change in epistemic states represented as belief sets. The theory outlines how agents should expand, contract, or revise their beliefs while preserving logical coherence. For more detail, see Carlos E. Alchourrón, Peter Gärdenfors, and David Makinson, "On the Logic of Theory Change: Partial Meet Contraction and Revision Functions," *The Journal of Symbolic Logic* 50, no. 2 (1985): 510–30.

[30] Fu, "Saving Supervaluationism from the Challenge of Higher-Order Vagueness Argument," 149-152.

and thus, the p-set must be revised to maintain consistency in our judgments. When we reconsider the series at t_2, the earlier categorisations from t_1 turn out to be indefinite, as case memberships shift.

5.2. **Applying dynamic supervaluationism**

Having formalised the view, I will now apply it to the challenges of HOV, starting with Williamson's D^* argument. To attack the dynamic approach, D^* could be restated as the conjunction 'DA at t_1 & DA at t_2 & DA at t_3 & … & DA at t_n'. As discussed in section 3, the D^* argument establishes that, if D^* is not shown to be vague, then the cases where D^* is true and the cases where D^* is false will both be ultimately definite. Hence, there will be no borderline cases between D^* categories, which provides a sharp boundary. This contradicts the foundational supervaluationist claim that there are no sharp boundaries. However, this argument loses its force under the dynamic view. The dynamic framework allows us to easily account for the vagueness of D^*. Just as in the case of any D, we need to progress in time to express D^*'s vagueness. Thus, while D^* may initially appear to be non-vague, this is because we need to move to t + 1 to realize its vagueness.

Secondly, Keefe's view faced concerns about rigid hierarchies, but the dynamic approach eliminates these. When two speakers disagree over a case's definiteness, neither statement must be 'prior'. They are simply speaking from different p-sets that underwent different evolutions. There is no rigid hierarchy of metalanguages since each discusses categorizations in another metalanguage, and no pair can be clearly ranked as 'prior'.

This lack of priority arises because it would be impossible to assign it to any particular metalanguage. Surely, the metalanguage at $t + 1$ must be a metalanguage of the metalanguage at t, since it is able to express facts about t. Therefore, it is more 'privileged' in this sense. However, suppose that the p-sets evolve over time such that, when moving from $t + 1$ to $t + 2$, we go back to the original p-set from t. Then, the t and $t + 2$ metalanguages gain their truth conditions from the same p-set. Therefore, in a sense, the t metalanguage becomes 'prior' to the $t + 1$ metalanguage. This would undermine the strict, unidirectional Tarskian hierarchy.

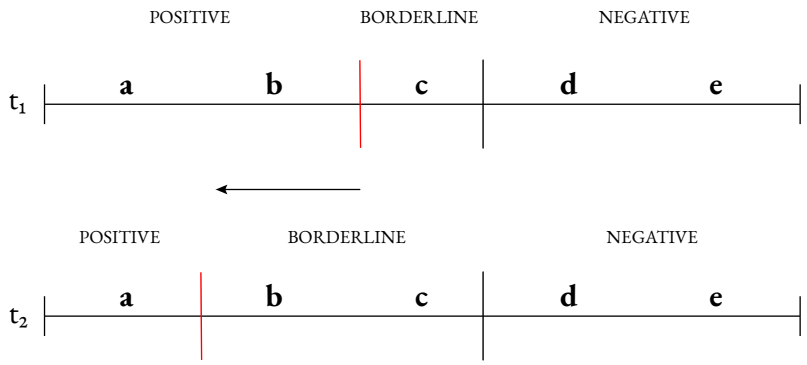
One could further argue that we could suppose a scenario in which two identical people, A and B, undergo identical p-set evolutions. However, A's evolution stops at t and B's evolution stops at t+1. On the one hand, we might be tempted to assign priority to B's statements, which would be counter-intuitive on the natural language objection. However, there is no reason to suppose that A's evolution should go the same way; she might consider a different part of the Sorites spectrum. Therefore, although the metalanguages are in some sense hierarchical, none has a clear priority in determining the truth of one classification over another. Thus, the objections, such as the ones made by Kripke, do not apply here.

Thirdly, the dynamic view can help tackle the finite series paradox, which was a critical blow to Keefe's account. I will explain how it could achieve this through an example. Consider a 5-element Sorites with objects **a**, **b**, **c**, **d**, and **e**. Suppose that Alfred's initial categorizations are:

$$\begin{aligned} D_1F &= \{\mathbf{a}, \mathbf{b}\} \\ \sim D_1F \,\&\, \sim D_1{\sim}F &= \{\mathbf{c}\} \\ D_1{\sim}F &= \{\mathbf{d}, \mathbf{e}\} \end{aligned}$$

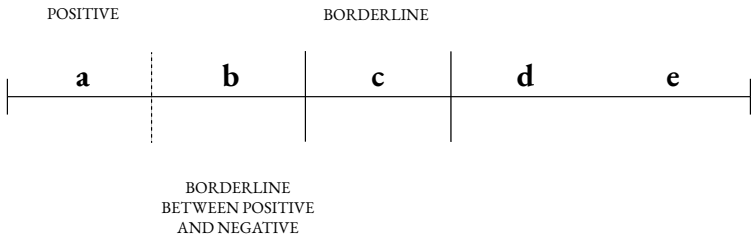
Alfred considers the pair **b** and **c** again. He realizes that he cannot tell the difference, concluding that **b** is also borderline. He adjusts his p-set accordingly, forming a new t_2 p-set.

$$\begin{aligned} D_2F &= \{\mathbf{a}\} \\ \sim D_2F \,\&\, \sim D_2{\sim}F &= \{\mathbf{b}, \mathbf{c}\} \\ D_2{\sim}F &= \{\mathbf{d}, \mathbf{e}\} \end{aligned}$$



The t_1 division, from the perspective of t_2 becomes:

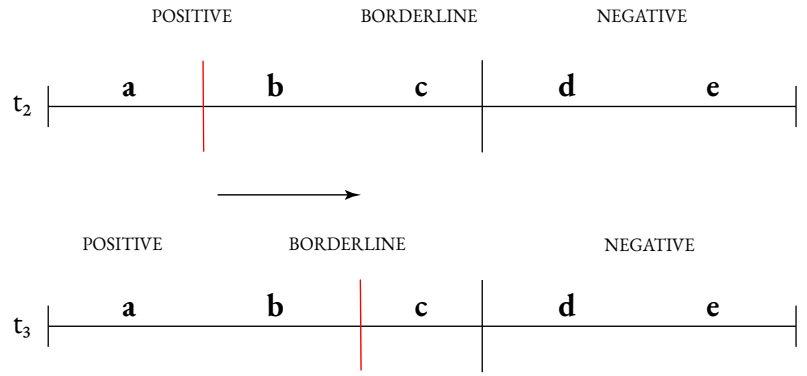
$$\begin{aligned} D_2D_1F &= \{\mathbf{a}\} \\ \sim D_2D_1F \,\&\, \sim D_2\sim D_1F &= \{\mathbf{b}\} \\ D_2\sim D_1F \,\&\, D_2\sim D_1\sim F &= \{\mathbf{c}\} \end{aligned}$$



Hence, in this part of the series, the vagueness of D_1 is fully accounted for since all D_1 categories have borderline cases.

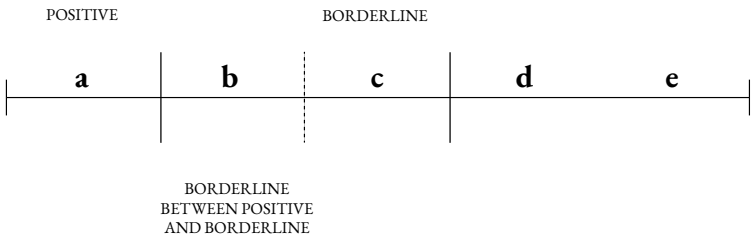
Now suppose that at time t_3, he looks at the pair **a** and **b**. Since he cannot tell the difference, he decides that b is also a definite case, adjusting the p-set again.

$$\begin{aligned} D_3F &= \{\mathbf{a}, \mathbf{b}\} \\ \sim D_3F \,\&\, \sim D_3\sim F &= \{\mathbf{c}\} \\ D_3\sim F &= \{\mathbf{d}, \mathbf{e}\} \end{aligned}$$



Since **b** changed its category membership, from the perspective of t_3, **b** was not a definite borderline case at t_2. Thus, the t_2 division, from the t_3 perspective, is:

$$\begin{aligned} D_3D_2F &= \{\mathbf{a}\} \\ \sim D_3D_3F \,\&\, \sim D_2\sim D_2F &= \{\mathbf{b}\} \\ D_3\sim D_2F \,\&\, D_3\sim D_2\sim F &= \{\mathbf{c}\} \end{aligned}$$



Thus, vagueness of D_2 is accounted for.

In general, any bordering pair will exhibit change when reassessed. Thus, any categorization at t can prove indefinite at $t+1$. In effect, you will never reach a point where there are more categories than members in the series since the fluid categories will always be filled. An object can fill different categories at different times. This also does not mean that the t_1 categories are definite at t_3, only that their vagueness cannot be expressed from the t_3 perspective.

6. Addressing possible objections

Dynamizing supervaluationism provides new methods to tackle the paradoxes of HOV and other problems, for which standard supervaluationism struggles to account. However, it also presents new worries, which I will explore and sketch responses to in this section of the essay.

6.1. Fixed time worry

The first possible objection to the view is that it breaks down when time is fixed. This is because the account of HOV relies on shifty p-sets, which in turn rely on the progress in time. More precisely, the vagueness of some set of categories drawn in period t can only be expressed in period $t+1$. Thus, if we hold the time fixed, the view breaks down: the categories drawn in period t appear to be sharply bounded, which contradicts the foundational claim that there are no sharp boundaries.

Although this might seem like a critical blow to the view, there are two possible lines of response. First, we could simply reject the inference from our inability to express the vagueness of some order when time is fixed, to the claim that there are sharp boundaries. After all, the fact that we cannot express it does not imply that it does not exist. This, however, demands further explanation of why we cannot express it. One response is that at a certain time, we are just using a well-defined but arbitrary set of precisifications. However, this division is surely wrong; it is made under one of many sets of equally good precisifications. Thus, there is no reason to believe that the term was made precise – we just have not realized our mistake yet.

A second and more powerful response is to deny the possibility of fixing time in this sense. This could supplement the above argument. Suppose that the critic of the view wants to prove to us that there are sharp boundaries. However, in order to show that there are sharp boundaries, they would have to find them in the series. Suppose that you manage to find the extension-switching pair. Even if you do this, you will realize, per tolerance, that you cannot tell the difference between the two cases. In effect, you must conclude that one of the cases was falsely classified when you made the division in the previous period. Thus, your p-set changes. Therefore, the very considering of the sharp distinction would automatically progress us to $t+1$, ensuring that there was no sharp boundary. In conclusion, the fixed time objection is not a significant worry to the dynamic view.

6.2. **Collapse to contextualism worry**

There is a second and more dangerous worry: one could argue that the supervaluationist aspect of the dynamic view seems unimportant. By this, I mean the use of supervaluationist semantics and classification of vagueness through indecision between precisifications. It is only directly applied to resolve FOV, and one could argue that the relativity of classifications over time, which accounts for HOV, could be equally applied to FOV. In effect, the supervaluationist method would disappear. If this argument is accepted, and if we further assume that the functioning of p-sets is sufficiently similar to that of contexts, then the dynamic view risks collapsing into a contextualist one. This could have some benefits, such as the preservation of bivalence (which contextualists keep) and making the view more parsimonious by unifying the approaches to vagueness at different orders.

