Apophansis

Volume 1 No. 1 June 2024

The Postgraduate Journal for Academic Philosophy from King's College London

Apophansis

The Postgraduate Journal for Academic Philosophy from King's College London

Contents

History of Philosophy

Aquinas on Angelic Difference
– Brandon Smith 2

Pleonexia and Political Decline in Plato's Republic
– Arthur Kleinman 12

Moral and Political Philosophy

On the Slavery Objection to Political Authority
– Jonathan Anderson 24

Simone Weil and the Ethics of Attention: An Investigation on Compassion, Love and Justice
– Anna Maria Robles Fumarola 36

Philosophy of Mind and Epistemology

Justification as an Evaluative Concept: A New Argument for Externalism
– Guy Dinwiddy Smith 44

On the Logical Character of Emotive Language:Evaluating Williamson's Anti-luminosity Argument
– Mette Andersen 54

The Blind Roboticist: The Knowledge Argument Against Qualia
– Dan Ayres 68

Ontology

Exploring the Boundaries of Word Identity through a Nominalist Ontology of Words
– Sophie Keay Rodger 80

Letter from the Editors

We are very proud to present the inaugural edition of *Apophansis,* a postgraduate journal for academic philosophy based in King's College London. First and foremost, we extend our heartfelt thanks to Andrea Sangiovanni for his dedication to securing funds for aspiring philosophers. Additionally, our thanks go to our faculty advisor, Zita Toth, for her support and guidance. Finally, we wish to express our appreciation to the students who entrusted us with their work– it's a privilege to provide a platform for them.

Over the past six months, the *Apophansis* Editorial Board has come together to read and discuss over 50 quality submissions spanning three continents. Among the many strong candidates, the final articles have been chosen for their innovative approaches, precise argumentation, or insightful philosophical analysis, each of which will –we hope– offer interesting perspectives for readers to engage with.

Our first section tackles problems in the **History of Philosophy**, featuring work from Brandon Smith and Arthur Kleinman. Our second section consists of essays discussing issues within **Moral and Political Philosophy**, with contributions from Jonathan Anderson and Anna Maria Robles Fumarola. The third section engages in debates in **Philosophy of Mind and Epistemology** showcasing work from Guy Dinwiddy Smith, Dan Ayres, and Mette Andersen. Our fourth and final section presents a stand-alone work of **Ontology** by Sophie Keay Rodger. This is our preferred ordering. Nevertheless, for those yearning for more philosophical depth, we recommend buying the online version, ordering the sentences alphabetically, and considering *Apophansis* with a glass of wine.

As we reflect on the journey of founding *Apophansis,* we're prompted to consider the status of philosophy today. We find ourselves in an age when the societal value of philosophical pursuit is increasingly questioned. There is no shortage of burgeoning young philosophers vying for inclusion in the academic job market and yet almost everywhere universities are scaling back philosophy departments. Despite these challenges, we, as young philosophers, remain resolute in our pursuit. In answer to those who question our dedication to philosophy in this unwelcome era, all we will allow ourselves to do is simply echo Apollodorus

> *Perhaps, in your turn, you think I'm a failure, and, believe me, I think that what you think is true. But as for all of you, I don't just think you are failures—I know it for a fact* (Plato, Symposium, 173d).

Without further ado, we invite you to enjoy the inaugural edition of *Apophansis*. We hope it sparks contemplation.

Sincerely,
Isabella Lucero and Nathanael Dorsch
Editors-in-Chief

Aquinas on Angelic Difference

Brandon Smith

King's College London

Consider the following two differences: (1) the difference between a dog and a ferret and (2) the difference between my dog and some other dog. In the first case, there is a difference in species. In the second, the two different things have the same species, but they are numerically distinct. Aquinas maintains that the difference between two angels amounts to a difference in species. A standard approach to Aquinas's position explains that, because angels are immaterial, and because things are numerically distinct in virtue of their matter, angels are not numerically distinct. This view is almost certainly correct. But, in Article VIII of his *De Spiritualibus Creaturis*, Aquinas argues differently. There, he concludes that *even if* angels were material, they would differ in genus, and therefore species. In §1, I explain the terms genus, species, and essence. In §2, I use the discussion of essence to introduce Aquinas's notions of signate matter and numerical distinction, and then I discuss the standard approach to Aquinas's argument for angelic specific difference. In §3, I present and explain Aquinas's peculiar argument in Article VIII of *De Spiritualibus Creaturis*. The argument's peculiarity renders it worthy of notice; its relation to multifarious other facets of Aquinas's thought makes it worthy of investigation.

1. Species and Essence

So far, when discussing differences in number and species, I have appealed to the ordinary understanding of species of animals—like dogs and ferrets. But I should say more precisely what a species is.[1] In the *Categories*, Aristotle distinguishes between primary substances, each of which is a particular being - a 'this something' (τόδε τι) - and secondary substances,which are species and genera, each of which is a universal.[2] Species are predicated of primary substances and genera are predicated of species.[3] Aquinas, following Aristotle,

[1] I assume familiarity with the notions of substance, form, and matter, as they occur in Aristotle, and substantial form and prime matter as they occur in Aquinas. Nevertheless, I will *briefly* explain them here. **Substance** (οὐσία) is (primarily) some being, i.e., a 'this something' (τόδε τι) (*Categories* 2^a10-17, *Metaphysics* B5 1001^b33 and Z4 1030^a5); **form** (μορφή or εἶδος) is the account (λόγος) of what the being is (*Metaphysics* Δ2 101325), is what the being changes to (*Metaphysics* Λ3 1070a1), and is its actuality (*Metaphysics* Λ8 1074^a31-6, *On the Soul* II 412^a9-10 and 412^a20-3); **matter** (ὕλη) is the potential of a being to change and be what it can be (*Metaphysics* H2 1042^b7-10, On the Soul II 412^a9-10), the potential to be informed (*Metaphysics* Θ8 1050^a15), and is not said to be anything determinate(*Metaphysics* Z3 1029^a15-26); **substantial form** is that form which causes the actual existence of a substance (*DPN* 1.5), and that form by which something comes to be simpliciter (*DPN* 1.6-7); and **prime matter** is matter insofar as it lacks any form or privation (*DPN* 2.14) and is therefore purely potential (*DPN* 2.17).

[2] Aristotle *Categories*, 2^a11-19.

[3] *Ibid.*

says that a species is 'contained under a genus and constituted of differences.'[4] For example: man is a species; it is contained under the genus, animal; and it is constituted by the difference, rational. Man is said of some particular man, say, Porphyry. Animal is said of man, and animal is determined as man by what differentiates man from animal as a whole, namely, rational.