In what follows, I will defend the dynamic view from this objection. See footnotes for background on contextualism[31] and their solution to the Sorites.[32] The first point that I address is the idea that supervaluation is obsolete. On this view, its role at the first level could be replaced by the context-reminiscent p-sets. The intuitive idea is that, since shifty p-sets account for HOV, why not apply them to FOV and get rid of additional semantic claims and concessions altogether? However, this intuition is misguided, since the supervaluationist solution to FOV is required to make the shifty p-set account of HOV work. This is because the first-order divisions allow for the p-sets to shift in the first place. At the first stage, we implicitly categorize objects into positive, negative, and borderline cases. These categories are directly determined by the p-set, which sets out the supervaluationist truth conditions (i.e., *DF* iff true for all precisifications and so on). These categorizations are provisional: they impose sharp boundaries where none truly exist. This tension allows for future revisions of p-sets, and thus for p-sets to shift. Hence, without supervaluation in the beginning, the p-sets cannot shift. And if they cannot shift, they cannot account for any order of vagueness.

A stronger claim could be made that the p-sets are entirely purposeless if we do not allow for supervaluation. To see the point, imagine that you have some set of precisifications of tall {> 170cm, > 180cm, > 190cm} and you use them to categorize a group of people in the series. Without supervaluation, you end up with six extensions, i.e., three positive and three negative extensions, one per precisification. There are no borderline cases, since without supervaluationist truth conditions — where borderlines are true under some precisifications and false under others — such cases are not defined. Since this is a key symptom of vagueness, as stressed in the beginning, this result would require further explanation of why we think there are borderlines at all.

An enemy of the view could argue that this response misses the point — vagueness did not fail to arise due to the absence of supervaluation, but rather because the p-sets did not shift. After all, on the dynamic

[31] Contextualism rests on the claim that vagueness is a species of context-sensitivity. This roughly means that, in its application across different contextual circumstances, a vague term maintains a constant *character* but shifts in *content*. Therefore, vague terms function like indexical terms. The relationship of vagueness and indexicality is a contested matter for contextualists. Some hold that vague terms behave *like* indexicals, while others claim they *are* indexicals. However, this distinction is not directly relevant to the discussion, and the objections raised here apply equally to both views. Consider the word *now*. It adheres to the same grammatical rules (i.e., has the same *character*) when uttered today and tomorrow. However, when said today, it picks out a different time than it does when used tomorrow (i.e., has different *content*). Similarly, a vague predicate like *tall* is used in the same way when applied to members of a group of pygmy peoples, as when applied to a group of Dutch people. However, it would pick out radically different people. In the first case, the extension of *tall* likely includes some of the world's shortest people; in the second, some of the tallest. See Roy Sorensen, "Vagueness," *The Stanford Encyclopedia of Philosophy* (Winter 2023 Edition), ed. Edward N. Zalta and Uri Nodelman.

[32] Contextualists exploit this idea of unstable extensions over contexts to solve the Sorites by accusing it of equivocating different meanings of a vague term. Similarly to the supervaluationists, the contextualists target the inductive premise (2). The contextualist is committed to the claim of weak tolerance (WT), which states that when two members of a bordering pair are considered in the same context C, they will belong to the same extension. However, WT permits that when one member is considered in context C and the other in C', then they might belong to a different extension. See Jonas Åkerman and Patrick Greenough, "Hold the Context Fixed—Vagueness Still Remains," in *Relative Truth*, ed. Manuel García-Carpintero and Max Kölbel (Oxford University Press, 2010), 275–76.

The WT explains why the inductive premise seems to hold. If we consider any pair in the series, we will conclude that both members belong to the same extension. But this is just because we are disposed to view them in the same context C. The contextualist says that, in fact, the context will gradually change across the series. This means that even if we classify neighbouring terms the same at first, this classification will not persist throughout the series. Thus, the inductive premise of the sorites, such as 'if **n** is short, then **n** + 1 is short', fails since the meaning of 'short' is not the same for every member **n**. This is because, the shift of context C into C', enables cases where '**n** is short' is true (in C) but '**n** + 1 is short' is false (in C'). See J. Åkerman, "Contextualist Theories of Vagueness," *Philosophy Compass* 7 (2012): 470–75.

account, it is the shiftiness of p-sets that allows for HOV. To address this, let us suppose, for the sake of the argument, that the p-set can somehow shift without supervaluation. Imagine, for instance, that the p-set expands by incorporating an additional precisification to the set. You now have eight extensions, yet still no explanation for either first-order or higher-order vagueness. Thus, even with shifty p-sets, the dynamic view cannot function without supervaluation, showing it to be an essential, not merely supportive, component of the account.

Therefore, the case for the contextual collapse breaks down in the very beginning. We simply cannot make the p-sets shifty without maintaining the baseline supervaluationist aspects of the theory. If we cannot make the p-sets shifty, they cannot resolve FOV, let alone HOV. Hence, supervaluation is by no means obsolete. However, to strengthen the defense, I will demonstrate that the next step needed for the contextualist collapse fails. That is, I will show that p-sets and contexts behave very differently.

Although they might appear similar, the former crucially relies on the characterization of vagueness as semantic indecision, while the latter depend on context sensitivity. We might express this difference by saying that the p-sets are inward-oriented, while contexts are more outward-oriented. This is because the former shifts due to our indecision among several equally good precisifications at the initial stage. This indecision prompts us to make mistakes, which we subsequently correct by revising the p-set into another equally acceptable p-set. Thus, the changes directly follow our judgments. By contrast, shifts in contexts seem to have an effect on our judgments - contexts shift first, and judgments follow. Thus, the machinery appears to be quite different.

One could even argue that shifty p-sets rest on a firmer theoretical ground – their shiftiness is caused by our inconsistent judgments. On the other hand, the contexts appear to shift arbitrarily. Thus, the contextualist requires some external justification for this instability. Additionally, the contextualist needs to show how contexts could become shifty enough to prevent every instance of the Sorites. In other words, enough shiftiness must be generated. I do not intend to digress further, but the key takeaway is that despite their apparent similarities, p-sets and contexts differ significantly. Thus, the threat of the 'collapse' does not seem to be so imminent.

As a final point to strengthen my argument, I will provisionally assume that the dynamic approach could collapse into contextualism. Even in such a scenario, there remain independent reasons to prefer the former view over the latter. One significant reason is that contextualism undermines some of our most basic approaches to reasoning. Contextualism requires extensions of vague terms to be unstable, which is precisely what enables it to defeat the Sorites. However, these shifty contexts become deeply problematic when applied outside of the paradoxical setting.

To see this, consider the following example. Saul and Jan are borderline cases of tall. The former is 176.1cm, and the latter is 176cm. Suppose you judge both of them to be tall. Now consider applying the following instance of conjunction introduction:

$$\frac{\text{Saul is tall} \qquad \text{Jan is tall}}{\text{Saul and Jan are tall}} \wedge I$$

However, if the extension of the vague predicate *tall* is unstable, we can easily imagine a situation in which both premises are individually true, yet the conclusion turns out false. This would happen if the context shifted midway through the argument. Thus, although context sensitivity is useful for solving the Sorites, it is dangerous when applied to everyday reasoning. Specifically, how can contexts remain sufficiently stable to ensure our logic does not fail even in such simple cases?[33]

In contrast, dynamic supervaluationism does not provoke such worries. Under supervaluationism, the rule of a conjunction introduction always preserves validity. To illustrate, consider a p-set representing precisifications for *tall*: {>170, >175, >176}. First two precisifications make both premises true and the conclusion true as well. The third precisification makes one of the premises true, the other false, and the conclusion false. This will work for any possible precisification. Consequently, it applies to every p-set.[34]

[33] J. Åkerman, "Contextualist Theories of Vagueness," *Philosophy Compass* 7 (2012): 475–76, .

[34] This follows the exact same reasoning as that applied to the failure of the inductive premise or the truth of the law of excluded middle discussed in more detail at the beginning of the essay.

One might argue that, similarly to a shifting context, the p-set could shift over the course of an argument. For example, we might initially classify both premises as true (e.g., using the set {>170, >175}, but later we classify the conclusion as false (e.g., shifting to the set {>177, >180}). However, this objection reflects a misunderstanding of supervaluationist semantics, since arguments must always be evaluated relative to a single p-set. If we shifted the p-set to the second one, both premises would become false along with the conclusion. Therefore, the validity of conjunction introduction would remain intact.

Why is this strategy not available to the contextualist? The contextualist could simply deny that contexts can shift in such ways, insisting instead that we always evaluate the premises and the conclusion within a single context. However, this directly contradicts the contextualist's equivocation strategy to the Sorites paradox. That is, the strategy according to which bordering cases may differ in truth value because their evaluation contexts differ. Hence the contextualists need contexts to shift. In effect, they cannot deny that the above scenario is possible. Instead, their strongest response would likely be to argue that such cases rarely happen.

I do not intend to argue that supervaluationism, or its dynamic version, is superior to contextualism. Such a claim is clearly beyond the scope of this essay and perhaps beyond the scope of any single essay. Rather, my point is simply that there are independent reasons to prefer the dynamic view over contextualism. Therefore, the claim that contextualism explains everything that the dynamic view explains - but more simply, and thus more parsimoniously - is clearly not accurate.

Taking stock of these considerations, the collapse argument fails not only at its initial stage but also on all subsequent fronts. Dynamic supervaluationism is by no means contextualism in disguise; rather it is its own theory, deeply grounded in Keefe's original supervaluationist framework.

7. Conclusion

While Keefe's supervaluationism remains an attractive account of vagueness, it ultimately struggles to account for higher-order vagueness. Her adoption of a rigid, Tarskian infinite hierarchy may block Williamson's D^* argument, but at the cost of disconnecting the theory from natural language. Even if, as I briefly explored, she could respond to these problems, adopting an infinite metalanguage hierarchy still leaves Keefe subject to a seemingly unresolvable finite series paradox. I argued that Keefe's account could be dynamized by incorporating ideas from Fu, thereby resolving the finite series paradox and avoiding issues associated with a rigid hierarchy. Yet, the dynamic model itself introduces new difficulties, notably the 'fixed time' and 'collapse to contextualism' problems. To defend the view, I briefly outlined potential replies to these issues, showing that they are not fatal. Dynamizing supervaluationism may not resolve all problems, but it is a promising development of the supervaluationist theory and would be worth elaborating on and defending in future enquiries.

Appendix

Why must Keefe deny the S4 and S5 principles?

(1) The S5 principle: If $\sim DF$ then $D\sim DF$.

(2) The S4 principle: If DF then DDF.

Suppose that (1) and (2) hold and that we have the first-order classification:

(i) DF for definite positive cases.

(ii) $\sim DF$ & $\sim D\sim F$ for borderline cases.

(iii) $D\sim F$ for negative cases.

If (1) holds, it implies that at the second level, DF and $D\sim F$ transform into DDF and $DD\sim F$ (see proofs a and b). That is, the definite positive and definite negative case is definitely definite positive and definitely definite negative, subsequently. If (2) holds, it implies $\sim DF$ & $\sim D\sim F$ $\sim DF$ & $\sim D\sim F$ transforms into

$D{\sim}DF$ & $D{\sim}D{\sim}F$ (see proof c). That is, the borderline case is definitely a borderline case. However, second-order vagueness would require two more categories – the borderline between positive and borderline (${\sim}DDF$ & ${\sim}D{\sim}DF$) and the borderline between borderline and negative (${\sim}DD{\sim}F$ & ${\sim}D{\sim}D{\sim}F$). As a result, sharp boundaries are drawn between the three categories since there are no cases between them.

Proof a:

$$\frac{DF \qquad DF \rightarrow DDF}{DDF} \rightarrow E$$

Proof a:

$$\frac{D\neg F \qquad D\neg F \rightarrow DD\neg F}{DD\neg F} \rightarrow E$$

Proof c:

$$\dfrac{\dfrac{\neg DF \rightarrow D\neg DF \qquad \dfrac{\neg DF \neg D\neg F}{\neg DF}\wedge E}{D\neg DF}\rightarrow E \qquad \dfrac{\dfrac{\neg DF \wedge \neg D\neg F}{\neg D\neg F}\wedge E \qquad \neg D\neg F \rightarrow D\neg D\neg F}{D\neg D\neg F}\rightarrow E}{D\neg DF \wedge D\neg D\neg F}\wedge I$$

Bibliography

Åkerman, Jonas. "Contextualist Theories of Vagueness." *Philosophy Compass* 7 (2012): 470–80.