But not any predication of an individual will be a species, and not any predication of a species will be a genus. As Aristotle explains, species and genera do not merely say something accidental about a substance. Rather, they signify 'what sort' of substance (ποιὰν τινα οὐσίαν) something is, insofar as it is substance.[5] In other words, species and genus signify something *essential*. Now, essence, roughly, is what is signified in the definition of something, and it is that through which a thing is what it is.[6] A comprehensive account of essence according to Aquinas is beyond the scope of this paper. For our purposes, which include introducing the notion of species in a way relevant to specific and numeric difference, we need only note that Aquinas holds that the essence is 'something common to all natures thanks to which beings are placed in the different genera and species'.[7] As Shields and Pasnau read this remark, to say that y is the essence of y is therefore to say that '$x$makes y be a member of species $z$and y is held in common by all members of z.'[8] Thus, Shields and Pasnau conclude that members of the same species have the same essence.[9] Although they do not argue for this conclusion, it is easy to see why they maintain it, for an essence determines the existence of a substance to one species,[10] such that if two things have different essences, they must be determined to different species.[11]

2. Matter and the standard account of angelic difference

How are two members of a species differentiated, if they share the same essence? In other words, how can things be numerically distinct? According to Aquinas, they are individuated by signate (*signata*) matter. In this section, I will attempt to explain what signate matter is by considering Aquinas's account of it in *De Ente et Essentia*.

Aquinas argues that the essence of a composite substance - i.e., one composed of matter and form - is neither matter nor form alone. It is not matter alone because matter is unknowable and potential, while essence is what is known and actualizes.[12] It is not form alone because then 'natural and mathematical definitions would not differ.'[13] Aquinas does not explain this remark here. I take it that he is drawing from Aristotle, who holds that the natural scientist, who studies composite substances insofar as they are composite, defines

[4] Aquinas *Commentary on the Metaphysics,* Bk. VII, L. 3, C. 1327, vol. 2, p. 28.
[5] Aristotle *Categories*, 3^b18-23.
[6] Aquinas *On Being and Essence*, 1.7-11, p. 31-2.
[7] *DEE* 1.6, p. 31.
[8] Shields and Pasnau 2016, p. 66.
[9] *Ibid.*, p. 67.
[10] ST I.Q54.A2.co.
[11] Shields and Pasnau do not include genus in their formulation of essence because the genus *indeterminately* signifies essence, such that when a difference is added, it is determined to signify some essence (and constitutes a species). Many differences will determine many essences, or many 'species essentially different,' within one genus (*DEE* 2.41).
[12] *DEE* 2.15, p. 32.
[13] *DEE* 2.16, p. 32.

the essences of composites insofar as they are not without matter (like the snub), while the mathematician considers composites insofar as they are form alone (like the concave) and accordingly defines only form.[14] So if the essences of composite substances were merely formal, then the natural scientist and the mathematician would define things in the same way; but they don't because the natural scientist takes stock of the materiality of what they study. Aquinas concludes that the essence of a composite substance signifies what is composed of matter and form.[15]

If the essence of a ferret includes its matter and form, and if all members of a species share the same essence, it seems like all ferrets should share the same matter. How, then, can matter individuate members of the ferret species? According to Aquinas, matter is not unqualifiedly the principle of individuation, but only "designated" or signate (*signata*) matter is.[16] Now, signate matter is matter under determinate dimensions (that is, the three dimensions).[17] Signate matter is not universal and is therefore not included in the account of a species. Thus, it is not signate matter that is signified by the essence of a composite substance, but non-signate matter. Thomas writes,

> 'Non-designated matter is included in the definition of man, for it is not this bone and this flesh that are put into the definition of man, but bone and flesh as such which are the non-designated matter of man.'[18]

Bones and flesh considered as they are in the essence of the ferret are not extended in three dimensions, but *these* bones and *this* flesh of this particular ferret are. Moreover, the bones of *this* ferret are not the bones of *that* ferret, and thus ferrets are individuated within their species insofar as their matter is signate.

Because only material beings have signate matter, it follows that only material beings are individuated within a species. The standard account of angelic difference according to Aquinas readily follows once it is supposed that angels are immaterial. Arguments that express the standard account have the following structure:

(A) Angels have species.

(B) Things can be many and have the same species if and only if the members of the species have matter.

(C) Angels do not have matter.

(D) By (B) and (C), there cannot be many angels of the same species.

(E) By (A) and (D), if there are many angels, they must have different species.

(F) There are many angels.[19]

∴ By (E) and (F), angels have different species.

[14] Aristotle *Metaphysics*, $1025^{b}29$-$1026^{a}10$.
[15] *DEE* 2.18, p. 33.
[16] *DEE* 2.23, p. 33.
[17] *Ibid.*
[18] *Ibid.*
[19] *ST* I.Q50.A3.co.

I have called this argument standard because many scholars attribute a version of it to Aquinas. Pasnau and Shields explain (A) and (C) by noting that Aquinas's angels have simple essences and are therefore members of a species and devoid of matter and conclude that angels lack the ground for numeric distinction expressed in (B).[20] J. Collins explains (A) by arguing that, since angels have forms, they have species, and concludes (D) from (B) and (C).[21] J. F. Wippel explains (C) by noting that Aquinas maintains that angels are not matter-form compounds, and asserts that, given (B), (D) and the Conclusion follow.[22] A. Kenny concludes (D) from (B), presupposing (A) and (C).[23] All of these philosophers presuppose (F) because Aquinas unquestionably maintains it.

In calling the approach "standard", I do not mean to imply that it is wrong or insufficient. Aquinas *does* argue in the standard way.[24] In the next section, I will consider an argument that differs from the standard approach in one significant way: therein, Aquinas does not aver (C).

3. The argument in De Spiritualibus Creaturis VIII

This section has two parts. In the first, I reconstruct Aquinas's argument, the entirety of which is found in *QDSC A8.co*. In the second, I consider the argument.

3.1 The Argument Reconstructed

(1) Angels are either simple forms or composed of matter and form.

(2) If angels are simple forms and do not differ in species, then their simple forms will be multiplied in that species.

(3) No simple form is multiplied in a species.

(4) By (2) and (3), if angels are simple forms, they differ in species.

(5) If angels are composites of matter and form and differ from each other, their matters differ.

(6) Matters differ either according to their particular dimensions or with respect to their potencies in relation to acts.

(7) Only the matters of corporeal beings can differ according to their particular dimensions.

(8) Angels are not corporeal beings.

(9) By (7) and (8), the matters of angels do not differ according to their particular dimensions.

(10) By (5), (6), and (9), if angels are composites of matter and form, their matters differ according to their potencies to acts

[20] Shields and Pasnau 2016, p. 77-8.
[21] Collins 1947, p. 96.
[22] Wippel 2000, p. 311.
[23] Kenny 2002, p. 166.
[24] See *DEE* 4.75, p. 41-2 or *ST* I.Q50.A4.co.

(**11**) If angels' matters differ according to their potencies to acts, then angels are generically distinct.

(**12**) Whatever is generically distinct is specifically distinct.

(**13**) By (11) and (12), if angels' matters differ according to their potencies to acts, then angels differ in species.

(**14**) By (10) and (13), if angels are composites of matter and form, then angels differ in species.