Åkerman, Jonas, and Patrick Greenough. "Hold the Context Fixed—Vagueness Still Remains." In *Relative Truth*, edited by Manuel García-Carpintero and Max Kölbel, 275–288. Oxford: Oxford University Press, 2010.

Alchourrón, Carlos E., Peter Gärdenfors, and David Makinson. "On the Logic of Theory Change: Partial Meet Contraction and Revision Functions." *The Journal of Symbolic Logic* 50, no. 2 (1985): 510–30. https://doi.org/10.2307/2274239.

Bobzien, Susanne. "In Defense of True Higher-Order Vagueness." *Synthese* 199, no. 3–4 (2021): 10197–10229.

Fine, Kit. *Vagueness: A Global Approach.* Rutgers Lectures in Philosophy Series. New York: Oxford Academic, 2020. Online edition, May 21, 2020. https://doi.org/10.1093/oso/9780197514955.001.0001. Accessed November 15, 2024.

Fu, Hao-Cheng. "Saving Supervaluationism from the Challenge of Higher-Order Vagueness Argument." In *Philosophical Logic: Current Trends in Asia*, 139–52. 2017.

Greenough, Patrick. "Higher-Order Vagueness." *Proceedings of the Aristotelian Society, Supplementary Volumes* 79 (2005): 167–90. http://www.jstor.org/stable/4106939.

Hyde, Dominic. "Why Higher-Order Vagueness Is a Pseudo-Problem." *Mind* 103, no. 409 (1994): 35–41.

Keefe, Rosanna. *Theories of Vagueness.* Cambridge: Cambridge University Press, 2000.

Keefe, Rosanna. "Vagueness: Supervaluationism." *Philosophy Compass* 3, no. 2 (2008): 315–24.

Kripke, Saul. "Outline of a Theory of Truth." *The Journal of Philosophy* 72, no. 19 (1975): 690–716. https://www.jstor.org/stable/2024634. Accessed November 30, 2024.

MIT OpenCourseWare. *Session 11: Mathematics for Computer Science. 6.042J: Mathematics for Computer Science, Spring 2015.* Massachusetts Institute of Technology, 2015. Accessed December 8, 2024.

Sainsbury, Mark. "Concepts without Boundaries." Chapter three of *Departing From Frege*. London: Routledge, 1990.

Sorensen, Roy. "Vagueness." *The Stanford Encyclopedia of Philosophy* (Winter 2023 Edition), edited by Edward N. Zalta and Uri Nodelman. .

Williamson, Timothy. *Vagueness.* London: Routledge, 1994.

Defending Williamson's Explanatory Challenge to Contingentism[1]

Koda Li, *Brown University*

1. Introduction

In his book *Modal Logic as Metaphysics*, Timothy Williamson developed a series of arguments against contingentism and in favor of necessitism. I outline the two theses in the following:

(Contingentism) $\Diamond\exists x\Diamond\neg\exists y x = y$

Informally, some things could have not existed.

"The table could have been destroyed in the making process and therefore does not exist."

(Necessitism) $\Box\forall x\Box\exists y x = y$.

Informally, everything necessarily exists.

"This table, the person John, and all other things exist necessarily."

Williamson's arguments are complex and intricate. This paper will focus on one particular challenge he raised to contingentism in Chapter 6 of his book and various responses toward this challenge. The paper is structured as the following: Section 2 reconstructs Williamson's challenge; Section 3 explains two "trivialization" worries about this challenge and respond to them on Williamson's behalf; Section 4 develops a substantive response to Williamson's challenge and criticize it on Williamson's behalf. I argue that Williamson's challenge is successful and contingentists have considerable dialectical disadvantages.

2. Williamson's Explanation Challenge to contingentism

Williamson raised a challenge to contingentists who accept (Comp_M) in high-order modal logic.[2] Below is (Comp_M):

(Comp_M) $\vdash \exists X\Box\forall x(Xx \leftrightarrow A)$

where A is a metalinguistic variable ranging over formulas.

Informally, (Comp_M) says that for any formula A, there is some property that something instantiates just in case A is true.

I will first say something to motivate (Comp_M) before getting into Williamson's challenge using this principle. For one, (Comp_M) is a very attractive higher-order logic principle, for it says roughly that given any formula A, one can define a property such that necessarily, something has it just in case A is true. Intuitively, this seems true. We frequently define complex properties using this way. Given an open formula, for example, "x is white and x is big", certainly there is some property P such that necessarily, a thing y has P iff y is white and y is big. In other words, it seems that we should be able to use any formula A to give the necessary and sufficient conditions for something having a certain property.

Further, we need (Comp_M) to capture compelling natural language inferences, for example the following:[3]

[1] I am grateful to Prof. Eric Guindon, Shichen Wang, Caleb Camrud, and Claire Robertson for very helpful discussion on issues in the paper. I thank anonymous reviewers at Aporia for their feedback on the paper. Finally, I am grateful to my very good friends for their constant support.

[2] The background logic Williamson assumes is the one developed in his Chapter 5, p. 225. What is of significance is that the underlying modal logic is S5. This paper will not tap into the debate of which modal logic is the correct modal logic. I will assume Williamson's logic and develop challenges and responses.

[3] Timothy Williamson, *Modal Logic as Metaphysics* (Oxford University Press, 2013), p. 227.

P1. Alice doesn't smoke a cigar, but she could have done so. $(\neg Sa \wedge \Diamond Sa)$
C. Alice doesn't do something she could have done. $(\exists X(\neg Xa \wedge \Diamond Xa))$

This inference is valid. To capture this, we need precisely an instance of (Comp_M): $\exists X\Box\forall x(Xx \leftrightarrow \neg Sx \wedge \Diamond Sx)$.[4] Finally, more generally, (Comp_M) is an example of comprehension principles for higher-order logic (even for non-modal logic). Standard second-order non-modal logic usually has comprehension principles of similar form: given any formula A, $\exists P\forall x(Px \leftrightarrow A)$. This ensures that the logic has enough power to prove important theorems that intuitively needs to be provable. For example, second-order Peano arithmetic typically contain the following Induction axiom:

(Induction Axiom) $\vdash \forall P\forall x(P(\mathbf{0}) \wedge (P(x) \rightarrow P(x+1)) \rightarrow \forall x P(x))$

Now suppose I have the following formula: x is even or x is odd. Certainly, every natural number has this property: being either even or odd. However, the formula itself cannot instantiate the induction axiom given above, as it is a formula not a predicate. With the comprehension principle, we have: $\exists Q\forall x(Qx \leftrightarrow x$ is even or x is odd). Then we can fix on this property Q and use it to instantiate the induction axiom.[5] This shows again that (Comp_M) is not some novel/strange principle that Williamson cooked up but a typical example of logical principles in higher-order logic. So to sum up, (Comp_M) is a very natural and useful logical principle that we want to add to our higher-order modal logic.

Now we can move on to reconstruct Williamson's challenge. Suppose we instantiate A with $x = y$. We will derive the following:

(Haecceity) $\Box\forall y\Box\exists X\Box\forall x(Xx \leftrightarrow x = y)$ [6]

Informally, this says that necessarily everything necessarily has some property such that having this property is necessary and sufficient for being that thing. This property (of necessary and sufficient for being this thing) can be called the haecceity of that individual, following previous literature.

Next, we can introduce some terminology: let $Haec(X)(y)$ abbreviate $\Box\forall x(Xx \leftrightarrow x = y)$, informally, "X is the haecceity of y"; $Tra(X)(y)$ abbreviates $Haec(X)(y) \wedge \neg\Diamond\exists z(Haec(X)(z) \wedge y \neq z)$, informally "X tracks y". Then, we have:

(Tracking) $\vdash Haec(X)(y) \rightarrow \Box Tra(X)(y))$[7]
Informally, "my haecceity necessarily tracks me."

Then consider an individual o (say, John). By the above theorems, we have:

(o-Haecceity) $\vdash \Box\exists X Haec(X)(o)$[8]

Putting the above two theorems together, we can derive:

(o-Tracking) $\Box\exists X Tra(X)(o)$

Now the challenge according to Williamson is this:

> "Even if I had never been, [...], there would still have been a property tracking me (and only me). But how can it lock onto me in my absence? In those circumstances, what makes me rather than something else its target?"[9]

[4] One might argue that this valid inference can be equally captured by adding an existential generalization axiom to the logic. I just want to point out that this EG axiom is in the exact same spirit as (Comp_M) here: they are both saying that we can form complex properties from simpler ones. So they are not in tension: if one accepts one, one should have reasons to accept the other.

[5] I am using a very informal argument here to motivate and illustrate the use of comprehension principles. For one, "even" and "odd" are not primitive in the formal language of arithmetic, but must be defined. For another, the exact proof does not go the way the informal illustration went. However, these are technical details irrelevant for illustrating the use of comprehension principles, so I will not go into them here.

[6] Necessitation is: if $\vdash A$, then $\vdash \Box A$; Universal Quantifier Rule is: if $\vdash A$, then $\forall x A$.

[7] I include a full proof in the Appendix (Section 6), and say more about the significance of the proof.

[8] We here instantiate (Haecceity) with o.

[9] Williamson, *Modal Logic as Metaphysics*, p. 269.

In other words, there is a challenge to contingentists who accept (Comp_M) to explain how the haecceity of an individual can track this individual in a situation where that individual does not even exist. More intuitively, one might identify haecceities ostensively: when John is here, I can point to him, and say *the property of being John, that person*. However, in a case where John does not even exists, how can you identify such a property? How can a property in that situation manage to behave like a haecceity of John? Even if some property manages to do that, what can possibly explain why it necessarily targets this non-existing individual but not some existing individual? What would the identity condition be when comparing an existing individual and non-existing one? Put in more formal terms, the contingentist needs to explain why (o-tracking) is true while $\Diamond\neg\exists y o = y$. The same challenge can be given for anti-haecceity of individuals, the property X such that $\Box\forall x(Xx \leftrightarrow x \neq y)$. I will not reiterate the argument here. I call this challenge the Explanation Challenge since it is demanding contingentists to offer an explanation of some sort about consequences of their view.

I will end this section with a final clarification note on the broader dialectical situation in Williamson' book. The above Explanation Challenge is what Williamson deemed as "the first horn" in a dilemma for contingentists. The "other horn" is when contingentists attempt to weaken (Comp_M), which is a natural response if one finds the Explanation Challenge a genuine problem. Williamson in the second half of the chapter argued that this weakening also faces serious problems. Thus, the Explanation Challenge is only a part of a larger argument against contingentism. I do not attempt to survey and evaluate the other horn in this paper.

3. Two worries about Williamson's challenge

3.1. The Minimalist Response

The minimalist response is motivated by the intuition that there is not really much to explain. In other words, they want to insist that some metaphysical claims do not require substantive explanation. This is a response *on behalf of* contingentists adopting (Comp_M). There are two specific strategies implementing this response: (i) insisting that no explanation is required, and this does not render contingentism in a dialectically weaker position; (ii) insisting that there is a trivial explanation, and so contingentism again does not fair worse. I will develop these two strategies in more detail and respond to them on Williamson's behalf.

Strategy 1 can be developed in the following, Williamson pointed out that individuals can have contingent existence, but there is always a property tracking them. However, this could be seen just as a brute fact of the modal structure of the world and requires no further explanation. Contingentists do not need to be impressed by this phenomenon at all. This response can be generalized to respond to Williamson's challenge concerning the asymmetry between first-order and higher-order necessitism. The challenge is that given higher-order comprehension principles like (Comp_M), one can prove higher-order analogues of first-order necessitism (which is shown in Section 1), like the following second-order version: $\Box\forall X\Box\exists Y\Box\forall x(Xx \leftrightarrow Yx)$.[10] Thus, contigentists will need to endorse this systematic asymmetry between first-order and higher-order claims. A contingentist can just say that this is exactly what they adopt, and the consequent asymmetry is just a fact that does not call for any further explanation. It is worth emphasizing that this minimalist should not be thought of as "resisting" or "refusing" to explain (o-tracking), but does not see the need to explain in the first place.

I find this minimalist response unconvincing. My criticism will be different from Williamson's, so I will not reiterate his arguments here. First, adopting a minimalist response does not refute or falsify contingentism. It is just that in this dialectical context, contingentism will look much less attractive because there is an alternative theory that has a perfectly simple explanation of (o-tracking). More generally, not being able to explain something is no defeat for the theory (probably every theory has something that it has not yet explained), yet in comparing theories, a phenomenon that one theory can readily explain while the other cannot certainly favors one over the other. In this case, necessitism has a very simple explanation of (o-tracking): just as the usual case, they can point to o which exists necessarily, and identify the property of *being that thing*.

[10] Williamson, p. 264.

Here is an analogy with the Supervenience Challenge to metaethical non-naturalism, the thesis that moral properties are sui generic non-natural properties. The Supervenience Challenge is also an explanatory challenge. The Supervenience Thesis (abbreviated as "(Supervenience)") of the moral properties on the natural properties claims that two objects cannot differ in their moral properties unless they differ in some natural properties.[11] Here the challenge for non-naturalism is to explain why (Supervenience) holds. The key is not the first box since that is usually understood to be conceptual necessity but the second box representing metaphysical necessity.[12] In other words, why the instantiation of natural property necessitate the instantiation of some non-natural property? Just like the minimalist sketched above, some non-naturalists have tried to argue that (Supervenience) does not need an explanation. It is just a fact about the metaphysical structure of the world. This quietist response does not falsify non-naturalism. It just puts non-naturalism in a dialectically weaker position, especially when there are alternative theories which offer an explanation: naturalism does this by identifying moral properties and natural ones. The upshot is that the force of the Explanation Challenge does not derive from posing a counterexample/contradicting contingentism but identifying a source of explanatory weakness.

Further, I think when phrased in terms of explanation, the burden will be on to contingentism to say why the phenomenon does not demand explanation. Here is a parallel in the sciences. We encounter some natural phenomenon: water freezes in winter, leaves fall down in the fall, etc. The default is that these all call for explanation. The only exception might be that when we get to the most fundamental level of nature: only when we get to the fundamental particles can we say: those particles just have those properties they have, by nature. There is nothing more we can say. It is simply bad science if one look at a macroscopic phenomenon and just say it is just there and requires no special explanation at all.[13] The same applies for modal metaphysics as long as it aspires to be (explanatory) science. The default is to assume every modal facts about ordinary objects require explanation, and only when we get to the most fundamental level can we resist the explanation demand. The upshot of all these is that a minimalist adopting strategy 1 will simply be doing bad metaphysics.

Strategy 2, which is more interesting, can be developed in the following way. Contingentists can accept that there is an explanatory demand but argue that there is a trivial explanation. Specifically, contingentists can argue that higher-order necessitism and tracking follows *logically* from $Comp_M$ and the background logic, and that explains why there is tracking and "locking on to individuals." Again, there is nothing further to it. To take an analogy, suppose one was asked why he believes that A and B, he can answer: "I believe A, and I believe B, so I believe the logical consequence of my beliefs, namely A and B." He has indeed provided an explanation of his belief in the conjunction, though a trivial one. Further, the contingentists can even use Williamson's argument in the latter half of the chapter: Williamson discusses how (Comp_M) is the superior comprehension principle in higher-order logic and various technical reasons why one wants to adopt this as part of the logic rather than weaker principles: simplicity, elegance, and enough expressive power to serve logical/metaphysical purposes. Thus, contingentists can even maintain that they have independent reasons to adopt (Comp_M) and argue that (o-tracking) is a logical consequence.

I find this strategy to be problematic, too. Firstly, it seems that the logical consequence explanation does

[11] For a more concrete example, we can imagine John and Bill, who are students in the same class. They both arrive at class on time, handed in assignments on time, etc. Now if the teacher start to punish John for alleged moral reasons, he is rightly to object that the teacher's moral assessment is groundless: what could possibly distinguish him from Bill morally? For a more abstract example, one can imagine John in our world and John' in another possible world. Suppose they do exactly the same things and have the same intentions, etc. It seems that they must receive the same moral evaluation (whether that is virtuous or evil): what could possibly distinguish John from John' morally?
Here is (Supervenience) formulated in higher-order logic just to draw out the analogy with the current case more clearly:

(ST) $\quad \Box\forall X(\mathit{Moral}(X) \rightarrow \forall x(Xx \rightarrow \exists Y(\mathit{Natural}(Y) \wedge Yx \wedge \Box\forall y(Yy \rightarrow Xy))))$

[12] Note that contingentists cannot appeal to conceptual necessity or facts of meaning to explain (o-tracking) since all of the boxes in the theorems refer to metaphysical necessity. At least as Williamson framed the debate, contingentism and necessitism are full-blown metaphysical theories about the world. I suspect that there are ways to think about this debate using conceptual methods, which will be beyond the scope of this paper but interesting to explore.

[13] While for daily life/practical purposes, this attitude is entirely justified, it is not for scientific purposes. Otherwise, it is hard to see how explanatory science can ever get started. Consider vision science. One basic question is, what explains our visual capacity? If a person comes along and says "Well, I can see those things, and not some other things. That's just how I am evolved. What else is to explain?" That is just bad vision science.

not give us a deep-enough explanation. Suppose a contingentist does offer this explanation; a necessitist can just inquire further for an explanation of (Comp$_M$). Why is that for any condition whatsoever there exists a property such that that condition holds just in the case it is instantiated? Notice now, contingentists cannot use the Williamsonian justification for (Comp$_M$), since that will answer the wrong question - why we should *believe in/accept* (Comp$_M$). Thus, the explanatory demand just gets pushed further back. In general, one can always explain p by saying that it is logical entailed by some q. But then the explanatory demand just got pushed back to why q. Of course, if a theory has to push back infinitely, then it is not a good theory.

Secondly, I worry that logical consequence is too lax an explanatory basis; that it generates bad explanations. Here is an example that I have in mind:

> (Racist Explanation). The Racist believes that every member of race A is evil. Consequently, he believes that a member of A o is evil. When asked why he thinks o is evil and consequently refused to offer o equal payment/respect as other employees, the Racist says, "Well, this is a logical consequence of my belief. What more do you want me to explain?"

There is something wrong with both explanations.[14] One, what validates the Racist's universal belief is precisely the character of each individual. Thus, the Racist is not entitled to appeal to this universal belief as an explanation of his specific belief. Two, the Racist's belief attributes some structure to human beings in general and to particular individuals like o. He cannot simply ignore this structure by appealing to the general principle that attributes this structure. He needs to look at this specific person and explain why the "evil" structure is really there, as his theory claims it is. One can now see the same problems hold for contingentism too: (i) (o-tracking) is supposed to support (Comp$_M$), not the other round. Whether contingentism should accept (Comp$_M$) partially depends on whether they find its consequences compelling. The logical consequence explanation gets the order of explanation wrong; (ii) contingentists by adopting (Comp$_m$) attribute some metaphysical structure to individuals (like (o-tracking)), and they now need to find actual features of the world/individuals that support this structure in order to vindicate their theory. In either case, contingentists are not done. To sum up, logical consequence cannot serve as a good explanation. Consequently, Strategy 2 fails.[15]

3.2. **A worry about the notion of explanation**

The following worry is not so much directly on behalf of contingentists but trying to undercut Williamson's entitlement to raise the Explanation Challenge in the first place.

There are two parts to this worry. Firstly, one might be skeptical of the notion of explanation evoked here. At a first pass, the explanation demanded seems metaphysical.[16] However, Williamson is skeptical of notions like grounding and truth-making (the often-evoked notion in metaphysical explanation), so they cannot be used to give metaphysical explanation. Thus, how is Williamson's demand for a metaphysical explanation (potentially using these notions) from the contingentists legitimate? To put it in another way, suppose Williamson thinks that the notion of "metaphysical" explanation (as distinct from physical/causal, social, mathematical, etc) is bogus, and there is no genuine metaphysical explanation at all. Then, trivially, there is no genuine metaphysical explanation for the truth of (o-tracking) and (Contingentism). So how

[14] Of course the Racist is subject to my previous challenge as well: I can simply ask the Racist why he believes that every member of race A is evil in the first place. He is not discharged of explanatory demand. But here I am developing a different problem for the Racist.

[15] An anonymous reviewer asked whether there are previous contingentists literature defending the worry that I criticized in this section. The reviewer asked because my criticism of this contingentist response seems sweeping, and so this response may seem like a "low hanging fruit objection that contingentist philosophers would, in general, avoid". There are several things to say:
Firstly, the relevant literature does not seem to have focused particularly on the Explanation Challenge and the role of explanation. Second, although my objections can seem comprehensive, I think they only scratch the surface of the relevant problems related to explanation, the need for explanation, etc. I am sure there are insights that contingentists can bring in from the literature about explanation in general (and philosophy of science etc) to resist my arguments. However, my goal in this paper is to articulate more clearly a worry that was raised in the seminar discussing this book and give some compelling response to it. So I did not get into potential larger debates about explanation and related notions.

[16] It certainly is not causal or physical: there is no causal relationship or "physical" process in place. It is not constitutive either: nothing seems to constitute another. Normative explanation is certainly not what is being demanded: they are descriptive claims through and through.

can he demand contingentists perform this task? Perhaps more strongly, he should actually be happy with contingentists giving no metaphysical explanation as that is exactly what they should be doing.

Secondly, one might be skeptical of whether Williamson is entitled to this challenge given his methodology. Williamson is very insistent on evaluating metaphysical theories by appealing to their simplicity, strength, and elegance. One of the main threads of his book argues that necessitism allows us to have much simpler semantics for quantified modal logic by allowing us to treat possible world models realistically, adopt simple elegant axioms, avoid complex restrictions in the proof theory, and have an expressive high-order language. Those notions like simplicity/strength are easily demonstrated by observing the kind of formal apparatus present. In contrast, the notion of explanation is comparatively murky.

I think both worries can be dissolved. Regarding the first worry, Williamson can respond in the following ways. One is that we do not have to fix on the nature of explanation prior to giving one. There is an intuitive grasp of explanation that one can rely on. Consider the Supervenience Challenge again. There, a "metaphysical" explanation is also demanded. But one does not necessarily have to give a "metaphysical" explanation in terms of grounding, metaphysical laws, essences, truth-making, etc. Naturalists just say that moral properties are identical to natural ones, so (Supervenience) is a trivial truth; expressivists just say that I cannot explain (Supervenience) as it is about "inflated" properties, but I can explain why people have good reasons to commit to (Supervenience).[17] Those explanations are in no way "metaphysical". Similarly, the explanation of (*o*-tracking) does not need to end up being distinctively metaphysical.[18] In fact, it is actually the *contingentists* for whom we might have this kind of worry about problematic "metaphysical" explanations, because they presumably need to give some "metaphysical" explanations for (*o*-tracking), just like one might be skeptical of non-naturalism when non-naturalists invoke all kinds of metaphysical notions/relations (normative laws, grounding, normative essences, hybrid properties, etc) in their explanation of (Supervenience).[19] Thus, in fact, having to recourse to inflated metaphysical explanation is a disadvantage for contingentists, not for Williamson.