∴ By (1), (4), and (14), angels differ in species.

3.2 The Argument considered

Premise (1) is clear. Angels are substances, and they are either simple or composite. If they are composite, they are composed of matter and form. If they are simple, then they are forms alone, for, according to Aquinas, 'there is nothing to prevent a given form from subsisting (*subsistere*) without matter, although matter cannot exist (*esse*) without form.' [25] Premises (2) to (4) are consistent with the standard account of angelic difference. Premise (5) says that matter is the principle of individuation in matter-form composites, which we have discussed in §2. The argument from (7) to (9) is familiar. To say that a being is corporeal is to say it has a *body*. A body is a substance whose form permits its matter to be designated in three dimensions, i.e., a body's matter is signate.[26] Angels are incorporeal, and thus their forms do not permit their matters to be so designated. So, while the matters of corporeal substances within one species are individuated insofar as they are signate, angelic matters could not be.

Aquinas does not wrap things up by way of (9). Rather, by way of the distinction in (6), he argues that if the matters of angels differ, they must differ according to their potencies to acts. By way of (11), he argues that they would then differ in genus, and therefore, by (12), in species. Aquinas does not argue for (12), but it follows from the fact that a genus is said of all the members of its species. If genus G is said of species S, then whatever S is said of, G is said of.[27] Therefore, if S is said of substance y, then G is said of y. Suppose substance y is generically distinct from substance y such that G is not said of y. Thus, if S is said of y, then G is said of y. But G is not said of y. Therefore, S is not said of y, i.e., y and y differ in species. (13), (14), and the Conclusion follow from what has been said.

Aquinas's explanation of premises (6) and (11) is brief. He asserts that (6) is an exhaustive distinction. He gives one example of a distinction according to potencies to different acts: while lower bodies are in potency to 'actual being' (*esse*),[28] heavenly bodies are in potency to "place". He supports the assertion that things distinct in this way are generically distinct merely by claiming that 'different things are generically different on a basis of

[25] *QDSC* A1.ad6, *DEE* 6.104, p. 47.
[26] *DEE* 4.28, p. 34.
[27] Aristotle *Categories*, 1^b9-15.
[28] My translation has 'the matter of lower bodies, which is a potency to actual being' for *modo materia inferiorum corporum, quae est potentia ad esse*.

matter.'[29] He supports this claim by citing (all of) Book Δ of Aristotle's *Metaphysics.* In what follows, I will attempt to explain what is going on in these two crucial premises. I will treat (11) first.

Aquinas uses lower bodies and heavenly bodies to explain generic difference in his Commentary on Book Δ of Aristotle's *Metaphysics.* There, he writes, 'the celestial bodies and lower bodies are diverse in genus inasmuch as they do not have a common matter.'[30] In the commented-on text, Aristotle says that 'things are said to be distinct in genus whose primary underlying subjects are distinct.' [31] Aristotle says in the same chapter that such an underlying subject is the matter which is related to the differentia.[32] When Aquinas cites the entirety of Book Δ of Aristotle's *Metaphysics* to support his claim that 'different things are generically different on a basis of matter,'[33] *it* is this passage in Δ28 that is most relevant.

What does it mean to say that lower bodies are in potency to being *(esse)* while heavenly bodies are in potency to place, and why should this distinction preclude them from having a common matter or underlying subject? Aquinas argues in his Commentary on the Physics that motion with respect to place is prior to generation and corruption inasmuch as the former can be without the latter but the latter requires the former.[34] In his Commentary on *De Caelo,* Aquinas says that for the lower bodies to be in potency to *esse* is for them to be in potency to generation and corruption, and he argues that the heavenly bodies are not in potency to generation and corruption because they are in potency to place alone.[35]

Because the heavenly bodies are not in potency to generation[36] or corruption, they are *incorruptible.* Because the lower bodies are in potency to generation and corruption, they are *corruptible.* According to Aristotle, the corruptible and the incorruptible are contraries that differ in genus.[37] In his Commentary on this point, Aquinas distinguishes between two ways in which contraries can relate to a genus: in the first, the genus *can* be both contraries (as animal can be rational or irrational) because the same matter is in potency to both, while in the second, the genus *cannot* be both contraries because the same matter cannot be in potency to both.[38] For, according to Aquinas, the genus signifies the whole of matter in potential to its different determinations[39]—the matter signified by a genus must be in potency to *all* of the determinations of the genus.

Now, to say that something is corruptible is just to say that its matter is in potency to

[29] *QDSC* A8.co.
[30] *InMeta* V.22.1125 (1:525).
[31] Aristotle, *Metaphysics* 1024^{b}9-10.
[32] Aristotle *Metaphysics*, 1024^{b}6. The Greek is awkward to translate: 'οὗ γὰρ ἡ διαφορὰ καὶ ἡ ποιότης ἐστί, τοῦτ' ἔστι τὸ ὑποκειμενον, ὃ λέγομεν ὕλην, 'for that to which the differentia and the quality is, this is the underlying subject, which we call the matter.'
[33] *QDSC* A8.co.
[34] *InPh* VIII.114.1093.
[35] *InDC* 1.6.62-3.
[36] Aquinas notes that, 'according to the Catholic faith,' it is not said that the heavenly bodies have always been, but neither is it said, 'that they began to be through generation, but through an efflux from the first principle' (*InDC* I.6.64).
[37] Aristotle *Metaphysics* 1058^{b}26-8.
[38] *InMeta* X.12.2142 (2:358-9).
[39] *DEE* 2.23-2, p. 35.

taking on another substantial form. All lower (elemental) bodies are corruptible. Aquinas argues that the same matter cannot be determined to be corruptible and incorruptible, even at different times, for the incorruptible can never change into the corruptible, else it wouldn't have been incorruptible.[40] So, for example, if the matter of an incorruptible heavenly body were in potency to acquiring the form of a corruptible lower body, then it would be in potency to acquiring another substantial form—but this is just what it means to be corruptible.[41] So the matter of the heavenly body must not be in potency to acquiring the forms of any of the lower bodies. The genera of the corruptible and the incorruptible must be distinct, for no one matter can be in potency to both.[42] Premise (11) thus means: if the matters of angels are distinct insofar as they are in potency to acts which no one matter can stand in potency to, then they differ in genus.

Before concluding, I will construct an argument for (6), since Aquinas only asserts it. Matters are either in potency to different acts, or they are not. If they are, then they differ accordingly. If they are not, then, *qua* matter, they are one. One matter is either 'without the dispositions (*dispositionibus*) which would cause it to differ numerically,'[43]or it is with such dispositions. If it is without such dispositions, then it can only be said to be one.[44] If it is not said to be one, it is thus with such dispositions, and therefore differs numerically from other such matters. But things which differ numerically differ according to their dimensions. Therefore, if matters are in potency to the same act and are not said to be one, then they differ according to their dimensions, and the exhaustiveness of (6) is secured.