Regarding the second worry, I think the best response is to argue that theory comparison cannot *all* be about simplicity, strength, and elegance especially when the theory is about the actual world. This is true for scientific theories. Comparison between physical theories cannot all come down to which one has more simplicity, strength, and elegance. These formal criterion are very important and rules out many strange theories, but they cannot be all there is to it. Physicists look at fit with empirical data, explanatory power, etc. One cannot decide between Newtonian mechanics or Relativisitic physics just on their formal features. Here is a more specific version for contingentism vs. necessitism. One cannot just decide between them based on formal features. Take the argument that necessitism provides a simpler formal semantics for quantified modal logic. Whether this is true depends on what language one is speaking. Suppose one is speaking a contingentist-commiting language, then of course contingentist semantics will be a better fit and simpler.[20] A similar reverse argument applies to logical strength. Contingentists might say if you adopt all those principles, it will vastly over-generate statements that are false or suspicious so one has to block their derivation. That surely does not look elegant. The upshot is that for empirical theories, we need to look at their empirical content and whether that fits with the world in the right way. Thus, Williamson need not only appeal to formal notions to adjudicate between theories. He is completely entitled to appeal to

[17] Gibbard had given one such account. Now whether expressivists are thereby discharged is actually not trivial. See Dreier, 2015. Dreier argues that quasi-realist formulation of expressivism, which one may reject, is not off the hook. However, this is beyond the scope of this paper.

[18] The necessitists' explanation is not metaphysical at all. One can just imagine the counterfactual circumstance and point to the existing object, concrete or not, and say the property of being that object.

[19] To illustrate, Stephanie Leary posited hybrid properties that have both normative and natural essences to explain (Supervenience) — see Leary, 2017. In general, I think that her additional ontology is not independently motivated and appears very ad hoc as it does not handle any other major (explanatory) challenges to non-naturalism like explaining our moral knowledge or making sense of alien communities and their "moral" language. Of course, these objections need to be further elaborated and defended, which is beyond the scope of this paper. The point I am making here is that one can *easily* come to worry about whether these metaphysical entities/postulates are real, whether they are really motivated, whether they really explain anything, etc. The upshot is that these (potential) worries about non-naturalist evoking heavy-duty metaphysical notions are exactly parallel to worries about contingentists evoking heavy-duty metaphysical notions.

[20] To see the challenge: consider religious discourse. There is this word "God" that appear in the discourse very often. What is its meaning? In some sense, a semantics that assigns it a supernatural being is the simplest as it will validate all the sentences in which the word occur, compared to a semantics that does not. But of course, this cannot be an argument for the existence of "God".

explanatory power and fit with data to evaluate the two theories. This accords with Williamson's general methodology in any case: modal metaphysics is a science.

4. An Anti-Haecceitist response

In this section, I explore a substantive contingentist response to Williamson's challenge. I argue that if true, this successfully answers Williamson's challenge, but this response ties contingentism to another controversial metaphysical doctrine.

The response is to first adopt anti-haecceitism and then use that to explain (o-tracking) even when o does not exist. Anti-haecceitism (abbreviated as AH) is roughly the thesis that there are only qualitative properties (consequently all seemingly non-qualitative properties are reducible to qualitative ones). Qualitative properties are everyday familiar properties like volume/size, color, material composition, or relational properties (*being the mother of*, etc).[21] Non-qualitative properties are harder to describe directly for they are intended to pick out precisely those properties that are not qualitative. Perhaps the most intuitive example is the property of *being this very object*.[22] The point of debate is whether for a particular individual o, the property of being o is qualitative or not. AH says yes, haecceitists say no. In other words, AH maintains that this property of *being o* can be reduced to a complex qualitative property like the property of *being the individual that occupies this location, has this height/weight/taste/wealth, ...*. Now we can see the AH response to Williamson's challenge. Suppose that o's haecceity X can be reduced to this complex qualitative property built out of simple qualitative properties. Suppose further that o does not exist in some counterfactual world. In this situation, very plausibly those simple qualitative properties still exist and so does this complex property. Then indeed there will be a haecceity of o, namely this complex property, and it locks on to o because by stipulation whatever instantiates it will be identical o instead of identical to other objetcs. o is "defined" by all and only these properties.[23] Note that o does not have to exist: this complex qualitative property is just not instantiated in this situation. So AH gives contingentism a *substantive* explanation of why o can fail to exist while (o-tracking) is still true.

One might be tempted to say that AH targets the wrong explanandum. The idea is that AH response is phrased in terms of properties and individuals instantiating them, but Williamson's (Comp_M) and (o-tracking) are all formulated in higher-order logic. Further, if one thinks that higher-order quantification is not first-order quantification over objects of higher-type, then the AH response has failed to explain the right thing. They have explained why properties of a certain kind track the individual, but they have not explained why $\Box\exists XTra(X)(o)$ given $\Diamond\neg\exists y o = y$. However, this response is not successful since one can easily formulate AH in higher-order logic: (AH) $\forall x\forall X(Haec(X)(x) \rightarrow Q(X)$ where $Q(X)$:=X is qualitative. Then one can reformulate the AH response above in those terms. In fact, after the formulation in higher-order logic, there can be a very clear path to a successful explanation. Suppose one has discovered that:

$\Box\forall x(x = o \leftrightarrow P_1x \wedge P_2x \cdots \wedge P_nx)$ where P_i is a qualitative property for any i.

Then one can form a complex predicate P_{total} such that:

$\Box\forall x(x = o \leftrightarrow P_{total}x)$ where $P_{total} := \lambda x.P_1x \wedge P_2x \cdots \wedge P_nx$.

Then one can easily, by Existential Generalization, derive:

$\exists X\Box\forall x(x = o \leftrightarrow Xx)$

[21] Robert Stalnaker, *Mere Possibilities: Metaphysical Foundations of Modal Semantics* (Princeton University Press, 2012).

[22] I think it is hard to identify non-qualitative properties in an entirely theory-neutral way, since AH precisely denies their existence! However, generally, non-qualitative properties are motivated by thought experiments like the following: Consider a possible world where there are only two balls. They are identical in every aspect: shape, size, color, material, etc. Suppose one brings in spatial locations or relationships to an observer, then one is "illegally" bringing in distinct qualitative properties or assuming the existence of an observer that the thought experiment stipulated not to exist. Thus, there seems to be no way of distinguishing between them but through properties like *being this very ball* and *being that very ball*.

[23] Here o does not have to be identified with those properties. So the AH explanation does not have to commit to a bundle theory of individuals.

Williamson has various responses to his anticipated AH response. Against a purely qualitative conception of properties, he said:

> "Thus the purely qualitative conception of properties may well require a highly contentious form of the identity of indiscernibles for individuals, on which qualitative identity entails numerical identity. That is a far less plausible claim than the trivial form of the identity of indiscernibles that permits non-qualitative properties such as identity with y. We have no serious evidence against the metaphysical possibility of a symmetrical universe in which every individual can be reflected (rotated, translated) onto its qualitative double."[24]

However, I think this response targets AH as a self-standing doctrine rather than the AH explanation of the Explanation Challenge, which I will come back to.

Against the previously developed AH response, he said:

> "The theory becomes still more elaborate once fitted out with an account of the persistence of individuals across times and possibilities, since an individual typically has many of its purely qualitative properties, such as shape and size, temporarily and contingently. Alternatively, if the theory denies identity through change and through contingency, not only is that yet another implausible consequence, which requires still more theoretical complexity to save the appearances, it also fits badly with the underlying motivation for contingentism, by treating a vast range of apparent contingency as an illusion. Thus the purely qualitative conception drags the contingentist into proliferating complications of metaphysical theory with no independent plausibility."[25]

I find the remarks here sketchy, too. For one, I am not sure if AH should regard having contingent qualitative properties as a problem for their theory. Either they can say that the qualitative properties include modal and temporal properties (so *being John* involves *being a possible lawyer*) or they can index haecceities such that one should really say: X is a haecceity of y at w, t. In this second way, the original theorems are still in place because, in any world X is still a haecceity of y at a particular indice w, t.[26] Further, and more importantly, I do not think his challenge shows that the AH explanation is unsuccessful. Williamson at best shows that AH has various undesirable consequences. Thus, overall, I think Williamson's own remarks in the chapter do not constitute the right kind of dialectic challenge to the AH response.

However, Williamson's skepticism brings us to what I think is a more successful general response to the AH explanation. The idea is to admit that the AH strategy does respond to Williamson's challenge, and to point out that this AH response forces contingentism to accept a controversial metaphysical doctrine, thereby having to accept its dialectical challenges.

Firstly, adopting the AH strategy forces contingentists to be anti-haecceitists, thereby having to answer challenges to anti-haecceitism itself. Contingentists need to accept AH to give the AH explanation. Granting that AH does offer a successful explanation of (o-tracking), we can say that if AH is indeed true, then we can solve Williamson's Challenge. Consequently, the plausibility of the theory package consisting of solving Williamson's Challenge and AH - will be the plausibility of AH itself. Assuming that Solving Williamson's Challenge amounts to vindicating contingentism, the plausibility of this version of contingentism (which will be the conjunction of AH and (Contingentism)) will be the plausibility of AH itself.[27] Now there is a problem if AH is not very plausible. If AH is not a compelling metaphysical thesis in the first place, contingentism will not fare well having to accept it. This is where many of Williamson's previous charges can be properly incorporated: skepticism about distinction between qualitative and non-qualitative properties, counterexamples from indiscernible objects, etc. The upshot is that now the dialectic cost for contingentism is no longer not being able to explain something but having to accept some controversial/implausible

[24] Williamson, *Modal Logic as Metaphysics*, pp. 271–272.

[25] Williamson, p. 272.

[26] Now I do not want to get into the debate between AH and haecceitis here. These two ways of responding to Williamson's challenge at least seem prima facie available. They could be false in the end, but at least they show AH is not easily defeated.

[27] Readers familiar with probability theory might think in terms of probabilities assigned to these propositions: the conditional probability P(Solving Williamson's Challenge | AH) is roughly 1. So P(Solving Williamson's Challenge $\wedge$ AH) will be equal to P(AH). So P(Contingentism $\wedge$ AH) = P(AH). These follow from basic probability axioms and logic.

doctrine in order to be able to explain something. This seems to be a cost. Contingentists give themselves a greater burden compared to a necessitist that can remain neutral on this issue.

I am personally not very worried about combining a theory with another controversial theory in itself. Intuitively, theories should be allowed to appeal to other resources (like other theories) in developing and defending itself even if those resources are controversial. Denying the legitimacy of this appeal would render theorizing very difficult and limited. We want to establish connection between theories across domains and explore how they can inform each other.[28] However, the real worry is whether this combination is the only viable combination (because we do not know which theory is ultimately right). That is, if contingentism can only be effectively defended relying on a particular theory of haecceities, then it looks less attractive than a view that is compatible with a variety of theories of haecceities. Contingentism is not supposed to be a global thesis that aims to provide complete answers to all metaphysical questions. It is not even aiming to be a comprehensive theory of metaphysical modality. Necessitism is the same. Thus, one would hope that it can remain local instead of having global consequences. However, if it can only be a good local theory when combining with a particular (global) theory, then one should be more skeptical as this local theory seems to demand too much packaged in along with it. Necessitism in contrast is compatible with both haecceitism and AH.[29] Thus, overall, while AH response is a good substantive explanation answering the Explanation Challenge, there will be considerable dialectical cost for contingentists to accept it.

5. Conclusion

In this paper, I have examined three responses to Williamson's Explanation Challenge and argued that each response faces their own problems. While I argue for the stronger conclusion that the first two challenges fail, I argue for the weaker conclusion that the last response succeeds but only with additional dialectical cost to contingentism. I hope this paper has helped to clarify the stake of Williamson's "first horn" to contingentism in Chapter 6 and strengthen his argument against contingentism.

6. Appendix

6.1. The proof for (Tracking)

First, we can observe the following proof:

$$\begin{array}{ll}
\forall x(Xx \leftrightarrow x = y), Haec(X)(z) \wedge z \neq y \vdash \bot & \text{(Reductio, Cond. proof, Universal generalization)} \\
\vdash \Box\forall x(Xx \leftrightarrow x = y) \rightarrow \Box\forall z\neg(Haec(X)(z) \wedge z \neq y) & \text{(K)} \\
\vdash \Box\forall x(Xx \leftrightarrow x = y) \rightarrow \neg\Diamond\exists z(Haec(X)(z) \wedge z \neq y) & \text{(Equivalence)} \\
\vdash Haec(X)(y) \rightarrow Tra(X)(y) &
\end{array}$$

Now I will show the first line.

$$\begin{array}{l}
\forall x(Xx \leftrightarrow x = y), Haec(X)(z) \wedge z \neq y \vdash Xy \leftrightarrow y = y \\
\forall x(Xx \leftrightarrow x = y), Haec(X)(z) \wedge z \neq y \vdash Xy \leftrightarrow z = y \\
\forall x(Xx \leftrightarrow x = y), Haec(X)(z) \wedge z \neq y \vdash y = y \leftrightarrow z = y \\
\forall x(Xx \leftrightarrow x = y), Haec(X)(z) \wedge z \neq y \vdash z = y \\
\forall x(Xx \leftrightarrow x = y), Haec(X)(z) \wedge z \neq y \vdash z \neq y \\
\forall x(Xx \leftrightarrow x = y), Haec(X)(z) \wedge z \neq y \vdash \bot
\end{array}$$

[28] I can give numerous examples. For example, expressivists in metaethics appeals to truth minimalism to recover the legitimacy of ordinary moral talk/thought, even if truth minimalism is controversial; non-naturalists appeals to post-modal/hyperintensional metaphysics in developing their theories, even if notions like grounding/essence invoked in hyperintensional metaphysics are very controversial — see Bengson, Cuneo, and Shafer-Landau, 2024. I think they can make these appeals. Metaethicists have made a lot of progress by doing this. Their theorizing would just be very limited if they cannot do this.

[29] This is exactly the same for expressivism and truth minimalism. Expressivism is meant to be a local thesis about moral language. However, to defend it, one would need to reject truth-conditional semantics (which is incompatible with truth minimalism) in general, then it no longer seems very attractive.

Then, from $\vdash Haec(X)(y) \rightarrow Tra(X)(y)$, we can observe that:

$\vdash \Box Haec(X)(y) \rightarrow \Box Tra(X)(y)$	(K)
$\vdash \Box\forall x(Xx \leftrightarrow x = y) \rightarrow \Box\Box\forall x(Xx \leftrightarrow x = y)$	(**4**)
$\vdash Haec(X)(y) \rightarrow \Box Haec(X)(y)$	(Chaining conditionals)
$\vdash Haec(X)(y) \rightarrow \Box Tra(X)(y)$	

6.2. **The proof for (o-Tracking)**

$\vdash Haec(X)(o) \rightarrow Tra(X)(o)$	(Proved above)
$\vdash \exists X Haec(X)(o) \rightarrow \exists X Tra(X)(o)$	(Derivable from $\forall$ rule)
$\vdash \Box\exists X Haec(X)(o) \rightarrow \Box\exists X Tra(X)(o)$	(K)
$\vdash \Box\exists X Tra(X)(o)$	(MP, *o*-Haecceity)

I include these proofs in detail for two reasons. One, Williamson did not lay out the proof at all in the book. So I think reconstructing it here will help the reader to see clearly how the seemingly strong principles are derived. Second, and more importantly, this proof shows how little background logic is needed to derive the later-shown-to-be-problematic (Tracking). This proof assumes only modal logic principles **4** and **K**, and the usual meta-rules like conditional proof, reductio, etc. Thus, it does not require a strong logic to prove (Tracking). The significance is that, suppose one accepts that (Tracking) has problematic consequences, one thing we can always see is if there is any logical principle we can reject which contributes to the proof. That would be a natural contingentist way out. However, this proof shows that it will not be easy to take this route. K is the least contentious axiom in modal logic; **4** is somewhat controversial, but not very, since intuitively, modal properties/facts should themselves be necessary and not mere accidental. Further, the controversial B axiom that actually bears on the necessitism and contingentists debates are not used essentially. Thus, there is not much reasonable/non-ad hoc room for contingentists to weaken their background logic to escape from Williamson's challenge. Williamson himself does not make this point, but I think it is important.

Bibliography

Bengson, John, Terence Cuneo, and Russ Shafer-Landau. *The Moral Universe.* Oxford University Press, 2024.

Dreier, James. "Explaining the Quasi-Real." *Oxford Studies in Metaethics* 12 (2017): 273–297.

Leary, Stephanie. "Non-Naturalism and Normative Necessities." *Oxford Studies in Metaethics* 12 (2017): 76–105.

Stalnaker, Robert. *Mere Possibilities: Metaphysical Foundations of Modal Semantics*. Princeton University Press, 2012.

Williamson, Timothy. *Modal Logic as Metaphysics*. Oxford University Press, 2013.

A Defence of the Interpretational Account of Validity

Audrey Hammer, *University of Cambridge*

Both the interpretational account and the representational account provide contrasting accounts of validity for natural-language arguments. While the interpretational account captures formal validity, unlike the representational account, it does not capture materially valid arguments. Therefore, materially valid arguments are viewed as counterexamples to the interpretational account. I motivate why we may want to defend the interpretational account over the representational account and then proceed to defend the interpretational account using the suppressed premise strategy. The first objection to the suppressed premise strategy is by Stephen Read, who argues that the supressed premise is redundant. My contribution is to demonstrate how his objection fails. I also discuss and defend the suppressed premise strategy against other objections, which concern the nature of the supressed premise and the problem of modus ponens.

Introduction

Validity, a key concept in logic, concerns whether an argument is truth-preserving. The interpretational account of validity defends the view that for an argument to be valid it must be formally valid. I turn first to the importance of logical form, its role in logic, generally, and validity, specifically. My discussion then moves to the interpretational account alongside its rival, the representational account. Both accounts face distinct issues. While I do not hold that the representational account is incoherent, I do hold that its formulation has weaknesses that are absent in the interpretational account, giving a motivation for preferring the latter rather than the former. Materially valid arguments, which are not formally valid, present counterexamples to the interpretational account. The remainder of the essay is devoted to showing how the suppressed premise strategy can defend the interpretational account against this main objection. The suppressed premise strategy will in turn be defended against pressing objections, primarily Stephen Read's objection that the suppressed premise is redundant. This objection to the supressed premise strategy aims to prove that there is a contradiction in adding a suppressed premise to an already materially valid argument, and my contribution is to show how this objection fails. I then go on to defend the suppressed premise strategy against a few other objections, including objections concerning the nature of the suppressed premise and the argument, and the problem of modus ponens. The result is a defence of the interpretational account of validity, using the suppressed premise strategy.

Understanding the Relation Between Logical Form and Validity

Logic is considered the science of deduction: it deals with arguments and their validity. In formal logical languages, like truth functional logic and first order logic, we can capture validity using the standard notion of logical consequence. A formal argument is valid if the conclusion is a logical consequence of the premises. As Owen Griffiths and Alexander Paseau put it, "A formal sentence ϕ is a logical consequence of a set of formal sentences γ just if every model of γ is a model of ϕ".[1] Thus, we can describe the formal notion of validity for a logical language, using a model-theoretic notion of logical consequence.

Once we have captured the notion of validity for logical languages, we can move on to understanding the concept of validity as applied to natural language, as the accounts of validity that will be discussed are

[1] Owen Griffiths and Alexander Paseau, *One True Logic* (Oxford: Oxford University Press, 2022), 8.

accounts of validity for natural language. To understand validity as applied to natural language, we must introduce the concept of logical form. Logical form is generally considered to be a property of a sentence of natural language. The logical form of a sentence is when, keeping the logical constants fixed, the non-logical expressions get replaced with variables of the appropriate sort. Thus, the logical form of a sentence can be expressed using a schema. Given this schematic representation of form, we can follow Alfred Tarski in the view that logic is topic neutral, because a schema abstracts from the content of the sentence, only retaining the form of the sentence. For example, take the following sentence:

(1) Pigeons wear vests and cats wear hats.

This sentence can be expressed using the following schema:

(2) $A \wedge B$.

This is because the logical expression in sentence (1) is "and" which can be formalised using the symbol "$\wedge$", and the non-logical expressions in the sentence are "pigeons wear vests" and "cats wear hats", and thus these expressions are replaced with variables.

One stipulation with this account of logical form, is that it requires us to have an understanding of what a logical constant is. Thus far formality has been captured by its topic neutrality, and since a demarcation of logical notions is crucial to form, it makes sense to construct this demarcation using this quality of topic neutrality. Here we can invoke Tarski's account of isomorphism invariance. Tarski defines logical notions using an analogy from geometry. Just as we may demarcate particular geometrical objects by their invariance under transformations, so too can we demarcate logical notions. Thus, "we call a notion 'logical' if it is invariant under all possible one-one transformations of the world onto itself".[2] To explain this further, we can consider an isomorphism to be a bijective function, so between two structures there is a one-one mapping, which preserves all the relevant relations. This isomorphism is the transformation that Tarski is speaking of. For a relation to be isomorphic invariant it must remain unchanged over this sort of transformation. A relation that is isomorphic invariant is thus indifferent to individual objects. The only notions that do this are logical notions, and this confirms neutrality. Thus, we can define a logical notion as being isomorphically invariant and non-logical notions as not being isomorphically invariant. This allows for the demarcation, which is necessary to define logical form.

This understanding of logical form can now aid us in capturing the notion of formal validity for natural language. It is common in the literature to equate an argument being formally valid with it being valid in virtue of its form.[3] However, using this as a definition for formal validity is unsatisfactory, for we still need to define being valid in virtue of form, which I find to be no more informative than formal validity. Therefore, I define formal validity to be the following: an argument is formally valid iff it has a form which has only valid instances. An example of a formally valid argument is:

(3) All men are mortal, Socrates is a man $\therefore$ Socrates is mortal.

The logical form of the argument can be captured using a schema, as described above. Given the use of quantifiers in (3), the schema of the argument is simply its first order formalisation (on the obvious formalisation key):

(4) $\forall x(Fx \rightarrow Gx),\ Fa \ \therefore\ Ga.$

There are no invalid arguments with this form, therefore all the instances of this form are valid, consequently the argument is formally valid. It is clear from this explanation that this definition of validity for natural languages coincides with the definition for formal languages, meaning that a natural-language argument is formally valid iff its formalisation is valid.

[2] Alfred Tarski, "What are Logical Notions?," *History and Philosophy of Logic 7,* (1986), 149.
[3] Mark Sainsbury, *Logical Forms: An Introduction to Philosophical Logic*, (Oxford: Blackwell, 2001): 37.

Two Accounts of Validity

We can now examine two model-theoretic accounts of validity for natural-language arguments. Generally, model-theoretic accounts of logical consequence are now viewed as more successful compared to other accounts of logical consequence, and the two accounts that are the focus of this essay are model-theoretic. As such the central thesis of both accounts understands logical consequence as concerning truth preservation across models.[4] The first account is the interpretational account of validity, which originates from Bolzano but was promulgated by Tarski.[5] This account holds that an argument is valid if there are no possible interpretations of the argument (except for a reserved class of logical interpretations) where the premises are true and the conclusion false. An interpretation of an argument is any argument that has the same logical form as the initial argument. The second account is the representational account of validity, which holds that an argument is valid if it is impossible for the premises to be true and the conclusion false.[6]

The interpretational account only accepts arguments that are formally valid. The account achieves this by examining different logical interpretations of the argument; if there is no interpretation that has true premises and a false conclusion then the argument is considered valid. On the other hand, the representational account allows for arguments that are materially valid, alongside those that are formally valid. Materially valid arguments are arguments in which the validity of the argument is in part due to the meaning of the non-logical terms involved. An example of a materially valid argument is:

(5) Jill is a paediatrician ∴ Jill is a doctor.