In sum, in other texts Aquinas argues that angels must differ in species because they are immaterial and therefore cannot be individuated within a species by matter. It is standard to explain Aquinas's account of angelic specific difference in this way. But, in the unusual argument of *De Spiritualibus Creaturs* VIII, Aquinas argues that, so long as angels are incorporeal and without the three dimensions, they will differ in species even if they are material. For, the only other sort of material distinction they could have would be the sort by which heavenly bodies and lower bodies differ, and their material distinction entails a generic distinction.

[40] *InMeta* X.12.2140 (2:358).
[41] I agree with the interpretation of this argument in S. Balder (2004, p. 457).
[42] At *InMeta* V.22.1127 and X.12.2142 Aquinas distinguishes between logical and natural genera. While the former correspond to the categories, the latter correspond to the order of potencies and acts. The corruptible and the incorruptible differ in natural genus. See also *ST* I.Q66.A2.ad2.
[43] *DPN* 2.16.
[44] *Ibid.*

Bibliography

Aquinas.(1963). *Commentaria in octo libros Physicorum.* Kenny, J. (ed.) & Conway, P. H. (trans.). College of St. Mary of the Springs. Available from: https://isidore.co/aquinas/Physics.html.

———. (2020). *Commentary on Metaphysics Books 1-6.* Rowan, J. P. (trans.). Green Bay: Aquinas Institute, Inc.

———. (2020). *Commentary on Metaphysics Books 7-12.* Rowan, J. P. (trans.). Green Bay: Aquinas Institute, Inc.

———. (n.d.). *De Principiis Naturae.* J. Kenny, J. (eds.) & Kocourek, R. (trans.). Available from: https://isidore.co/aquinas/DePrincNaturae.html.

———. (1998). "On Being and Essence". In: McInerny, R. (ed. & trans.). *Thomas Aquinas Selected Writings.* London: Penguin Books, pp. 30-49.

———. (1953). *Quaestio disputata de spiritualibus creaturis.* Alarcon E. (ed.). Available from: https://www.corpusthomisticum.org/qds.html

———. (1949). *St. Thomas Aquinas On Spiritual Creatures (De Spiritualibus Creaturis).* Translated by Fitz Patrick M. & J. Wellmuth, J.(eds.). Milwaukee: Marquette University Press.

———. (n.d.). *Summa Theologiae.* Knight, K., (ed.) & the Fathers of the English Dominican Province (trans.). Available from: https://www.newadvent.org/summa/.

———. (n.d.). *Thomas Aquinas on The Heavens.* Larcher, F. & Conway, P. (trans.). Available from: https://isidore.co/aquinas/english/DeCoelo.html.

Aristotle. (1938). "Categories". In: Cooke, H., & Tredennick, H. (ed. & trans.). *Categories, On Interpretation, Prior Analytics.* Cambridge: Harvard University Press Loeb Classical Library, pp. 12-111.

———. (2016). *Metaphysics.* Reeve, C. (trans.). Indianapolis: Hackett.

———. (1957). "On the Soul". In: Hett W. (ed. & trans.). *On the Soul, Parva Naturalia, On Breath.* Cambridge: Harvard University Press Loeb Classical Library, pp. 8-206.

Baldner, S. (2004). "Thomas Aquinas on Celestial Matter". *The Thomist: A Speculative Quarterly Review*, 68(3), pp. 431-467.

Collins, J. (1947). *The Thomistic Philosophy of the Angels.* The Catholic University of America Press.

Kenny, A. (2002). *Aquinas on Being.* New York: Oxford University Press.

Shields, C., & Pasnau, R. (2016). *The Philosophy of Aquinas.* New York: Oxford University Press.

Wippel, J. F. (2000). *The Metaphysical Thought of Thomas Aquinas.* The Catholic University of America Press.

Pleonexia and Political Decline in Plato's Republic

Arthur Kleinman

King's College London

What are the institutions, practices and mores that make societies peaceful, prosperous and resilient while preventing the depredations and social ills that have plagued humanity for much of its recorded history? This question – one that can be termed *the problem of political order* – has vexed philosophers, political theorists and social scientists since antiquity. However, it is especially salient at present given the myriad challenges facing modern societies, including political polarisation, mounting economic inequality and societal fragmentation. It is against this backdrop that Plato's account of the 'degenerate regimes', articulated in Books VIII and IX of his masterwork, the *Republic*, merits close analysis and consideration. In this paper, I put forward that the concept of *pleonexia* – the desire to want or have more – plays a crucial role in the construction of the degenerate regimes, and that an awareness of this is required to grasp the passage's full import for the problem of political order.

My argument will proceed as follows: In §1, I introduce Plato's tripartite psychology and reconstruct the argument of Era Gavrielides that degenerate regimes aren't inherently unstable, but are only problematic for Plato because they achieve stability by "force" instead of "persuasion". In §2, I argue against Gavrielides that the degenerate regimes suffer from *diachronic* instability because they lack the institutions – philosophic rule and rational education – that insulate polities from the damaging effects of pleonexia. In §3, I argue that it is precisely this diachronic instability that renders the degenerate regimes so problematic in Plato's framework, insofar as it distances them from what he identifies as a cardinal ontological virtue: insusceptibility to change.

1. Quasi-Rational Order in the Degenerate

1.1 Types of Rational Rule

In the *Republic,* Plato (through the voice of Socrates) outlines a tripartite model of human psychology, one reflected in the structure of human societies: a rational part that reasons and calculates, a spirited part that undergoes passions of anger, pride and shame, and an appetitive part that primarily generates bodily desires but is also concerned with wealth accumulation. The text's core proposition is that justice, both at the individual and societal level, consists in each of these components doing the work they are suited for or executing their *proper function* (Republic 433c). This means the rational part 'ruling', identifying what is good or advantageous for the whole and calculating the best course of action to that end (441e); the spirited part allying itself to the rational part and ensuring the psyche's obedience to reason's dictates (442c); and the appetitive part being firmly subordinated to the other

two parts and exerting itself non-disruptively (442c-d). The 'Kallipolis' is the society whose social structure mirrors this arrangement and thus instantiates full justice (441d).

The central preoccupation of Platonic justice is securing the rule of reason in the fullest sense. For John Cooper, this means one has 'worked out on exclusively rational and theoretical grounds what goals are worth pursuing' and 'patterned his life around the pursuit of these goals'.[1] In Books V-VII of the *Republic*, Plato identifies this 'theoretical ground' as **The Form of the Good (FG)**, which broadly speaking is the intelligible entity that underlies and orders all other metaphysical and normative truths (505a-b, 508e-509b). If the rational part is to rule, it must act on ends generated exclusively by reference to **FG** and organise the spirited and appetitive parts' conduct around these ends. For instance, the rationally-ruled person 'desires to preserve his health', not because he fears a painful bodily demise, but '*because* it is a good thing', the imperative for which derives its force and content from **FG** alone [2] **FG** affords moral agents an intellectual grasp of what each part of their soul is 'suited for' and habituates these parts into spontaneously gravitating toward their proper function such that 'faction' or motivational conflict is minimised (586e, 589b). In other words, the type of rational or *philosophic* rule (as I will henceforth label it, cf. 485a-b) being described doesn't unify the soul around reason's dictates by *force*, but by *persuasion*.[3]

This picture is complicated by G.R.F. Ferrari's distinction between the *activity* and *object* of soul-parts. The activity of the rational part is to calculate and engage in reasoning, while its proper object is **FG** and what follows from it, viz. the 'good' or 'best course of action'[4]. Philosophic rule involves the rational faculty engaging its characteristic activities, calculation and reasoning, for the sake of its proper object, **FG**. But what if the rational part undertook its activity in service of an object supplied by the appetitive or spirited part – a scenario that Ferrari's distinction renders feasible? In this instance, reason would still in a limited sense control one's actions, but not for the sake of endogenously generated rational ends. Rather, the result would be partially rational or *quasi-rational* rule, in which rationality would constitute 'a formal criterion from which no substantive choices and preferences can be derived'.[5]

There is much textual evidence that Plato identified and engaged with the notion of quasi-rational psychic order. In *Republic* VIII Socrates describes the oligarchic individual as one who establishes 'the appetitive and moneymaking element on the throne' and 'makes the rational and spirited elements' serve as its 'slaves', not letting the former 'calculate or consider anything except how little money can be made into more' (553c-d). The successful oligarch will quash the 'dronish' appetites – for instance, toward laziness or sexual over-indulgence – that might undermine their ability to make money (554b). For instance, even after they have enough income to secure an acceptable standard of living, they will resist the temptation toward idleness and work extra hours so they can accumulate more wealth. In this case, we don't merely see one appetitive desire successfully overpower another one, at least if the account here is to be consistent with **The Principle of Opposites** presented in *Republic*

[1] Cooper 2021, p. 132.
[2] Cooper 2021, p. 124; cf. Phaedo 68e.
[3] Gavrielides 2010, p. 210-211.
[4] Ferrari 2007, p. 165.
[5] Cooper 2021, p. 140.

IV. The latter dictates that 'the same thing cannot do or undergo opposite things... in the same respect, in relation to the same thing, at the same time'(436b).[6] When the moneymaking imperative that constitutes the oligarch's supreme practical principle intercedes against a frivolous, 'dronish' appetite, these competing motivations cannot simultaneously arise in the appetitive part. But the problem is avoided if the moneymaking imperative arises from the *rational* part of the soul, albeit one that has been 'corrupted' to calculate for the sake of an appetitive end. Motivational conflict here is thus between a hollowed-out rational part – one that pursues an exogenous, appetitively-sourced end instead of **FG** – and an appetitive part that urges unmediated impulsive action.

1.2 Gavrielides on the Degenerate Regimes

Although the 'necessary appetites' – the class 'we cannot deny' and 'whose satisfaction benefits us', like making money (558c-d) – supply the *object* of a successful oligarch's practical deliberation, nevertheless their rational part *formally* rules insofar as it calculates how best to satisfy these appetites. To this extent, the quasi-rationally ruled soul exhibits a kind of *order* whereby the rational part makes action conform to a coherent, long-term plan and motivates self-discipline to that end.

Against this backdrop, Era Gavrielides offers a novel interpretation of the degenerate regimes passage in Republic VIII-IX: that **the principal problem with these regimes *isn't* their instability and propensity to degenerate**. Rather, given the potential for quasi-rational rule under such regimes – timocracies, oligarchies, democracies, and tyrannies – they can attain 'a form of unity' and organise their action in accordance with 'short and long-term goals'.[7] She points to how the decline of the degenerate souls[8] in Book VIII is intergenerational: while Plato depicts a degenerating familial lineage and describes the mechanisms via which this degeneration occurs, each generation is depicted as internally stable.[9] The timocrat or spirit-governed individual, for instance, is depicted as able to maintain a timocratic psyche throughout their lives without degenerating to the oligarchic personality; the transition only occurs in their son, who is driven to valorise money after their father is persecuted by sycophants (553a-c). She concludes on these grounds that while 'the possibility of decline' might be an important part of the unsatisfactory picture painted of the degenerate souls, it 'cannot be the sole or even the main thing problematic about degenerate souls' – and by analogy, the degenerate regimes. If it were, Plato would lack the resources to critique timocratic, oligarchic or democratic regimes that exhibit robust quasi-rational order, threatening the passage's coherence.[10]

What normatively differentiates the Kallipolis and degenerate regimes for Gavrielides is that the latter achieve psychic or societal unity by force, while the former does so by persuasion – and crucially, 'a situation held together by force is simply uncomfortable and unpleasant *even when there is no risk that it might come apart*'.[11] As discussed in §1.1,

[6] Lorenz 2006, p. 22.
[7] Gavrielides 2010, p. 208.
[8] Not to be confused with the regimes, although the accounts run in parallel, and the soul and regime depicted at each stage are isomorphs.
[9] *Ibid.*, p. 209.
[10] Gavrielides 2010, p. 209.
[11] *Ibid.*, p. 210, emphasis added.

philosophic rule 'is concerned to bring it about that all the parts of the soul agree': the spirited part is trained to ally itself with reason and the appetitive part is habituated into moderation such that philosophic rule becomes a state of harmony in which each component 'does its own work', as **FG** dictates.[12] Quasi-rational rule, by contrast, entails that the rational part lacks the comprehensive intellectual grasp of each soul-part's proper role and thus fails to institute measures that habituate them into harmonious relations. Rather, the desires of the other parts are quashed only upon conflict with the ruling *object*: pride and reputation for the timocrat, money for the oligarch, freedom and lawful appetite satisfaction for the democrat, and 'lawless' appetite satisfaction for the tyrant (548c, 553c, 561b-d, 572d-573c.[13] When these impulses don't come into conflict with the individual's overriding imperative, they have no rational basis not to indulge them: the oligarch happily indulges their appetites when their reputation isn't at stake and they aren't 'terrified of losing... [their] possessions' (554c-d). The oligarch's unproductive or gratuitous appetites are resultantly 'kept in check through the force of law alone', but this is a tense, disharmonious and thus undesirable state for the soul to be in relative to the harmony of philosophic rule, **even if it turns out to be sustainable**.[14] The upshot of Gravielides's view is that degenerate regimes aren't *essentially* unstable or prone to degenerate, and even if they were, it wouldn't be the most important part of the *Republic*'s critique of them regardless. It is to this account's flaws we now turn.