The representational account intends to capture a more "intuitive" notion of validity. Defenders of this account hold that materially valid arguments are contained within this intuitive notion of validity, and so an account of validity must capture material as well as formal validity. This belief is rooted in the idea that there is an analytic connection between certain words or phrases, and these connections make the argument valid, even though the argument is not formally valid.

The main objection to the interpretational account is that it is subject to counterexamples, which take the form of materially but not formally valid arguments. To establish the success of the interpretational account we must meet this objection. One example of a materially but not formally valid argument is (5) above, and another is:

(6) Adam is taller than Bill and Bill is taller than Cathy ∴ Adam is taller than Cathy.

Neither of these arguments is formally valid, since there are invalid arguments with the same form as (5) and (6). The interpretational account would not accept that they are valid arguments given there are interpretations of (5) and (6) for which the premises are true and the conclusion false. A formalisation of these arguments in first order logic reveals their logical form:

(7) $Fa \therefore Ga$

(8) $(Tab \wedge Tbc) \therefore Tac$

Another interpretation of each of these arguments demonstrates the point further:

(9) Pat is a postman ∴ Pat is a father.

(10) Alice is friends with Bonnie and Bonnie is friends with Carl ∴ Alice is friends with Carl.

[4] This contrasts with proof-theoretic accounts which hold that the nature of logical consequence involves there being a proof from the premises to the conclusion.

[5] Jc Beall, Greg Restall, and Gil Sagi, "Logical Consequence", *The Stanford Encyclopaedia of Philosophy* (Summer 2024 Edition); Stephen Read, "Formal and Material Consequences", *Journal of Philosophical Logic* 23, no. 3, (1994): 249.

[6] Read, "Formal and Material Consequences", 250.

These arguments are clearly invalid, yet they have the same logical form as (5) and (6), respectively. It is due to these alternative interpretations that (5) and (6) are not valid.

However, the arguments (5) and (6) would be accepted under the representational account due to this account's use of modality. The representational account identifies logical consequence with metaphysical consequence. The reference to "impossible" in the representational account is a modal notion, whereas the interpretational account does not include such modal notions. The reference to "no possible interpretations" in the interpretational account may be made actual using substitutional classes, and thus does not need to rely on an analysis of modality.[7] Yet, it is because of its use of modality that the representational account can attribute validity to (5) and (6), for there is no possible world where the premises of (5) and (6) are true and the conclusion false.

On the other hand, modality is an issue for the representational account, for it requires that we have an analysis of modality.[8] Commonly, modality is cashed out in turns of possible worlds. This prompts the question of what a possible world is. The answers to this question are controversial. We have modal realists, like David Lewis, who endorse a view that possible worlds exist, as real concrete entities.[9] Adopting this analysis for our account of validity would also mean adopting the ontological commitments of this account. Other analyses of modality include modal sceptics, who deny that modal statements can be known. In adopting this approach, we could not know whether our arguments are valid, which is entirely counterintuitive. While there are some more modest approaches to modality, like those taken by Stalnaker[10] and Adams[11], there are still issues surrounding whether these accounts can provide a reductive analysis. This is all to say that while modality is often invoked in philosophical topics, the debate surrounding modal notions is not uncontroversial, and thus any time it is invoked in a theory, that theory faces the same controversies. This is not to say that modal notions should never be used in philosophical theories, but just that we should be aware of the commitment and, all things being equal, adopt theories without modal notions. This gives us a motivation to prefer the interpretational account over the representational account. Indeed, Read, who accepts the representational account over the interpretational account, admits that the lack of modal notions in interpretational account is a possible motivation to prefer this account rather than the representational account.[12]

While this general criticism concerning the use of modal notions is important to note, there is a more specific problem with the representational account; namely, the identification of logical consequence with metaphysical consequence then provides no account of the importance of formality in logical consequence.[13] Similarly, the account does not provide a basis for distinguishing between logical and non-logical vocabulary. This is because the representational account determines that all expressions used in the argument contribute to the validity of the argument. Consequently, the representational account undermines the topic neutrality of logic.

Given that the representational account faces the above challenges, I suggest that this should motivate us to adopt the interpretational account instead. While I do not view these issues as being insurmountable, I simply hold that if there is an alternative we should favour it. If the problem of counterexamples to the interpretational account can be overcome, then this account becomes a preferrable alternative to the representational account of validity. I devote the remainder of this essay to considering and defending a possible solution the interpretational account can adopt to resolve the problem of counterexamples. This solution is the suppressed premise strategy.

[7] Read, "Formal and Material Consequences", 252.

[8] It should be noted that this conversation concerns analyses of the metaphysical notion of modality, which is distinct from a discussion of modal logic, which is considered to be well understood. Metaphysical modality deals with the fundamental nature of modal notions, whereas modal logic is a formal system which reasons about sentences containing modal operators.

[9] David Lewis, *On the Plurality of Worlds*, (Basil Blackwell, 1986) 2-3, 86.

[10] Robert C. Stalnaker, "Possible Worlds," *Noûs* 10, no. 1, (1976): 65-75.

[11] Robert Merrihew Adams, "Theories of Actuality," *Noûs* 8, no. 3, (1974): 211-231.

[12] Read, "Formal and Material Consequences", 252.

[13] Beall, Restall, and Sagi, "Logical Consequence."

The Suppressed Premise Strategy

The suppressed premise strategy (hereafter SPS) can be employed by the interpretational account to overcome the problem of materially valid arguments. SPS holds that materially valid arguments have suppressed premises which when revealed make the argument formally valid, and thus valid under the interpretational account. These suppressed premises are true given they usually explicitly reveal true analytic connections between words.[14] Since they are true, the addition of the suppressed premise is largely unproblematic, although this claim will be defended further.

SPS applied to the argument (5) gives:

(11) Jill is a paediatrician, all paediatricians are doctors ∴ Jill is a doctor.

This argument can be formalised as follows:

(12) $Fa,\ \forall x(Fx \rightarrow Gx)\ \therefore\ Ga$

There are no possible interpretations of the argument (11) that will have true premises and a false conclusion, thus under the interpretational account (11) is valid, although (5) remains invalid. Of course, this strategy applies to (6), where the suppressed premise is that "taller than" is transitive. No suppressed premise can be added to (9) or (10), since it is not true that all postmen are fathers, there is no analytic connection between being a postman and being a father, and the relation "being friends with" is not transitive.

The Redundancy Objection

The first objection to SPS is put forward by Read and states that the suppressed premise is either false or redundant, and since it cannot be false it must be redundant. [15] Read gives his argument as follows:

> The extra premise is strictly redundant. For if the original argument were invalid, the added premise would not be logically true. Given that it is logically true, it follows that the unexpanded argument was already valid. Hence it was (logically) unnecessary to add the extra premise.[16]

This objection is best demonstrated using an example. Take argument (5), which is considered invalid under the interpretational account. Read says that because of its invalidity, it is possible for the premises of (5) to be true and the conclusion of (5) to be false. This entails that it is possible for Jill to be a paediatrician but not be a doctor. Yet the suppressed premise for this argument is that "all paediatricians are doctors", clearly contradicts the possibility Jill is a paediatrician and not a doctor. It follows if we accept that (5) is invalid, then we also accept that the suppressed premise is false. Yet this suppressed premise is true, so the initial assumption that (5) is invalid must be false, and therefore the addition of the suppressed premise is made redundant for it is not necessary for the argument to be considered valid. According to Read, the suppressed premise's redundancy means we should reject the interpretational account in favour of the representational account.

Read's objection, while presented convincingly, lacks any actual force. This is due to a key error it makes: it presupposes the representational account, when it should presuppose the interpretational account. It is not the case that (5) is invalid because the premise "Jill is a paediatrician" is compatible with it being false that "Jill is a doctor", which (if true) is what the representational account would suppose, rather (5) is invalid because there is an interpretation of (5) for which the truth of the premises is compatible with the falsity of the conclusion. (9) is an interpretation of (5) for which it is compatible that it is true that "Pat is a postman" and false that "Pat is a father", and therefore (5) is considered invalid under the interpretational account. Under the interpretational account, nothing specifically is said about the premises of (5), and so Read is

[14] Read views these suppressed premises not just as true but as logically true because he associates logical truth with analytic truth (Read, "Formal and Material Consequences", 258). Since I have not made this association, I will avoid understanding suppressed premises as logically true.

[15] Read, "Formal and Material Consequences," 257-9.

[16] Read, "Formal and Material Consequences", 259.

wrong to infer that attributing invalidity to (5) will make the suppressed premise false. Since Read is wrong to assert that the invalidity of the argument shows the suppressed premise's falsity, he cannot then infer that since the suppressed premise is true, it must therefore be redundant. Under the representational account, invalidity is saying something about the specific premises of the argument under consideration. Yet under the representational account a materially valid argument, like (5), would not be considered invalid.

Some may reply here that I am begging the question: why is it that we should assume the interpretational account and not the representational account? However, this line of thought is also mistaken. Read clearly starts by assuming that materially valid arguments are invalid, which is only the case under the interpretational account, not the representational account. From this assumption of invalidity, he attempts to prove a contradiction, but then uses the representational account's understanding of validity in this contradiction, even though the representational account would not attribute invalidity to something that is materially valid. However, if the interpretational account is used, then there is no contradiction in using SPS. In addition, this strategy is only used by the interpretational account. Thus, Read must assume the interpretational account if he is going to show a contradiction; given he does not use the interpretational account in his objection and that even if he did use the interpretational account there would be no contradiction, this implies that his objection holds no weight.

Objections about the Nature of the Suppressed Premise and the Argument

A second problem for SPS is that we have not been committed to the view that the suppressed premise is logically true. This may lead to the question: why is it acceptable to add to an argument an extra premise that is not logically true? Surely only logically true propositions may be added to the premises of an argument to retain the same argument. To answer this question, an important point must be reiterated: I do not agree that the argument prior to the addition of the suppressed premise is the same argument as the argument after the addition of the suppressed premise. To me this point is obvious, for the two arguments have different properties: one argument is valid, the other invalid, and they have a different number of premises. Since we are speaking of two different arguments, I do not need to prove that the first argument is "retained" in the second. However, this does not mean SPS can be used on any argument. If the premise "all postmen are fathers" is added to (9) then we have a new argument:

(13) Pat is a postman, all postmen are fathers ∴ Pat is a father.

(13) is a valid argument, but we should not consider (13) to be using SPS. Therefore, we must identify what differentiates (11) from (13), and why (11) is determined as using SPS and thereby linking it closely with (5) in a way that (13) is not linked with (9). The difference is that the suppressed premise revealed in (11) that "all paediatricians are doctors" is true, but the premise "all postmen are fathers" is not true. Indeed "all paediatricians are doctors" is an analytic truth. However, it is not necessary that this be considered a logical truth. To begin with, there seems to be no necessity to consider analytic truths to be logical truths, particularly if we retain the commonly held view that logic has no special content. And secondly, the goodness of an argument can be characterised by whether it is sound, i.e., it is valid and has true premises, which does not require the premises to be logically true. So long as the suppressed premise is true, its addition to the argument does not hinder the chances of the argument being sound and should in fact improve this since the argument will now be formally valid. Since one of the characteristics of a suppressed premise is that it is true, there is no issue that it is not logically true. Considering (13), the premise "all postmen are fathers" cannot be a suppressed premise of the argument (9) for it is not true. Therefore, the suppressed premise does not need to be logically true, but this does not mean that SPS can be applied to any argument to make it valid.

Moreover, we may consider that SPS might even allow us to consider contingent truths as suppressed premises. Let us suppose that it were a contingent fact that "all postmen are fathers", then it might make sense to consider this to be a suppressed premise of argument (9). Say Mr. Black presented argument (9) to Mr. White and both Mr. Black and Mr. White were aware that "all postmen were fathers", then the argument might be accepted as sound in the rhetoric (even though it is not formally valid) because both understood that the argument has a suppressed premise, and that Mr. Black in fact meant to make the

argument (13). Now suppose Mr. Smith questioned the validity of the argument because he was not aware that it was a contingent fact that "all postmen were fathers". Yet, once this would be revealed to him, Mr. Smith would certainly accept the validity of the argument. Therefore, we may accept that a suppressed premise may be contingently true, and it becomes clear that only truth, and not logical truth, is necessary for the suppressed premise.