2. Pleonectic Disorder in the Degenerate Regimes

2.1: Introducing the Concept of Pleonexia

A central concept in Plato's moral psychology is **pleonexia** (πλεονεξία): the desire to 'do better' or 'have more and more without limit'.[15] In the *Republic* it first appears in Book I, where Thrasymachus characterises the 'perfectly unjust' individual as someone whose willingness to adopt unscrupulous, immoral tactics means they always 'do better' – in the sense of acquiring more power and indulging more appetites – than their just counterpart, who is depicted as foolishly beholden to moral convention(343d-344a). For Thrasymachus's ilk, a recurring type in Plato's dialogues, the best life is one of **pleonectic** appetite satisfaction: one that, in the words of Callicles in the *Gorgias*, sees one's appetites 'get as large as possible' and eschews any kind of discipline or self-control (Gorgias 491e-492c). The concept articulates an expansionary model of human desire: unless subjected to rational control, our desires will be 'insatiable', inducing societal conflict and personal vice (586a-b).

We see this most vividly in *Republic* II. Glaucon asserts that if an ordinarily just person was given 'the freedom to do whatever they like' – that is, insulated from the negative reputational consequences that usually accrue to unjust conduct – they would end up behaving identically to the archetypally 'unjust' person, insofar as they too, despite their moralistic pretensions, are at root motivated by pleonexia, 'the desire to do better' (359c). The 'Gyges Ring' thought experiment, which envisions the realisation of said 'freedom' by making its subject invisible, is adduced by Glaucon as evidence for this conclusion: 'no one, it seems,

[12] *Ibid.*, p. 211.
[13] cf. Reeve 2006, pp. 47-49.
[14] Gavrielides 2010, p. 212.
[15] Reeve 2004, p. 327; Reeve 2006, p. 47.

would be so *incorruptible* that he would *stay on* the path of justice' when they could steal, murder, or engage in sexual impropriety without repercussions (360b, emphasis added; cf. Gorgias 466c).[16] The upshot is this: if unregulated human appetite wasn't inherently expansionary, Gyges's Ring wouldn't be such a deeply problematic and tantalising prospect. After all, the pre-Gyges shepherd will have appetites – predominantly appetites of the 'necessary' type outlined earlier, for food, sex, and the like – which he will sufficiently, if not maximally, satisfy. An inability to seduce the King's wife presumably wouldn't render his existence catastrophically frustrating. However, it is *possessing the ring* and having *available* the opportunities to commit injustice that so drastically amplifies and multiplies his appetitive desires, such that even if he initially has the will to stay on the path of justice, his morals are progressively eroded by the raw appeal of wanton indulgence. As we will see, the essentially expansionary or **pleonectic** character of human appetite constitutes a central motif of the *Republic*'s political theory.

2.2: Pleonexia and Diachronic Instability

Plato has Socrates assert that constitutions only change when 'faction breaks out' in the ruling class; when the elite is 'of one mind... change is impossible' (545c-d; cf. Laws 683e). Of the Kallipolis – the completely just city in which philosophers rule – Socrates says 'it is difficult for a city constituted in this way to change'; and it is only because 'everything that comes-to-be must decay' that Socrates contrives a mechanism, the failure of 'sense-perception' and the consequent breakdown of eugenics, to induce the elite faction that instigates the Kallipolis's decline (546a-b). As such, there is no internal contradiction that propels the Kallipolis's degeneration aside from the laws that bind all natural entities.

When the breakdown of eugenics 'dilutes' the ruling elite, however, it increasingly comes to be comprised of non-philosophic individuals who not only less ably govern the city but, more crucially, neglect its *educational institutions* – specifically, the musical and physical training that moulds its rulers' psyches (546d). Socrates stresses the upkeep of these institutions because in his terms, 'rhythm and harmony permeate the innermost element of the soul' and '*affect it more powerfully than anything else*' (401d, emphasis added). If an individual is to have the moderate, non-pleonectic appetites or *temperance* requisite for a fully just, philosophically ruled soul, they must receive the appropriate musical and physical training from a young age (402a, 403c, 410a); and Socrates goes as far as saying that a polity must be overseen by such harmonically trained individuals if 'its constitution is to be preserved' (412a).

To the extent that educational neglect instigates hitherto absent political change, we see why instability is built into the degenerate regimes: the only way to meaningfully curb pleonectic desire in human beings is through rational education and the temperance it engenders, and only the philosophically-ruled society or individual possesses the requisite knowledge to preserve the educational system. The degenerate regimes, lacking these institutions, can only curb pleonectic appetite expansion and maintain societal unity through *force*. As Gravielides correctly observes, Plato does see quasi-rational rule with its forceful suppression of the appetites as providing a measure of stability – but in Jerry Green's terms, the most it can offer is *synchronic* stability, stability in which 'citizens live at peace with one another' for

[16] Fissell 2011 p. 220.

a temporary period. Crucially, the degenerate regimes cannot provide *diachronic* stability – political stability that endures in the long term – because their attempts to curb pleonectic appetites unrestrained by the moderating influence of harmonic training ultimately prove futile.[17]

This can be seen in Book VIII's account of timocracy, a regime where the spirited part rules and 'victories and honors' are deemed the ultimate good (548c). Timocratic elites are averse to the gratuitous appetite indulgence praised by Thrasymachus and Callicles: strong feelings of pride and shame force them to eschew such a lifestyle (549a). In situations of motivational conflict like that of Leontius in Book IV, where appetitive desire and spirit are at odds, we can assume that (unlike Leontius) the timocrat's spirited part generally prevails (439e-440a); and the most effective timocrats are those exhibiting quasi-rational order, meaning reason exercising 'formal' control with the desires of spirit – 'victories and honors' – as its object, such that reason calculates how to maximise its attainment of spirited goods and the timocrat consistently obeys its dictates.

However, because the timocrat is not guided by **FG**, and because they have been educated inadequately, they can only suppress the impetus toward wanton appetite satisfaction by *force*, marshalling sensations of shame and disgust when it arises (548c). Unlike the philosophically-ruled individual whose appetites are temperate, the timocrat's appetites pleonectically expand and their spirited and rational parts must strenuously exert themselves to restrain them. But the forcefulness of this unity is precisely what renders it so fragile and ultimately impels its degeneration: Socrates describes the conflict between the spirit's *aversion* to wanton appetite indulgence and the pleonectic impetus *toward* the same as causing the timocratic elite to privately hoard wealth in 'storehouses and private treasuries' (548a). 'Reason mixed with musical training', the attributes identified by Socrates as the only 'lifelong preserver[s]' of individual virtue, are absent (549b), so as the timocrat ages and is less able to marshal the 'force' of spirit against their pleonectic appetites, they increasingly share in 'the money-lover's nature' (*Ibid.*). Socrates subsequently identifies the aforementioned 'storehouse filled with gold' – a metaphor for repressed but *active* pleonectic dynamics – as being what 'destroys' timocracy (550d).

We see this to even starker effect in the treatment of oligarchy, where the pleonectic dynamics that timocracy failed to repress are now granted a legitimate avenue: the 'necessary' appetite for wealth accumulation. Socrates makes quite clear that it is this legitimation of pleonexia that inexorably destabilises and induces the degeneration of oligarchy, a process he characterises as resulting from 'the *insatiable desire* for the good set before it – the need to become *as rich as possible*' (555b, emphasis added).

This unfolds in two ways: firstly, the oligarchic elite's pleonectic appetite for money deters them from enacting laws that might prevent intemperate people 'spending and wasting their wealth', for doing so would stop them profiting by 'making loans on the property of such people' (555c). As a result, younger elites face immense social disincentives, in the form of abundant luxuries and cheap credit, to exercise the self-discipline that the oligarchic lifestyle – with its ruthless suppression of all 'unnecessary' appetites (554a) – requires, and end up becoming 'incapable of effort, too soft to endure pleasures or pains, and lazy' (556b). Secondly, because the oligarchs neglect everything 'except making money', they keep 'their heads

[17] Green 2021, p. 66.

down' while their predatory economic practices reduce much of the population to poverty, leading to widespread disenfranchisement and the revolutionary ferment (555d-556c). In both cases the self-destructive, self-contradictory consequences of pleonectic dynamics are manifest: they simultaneously render elites decadent and the masses resentful, a combination ripe for societal divisions and inimical to political stability (556e). These divisions in turn induce the transition to democracy[18] in the *Republic*'s schema (557a). There is nothing contingent about this degenerative process's underlying logic: by making the pursuit of wealth its overriding practical principle, oligarchy unleashes centrifugal pleonectic dynamics that inexorably destabilise it in the long run.

Contra Gavrielides, this reading is reinforced – not undermined – by the intergenerational metaphor that Socrates deploys to characterise the transition between degenerate souls: the 'stable' or quasi-rational degenerate parents, because of their philosophic illiteracy and consequent inability to provide harmonic education, are unable to forestall the psychic degeneration of their children. Furthermore, at each degenerative stage, the parents are decreasingly able to transfer the semblance of virtue they possess to their offspring, rendering the latter increasingly susceptible to the influence and temptations of pleonectic appetite. For instance, the son of the oligarch, the young democrat, is targeted by 'drones'[19] and offered 'multifarious pleasures of every degree of complexity and sort'. Because the democrat's soul is 'empty of the fine studies and practices and the true arguments that are the best watchmen and guardians [of virtue]', they succumb to these advances and justify their indulgent lifestyle through 'imposter arguments' (560c-d). Although each degenerate individual down the chain can successfully maintain their condition through force, the very quality that essentially characterises quasi-rational rule – the crude instrumentalization of reason for a non-rational object – critically impairs their capacity to pass on the semblance of virtue they possess (viz., by adequately education their offspring). Thus, *diachronic* instability is an intrinsic aspect of degenerate souls, and by extension degenerate societies.

3. Is Pleonectic Instability inherently

Gavrielides could concede that descriptively speaking the degenerate regimes are diachronically unstable, but maintain that this doesn't by itself render them *problematic*. On her account, the discomfiting tension that forceful, quasi-rational rule involves renders it 'unsatisfactory in itself', independent of arguments around its stability or lack thereof.[20] This prompts the question: is there some respect in which the instability of the degenerate regimes renders them 'problematic' for the Platonic framework over and above the unpleasantness associated with their unification by force?

The answer lies in Plato's metaphysics of Forms. In Book V, Forms are characterised as entities that are themselves purely 'one thing', such as 'the beautiful itself' (476a-b): the form of beauty is unconditionally beautiful, and we only describe objects as beautiful by virtue of their 'participation' in that Form (476d). Crucially, the Form of any particular

[18] For reasons of space, I won't treat democracy and tyranny in detail, but given that these are also appetitively-ruled constitutions, the argument *a fortiori* applies to them too. Socrates at 562b: '... isn't democracy's *insatiable desire* for what it defines as the good [freedom] also what destroys it?' (emphasis added).

[19] Individuals ruled by 'unnecessary' – harmful and gratuitous – appetites (559b-d).

[20] Gavrielides 2020, p. 210.

property F must remain 'always the same in all respects' (478e-479a), for this is just what it takes for it to be *unconditionally* F and provide a fixed reference point for what constitutes F-ness in other things (484c). Philosophers, in turn, are those with cognitive access to the Forms, by virtue of which in Reeve's terms they 'judge a property to be F only if it actually is F' and thus possess the power of *knowledge*.[21]

In contrast, those possessing the inferior epistemic power of *belief* fail to isolate the Forms and are only capable of grasping entities that merely participate in or provide a 'likeness' of F (476c). These 'likenesses', lacking the fixed, immutable character of the Forms, invariably suffer from predicative instability such that they cannot reliably be said to 'be F' or 'not-F'. Returning to the example of beauty, Socrates and Glaucon agree that it is inevitable for objects in this class to 'somehow seem both beautiful and ugly' and veer between these states depending on one's perspective (479a-b). Plato (controversially) infers from this predicative or epistemic instability – that *we* cannot reliably predicate them – a more fundamental *ontological* instability, meaning that they have no fixed nature and are caught in a cycle of 'coming-to-be and decaying' (485b). In other dialogues Plato suggests this mutability or alterability is a defect of corporeal reality as such – one that Forms, by virtue of their incorporeality, are immune to (e.g. *Timaeus* 49c-50c.)[22]

It is because philosophers possess cognitive access to the highest or purest ontological entities, immutable Forms, that Socrates advocates their political rule (484b): unlike those who are 'blind' and lack knowledge, they are metaphorically 'keen-sighted', firmly grasping the ideals that a well-functioning regime should engender in society (484c-d). Likewise, when the philosopher rules, Socrates describes them as '[contemplating] things that are *orderly* and *always the same*, that neither do injustice to one another nor suffer it', *admiring* said things and *imitating* their virtues – order, harmony, immutability – to the greatest extent possible (500b-c, *emphasis added*). Manifestly, the immutability and consequent clarity exhibited by Forms is in large part what renders knowledge of them so desirable, especially in contrast to the opacity and changeability characteristic of the ontologically unstable entities grasped by belief.[23] In the *Timaeus* – telegraphed by Plato as building on the Republic's doctrines (Timaeus 17c[24]) – products of craft are designated as 'beautiful' only to the extent that they reproduce the 'look and power' of an intelligible entity that 'always *is* and has no becoming' and thereby is in a 'self-same [i.e. unchanging] condition' (27d-28a). In contrast, if the craftsman (δημιουργός) uses as a model what 'comes to be and perishes and never genuinely *is*' and is '*opined by opinion*' (28a, emphasis added), then their product cannot possess beauty (28b).