A counterexample to this argument has been pointed out to me.[17] This is that if we take the argument:

(14) I am a philosophy student ∴ puppies are cute.

This is clearly invalid. But if the conditional "If I am a philosophy student then puppies are cute" is added as a suppressed premise to (14), then we get the new valid argument:

(15) If I am a philosophy student then puppies are cute, I am a philosophy student ∴ puppies are cute.

It appears there is no problem with adding this conditional if we take the view that suppressed premises only need to be contingently true, and not analytically true, because considered as a material conditional it is true (the antecedent and consequent are true). This seems to be a problem for the strategy, as it might allow for many arguments like (14), that have true premises and true conclusions yet are not formally or materially valid, to be valid by adding these conditionals as suppressed premises.

My response to this argument is to say that these conditionals are indicative conditionals, not material conditionals, which means they involve a different treatment. An indicative conditional is the conditional of natural language, and the current discussion is about the validity of natural-language arguments, so it makes sense to speak of indicative conditionals rather than material conditionals. We may then consider views of indicative conditionals which hold that their truth values are different to those of material conditionals, and as such we can formulate a view that holds that "If I am a philosophy student then puppies are cute" is false. For instance, we might hold that an indicative conditional is true iff it is assertable and is in turn assertable iff it passes the Ramsey test. The Ramsey test is a test for the assertability of a conditional, it holds that a conditional is assertable if someone were to add the antecedent to her set of suppositions, she would also have to add the consequent. "If I am a philosophy student then puppies are cute" would clearly fail the Ramsey test. Thus, we can still consider that the suppressed premise may be true without the above presenting as a counterexample.

I have only given a rough sketch of a possible response to the objection suggested above, and while there are many problems with associating the truth conditions of an indicative conditional with those of the material conditional, there are still some who adopt this view. However, the conditional suggested is one where the antecedent and the consequent are both true and yet have nothing to do with each other. This sort of conditional is itself a problem case for someone who holds this truth-functional view of the indicative conditional, suggesting that there is something wrong with equating the indicative conditional with the material conditional. However, if the reader insists on the indicative conditional and the material conditional having the same truth value, even in cases where the antecedent and consequent have no relation to each other, then this reader may simply choose to reject this section on contingent truth and hold that the suppressed premise must be an analytic truth. This does not detract from the fact that the suppressed premise is not a logical truth. Of course, the reader may still object to the idea of analytic truth. However, this paper defends the interpretational account against the counterexample of material valid arguments, which themselves rely heavily on the notion of analyticity. So, if the reader places no importance on the analytic connections between words, then there is no forceful objection to the interpretational account and no need for SPS to begin with.

A third objection connects to my answer to the second objection. I have stated that the two arguments, the argument prior to the addition of the suppressed premise and the argument after this addition, are two different arguments. This may lead one to ask, "what connects the two arguments?" The answer to this is simple: they both have the same aim. The aim of an argument is an imprecise and informal notion; however, I want to use it to capture an intuitive idea. The two arguments share the same conclusion, and their aim is to

[17] By Owen Griffiths, in personal communication.

use true (and very similar) premises to arrive at this conclusion. Suppose that Jones is having a discussion of Jill's profession; he would be just as happy receiving the argument (11) as he would be receiving the argument (5), possibly even happier receiving (11) if he is unaware that a paediatrician is a kind of doctor (or if he is a logician who has a strong appreciation for formal validity). However, Jones would be disappointed if instead of receiving either of these arguments he received (3), for instance, which clearly has nothing to do with Jill or her profession. The aim of the arguments is informal, and the setting for which Jones might accept or reject them, as described, is also informal. The arguments are connected by this informality. The matter of validity in logic is strictly a formal matter, and thus there is a distinct difference between (5) and (11).

The Problem of Modus Ponens

The final problem I shall explore in relation to SPS is the problem of modus ponens. A modus ponens is a deductive argument of the following form:

(16) $A,\ A \rightarrow B\ \therefore\ B$

Modus ponens is discussed by both Read and Timothy Smiley, in very different ways.[18] They both view modus ponens as having a similar form to SPS but speak of different consequences related to this similarity. Below, I address both in turn.

The problem that Read notes with modus ponens is that the major premise of this argument (16) is either false or redundant. While his discussion of this problem is limited, he links it with SPS by arguing that in both cases the additional premise "adds psychological perspicuity [...] But at the same time, it is not essential".[19] To some extent I disagree with both points. Considering the second point, the suppressed premise and the major premise in the modus ponens argument are vital in making the argument valid, and thus are essential to the argument. On the first point, there is some sense in which adding the suppressed premise and the major modus ponens premise do add psychological perspicuity, but it does not necessarily always do this or do this to the extent Read may be suggesting. In cases where both parties implicitly know the suppressed premise, its addition to the argument may not provide any psychological clarity, only logical infallibility. This idea is strengthened when considering that most of the arguments we make in everyday life have suppressed premises and we do not seem to need to reveal these suppressed premises for psychological reasons.[20] Rather we tend to reveal suppressed premises for logical reasons. Given we are holding this discussion in the domain of logic, we may accept the resemblance between SPS and modus ponens while still rejecting Read's assertion of redundancy.

Smiley's discussion of this matter refers to a paradox that seems to be presented by modus ponens and the addition of the suppressed premises. The paradox in question originated from Lewis Carroll, who wrote:

> If I grant (A) All men are mortal, and (B) Socrates is a man, but not (C) The sequence "If all men are mortal, and if Socrates is a man, then Socrates is mortal" is valid, then I do not grant (Z) Socrates is mortal. Again, if I grant C, but not A and B, I still fail to grant Z. Hence, before granting Z, I must grant A and B and C. [Now consider] (D) If A and B and C be true, then Z is true.[21]

This becomes paradoxical when we observe an infinite regress occurring where we must grant (A), (B), (C), (D), and a further (E) If A and B and C and D be true, then Z is true, yet we can think of an infinite number of propositions that must be granted before it seems that Z is granted. We can view (C), (D), etc, as suppressed premises of the argument that Carroll reveals in his paradox. This leads Smiley to comment that "Lewis Carroll was doomed to detect suppressed hypothetical premises even in logically valid arguments, and incidentally to force them all into the straitjacket of modus ponens".[22] If these are considered to be suppressed premises then there is a problem for SPS, for these can be added to any argument, and make the

[18] Read, "Formal and Material Consequences," 259-62; Timothy Smiley, "A Tale of Two Tortoises", *Mind* 104, no. 496, (1995): 727.
[19] Read, "Formal and Material Consequences," 262.
[20] Smiley, "A Tale of Two Tortoises," 727.
[21] Charles Lutwidge Dodgson, *Lewis Carroll's Symbolic Logic*, W. W. Bartley III, ed., (Clarkson Potter, 1977), 472.
[22] Smiley, "A Tale of Two Tortoises," 727.

argument paradoxical. In addition, this does not seem to be what the strategy intends. To solve this, we can examine the characteristics of the suppressed premise, which is that its addition will make the argument formally valid. Yet the arguments that Lewis Carroll imagines are already valid arguments, thus SPS should not be employed in these cases. Smiley's examination of the problem also points out that the specific wording of the paradox is crucial for its paradoxical nature but is itself flawed. Lewis Carroll "lacked any distinct conception of a deduction as opposed to the assertion", and it is this confusion that leads to paradox. [23] By this Smiley means that (C) is not an assertion but a deduction, and so it must be granted, but Carroll seems to think that it is merely an assertion that can be accepted or denied. Hence, this paradox does not show that even valid arguments might have suppressed premises that lead to paradox, thus this objection presents no issue to the use of SPS.

The characterisation I have given of SPS prevents contradiction and I have shown how it is able to overcome objections about the redundancy of the suppressed premise, as well as more generally the nature of the suppressed premise and the nature of the arguments to which it pertains. Finally, I discussed the problem of Modus Ponens, showing two ways it relates to SPS, and that this does not impact the use of the strategy. Thus, SPS is a viable addition to the interpretational account and explains the relation of material validity to validity, without a need to adopt the representational account. Hence by defence of the interpretational account succeeds and preferred to the representational account.

[23] Smiley, "A Tale of Two Tortoises," 727.

Bibliography

Adams, Robert Merrihew. "Theories of Actuality." *Noûs* 8, no. 3 (1974), 211-231.

Beall, Jc, Greg Restall, and Gil Sagi "Logical Consequence", *The Stanford Encyclopedia of Philosophy* (Summer 2024 Edition), Edward N. Zalta & Uri Nodelman (eds.), https://plato.stanford.edu/archives/sum2024/entries/logical-consequence

Dodgson, Charles Lutwidge *Lewis Carroll's Symbolic Logic*. W. W. Bartley III, ed. Clarkson Potter, 1997.

Griffiths, Owen, and Alexander Paseau. *On True Logic: A Monist Manifesto.* Oxford University Press, 2022.

Lewis, David. *On the Plurality of Worlds.* Basil Blackwell, 1986.

Read, Stephen. "Formal and Material Consequences." *Journal of Philosophical Logic* 23, no. 3 (1994): 247-265.

Sainsbury, Mark. *Logical Forms: An Introduction to Philosophical Logic.* Blackwell, 2001.

Smiley, Timothy. "A Tale of Two Tortoises." *Mind* 104, no. 416 (1995): 725-736.

Stalnaker, Robert C. "Possible Worlds." *Noûs* 10, no. 1, (1976): 65-75.

Tarski, Alfred. "What are Logical Notions?" *History and Philosophy of Logic* 7, 1986): 143-154.

Contributors

Audrey Hammer

Audrey is finishing a BA in Philosophy at the University of Cambridge. She is taking papers in Mathematical Logic, Philosophical Logic, Wittgenstein and His Successors, and Kant. Her dissertation is on the implications of independence and incompleteness in results on realism set in theory. Next year she aims to pursue an MPhil in Philosophy at the University of Cambridge, where she wants to explore the nature of logic and mathematics.

Koda (Wenjing) Li

Koda is from China and is a third-year Linguistics and Philosophy student at Brown University. In philosophy, he is mainly interested in philosophy of mind, philosophy of language, metaethics, and philosophical logic. In linguistics, he is mainly interested in semantics and its interface with semantics, particularly with pronoun/binding and intensionality.

Wiktor Przybrorwski

Wiktor is a third-year MA Economics and Philosophy student at the University of St Andrews. His primary interests lie in the philosophy of logic and language, particularly in topics such as naming and reference, paradoxes, and non-classical logics.

Wilson Sugeng

Wilson is an Economics and Philosophy undergraduate at the University of St Andrews. He previously read Theology at Wheaton College, Illinois, and spent a semester at Wycliffe Hall, Oxford. With work experience in finance, microeconomic research, and church ministry, his research interests include religious epistemology and the ways different academic fields, religions, and philosophical traditions shape ethical and metaphysical frameworks. Outside of academics, Wilson enjoys cooking and hiking with the university's hillwalking club.

Aporia

Undergraduate Journal of the St Andrews Philosophy Society

VOLUME XXV

Aporia is funded by the University of St Andrews Philosophy Society, which receives funds from the University of St Andrews Department of Philosophy, the University of St Andrews Students' Association, and independent benefactors.
Aporia is published by The University of St Andrews Philosophy Society.

Aporia © 2025 is licensed under Creative Commons Attribution 4.0 International (CC BY 4.0). To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/.

Authors retain copyright, but give their consent to *Aporia* to publish their work.

aporia@st-andrews.ac.uk

Visit https://ojs.st-andrews.ac.uk/aporia to learn more.

Aporia
School of Philosophy
Edgecliffe, The Scores
St Andrews, Fife
KY16 9AL
Scotland