Now recall how in the *Republic,* Socrates describes the philosopher-ruler as a 'craftsman [δημιουργός] of temperance, justice, and the whole of popular virtue', whose principal project is stamping 'what is orderly and divine' into the characters of her subjects (Republic 500c-d). The best and most 'beautiful' society is that which instantiates the virtues associated with Forms – unity, order and immutability – to the greatest extent.[25] But on Plato's account, this is just the Kallipolis, the regime ruled by philosophers and possessing educational insti-

[21] Reeve 2006, p. 66; Republic 477a.
[22] cf. Reeve 2006, p. 67.
[23] "Belief" here denotes the epistemic power.
[24] Gill 2007, pp. 267-8.
[25] *Ibid.*, p. 266.

tutions that inoculate it from the diachronically destabilising dynamics of pleonectic desire (545c-546a; 547d; cf. §2.2). The Kallipolis, possessing the most harmonious but also the most *durable* unity out of all the regimes, is thus most Platonically 'good'. Conversely, the degenerate regimes, beset by pleonexia-induced diachronic instability and thus centrifugally impelled toward regime change, more closely resemble the ontologically unstable 'likenesses' grasped by opinion, rendering them less admirable (cf. 396c-e). Thus, contra Gavrielides, it is the instability of the degenerate regimes that renders them inferior (in the *Republic*'s eyes) to the Kallipolis.

4. Conclusion

Taking stock, we can see that although the degenerate regimes can hypothetically achieve synchronic stability under certain conditions, their failure to arrest pleonectic dynamics unavoidably destabilises them in the long run. Furthermore, it is precisely this instability that makes them so objectionable within the polemical context of the Republic, according to which any entity that exhibits instability or disharmony is of an ontologically and thus normatively lower class.

Interpreted this way, the degenerate regimes passage is simultaneously strikingly salient and markedly alien for the modern reader. As illustrated in §2.2, it articulates a proto-Marxian account of how the limitless and structurally incentivised accumulation of wealth by politico-economic elites centrifugally corrodes social order. It also anticipates Adam Smith in arguing that market economies must be moderated through the robust moral acculturation of their participants. Whether this critique works outside the Platonic metaphysical framework – one underpinned by a theory of Forms that at first blush appears distant from the more concrete considerations that conventionally inform our interest in politics – is another question entirely. The analysis of psycho-political dynamics presented in *Republic* VIII-IX is undoubtedly astute, even prescient, but it is difficult to shake the sense that its normative and metaphysical framing is dubious.

However, one of the reasons why Plato's political philosophy continues to fascinate millennia later is its explicit challenge to our deeply held but oft under-scrutinised intuitions around what the fundamental objectives of public life are. In this paper, I have argued that what we moderns find instinctively appealing about the Republic – the theory of psychology and politics – cannot be excised from its peculiar metaphysics and epistemology: in Plato's mind, any attempt to critique the former without some ground in the latter is doomed to fail. We might do well to heed his proposal.

Bibliography

Cooper, J. M. (2021). "Plato's Theory of Human Motivation". In: Cooper, J. (ed.). *Reason and Emotion: Essays on Ancient Moral Psychology and Ethical Theory.* Princeton: Princeton University Press, pp. 118–137.

——. (2021). "The Psychology of Justice in Plato". In: Cooper, J. (ed.). *Reason and Emotion: Essays on Ancient Moral Psychology and Ethical Theory.* Princeton: Princeton University Press, pp. 138–149.

Ferrari, G. R. F. (2007). "The Three-Part Soul". In: Ferrari, G. R. F. (ed.). *The Cambridge Companion to Plato's Republic.* Cambridge: Cambridge University Press, pp. 165–201.

Fissell, B. M. (2011). "Plato's Theory of Democratic Decline". *Polis.* 28(2), pp. 216–234.

Gavrielides, E. (2010). "What Is Wrong with Degenerate Souls in the Republic?". *Phronesis* 55(3), pp. 203–227.

Gill, C. (2007). "The Good and Mathematics". In: Cairns, D. & Herrmann, F. & Penner, T. (Eds.). *Pursuing the Good: Ethics and Metaphysics in Plato's Republic.* Edinburgh: Edinburgh University Press, pp. 251–274.

Green, J. (2021). The First City and First Soul in Plato's Republic. *Rhizomata.* 9(1), pp. 50–83.

Lorenz, H. (2006). *The Brute Within: Appetitive Desire in Plato and Aristotle.* Oxford: Clarendon Press.

Plato & Reeve, C. D. C. (2004). *Republic.* 3rd edition. Indianapolis: Hackett Publishing Company.

——. (2022) *Laws.* Indianapolis: Hackett Publishing Company.

Plato & Kalkavage, p. (2016). *Timaeus.* 2nd edition. Indianapolis: Hackett Publishing Company.

Plato & Zeyl, D. J. (1987). *Gorgias.* Indianapolis: Hackett Publishing Company.

Plato; Cooper, J.M.; Grube, G.M.A. (2002). *Five Dialogues: Euthyphro, Apology, Crito, Meno, Phaedo.* 2nd Edition. Indianapolis: Hackett Publishing Company.

Reeve, C. D. C. (2006). *Philosopher-Kings: The Argument of Plato's Republic.* Indianapolis: Hackett Publishing Company.

On the Slavery Objection to Political Authority

Jonathan Anderson

King's College London

Philosophical anarchists argue that there is no political obligation to obey the state. A classical argument in support of this claim is that the subjection to political authority is no different than the subjection to slavery; therefore, since slavery is morally intolerable, so is political authority.[1] To date, defenders of political obligation have offered surprisingly little by way of systematic reply to this argument.[2] The comparison between slaves and citizens is widely remarked upon, often taken as an illustration of the anarchist's challenge, yet the slave/citizen analogy is rarely treated as a formal objection to political obligation.

So considered, in this paper I analyse the slavery objection to political authority directly and defend a reply to this argument. I argue that slavery and political authority, although similar kinds of normative relations, are not morally analogous, for while the former is committed to the justification of its normative powers by the appeal to ownership, the latter is not. It follows that slavery and political authority are conceptually distinct and that, since this ownership-justification is itself problematic, slavery is wrong whereas the subjection to political authority is not (at least, not inherently for the same reason).[3]

The paper proceeds as follows. In §1, I set out the problem of the slave/citizen analogy, explaining how slavery and political authority appear normatively analogous by their features of command and coercion. From this, I outline the sort of explanation that will be needed in order to explain the difference between slavery and citizenry. §2 then turns to examine this distinction. In §2.1, I reject the traditional reply that slavery is wrong because it involves the alienation of a person's rights. In §2.2, I instead develop an argument that slavery is wrong because, unlike political authority, it involves a claim to the ownership of persons, which is morally unacceptable. Finally, §3 concludes that the comparison between slavery and political authority is only illuminating insofar as it reveals this justificatory difference. Philosophical anarchists would therefore be wrong to argue against political obligation on this basis.

[1] Perhaps the most notable articulation of the argument originates from Mikhail Bakunin's *Statism and Anarchy*. Bakunin writes: 'If there is a State, there must be domination of one class by another and, as a result, slavery; the State without slavery is unthinkable – and this is why we are the enemies of the State' (1873/1990, p. 178).

[2] To my knowledge, the only recent and sustained discussion in reply to the issue that can be found is Wellman & Simmons (2005, pp. 3–29).

[3] Since my focus here is to reply to a form of philosophical anarchism, it will not be my place to argue that there is such an obligation to obey the state.

1. The Problem of the Slave/Citizen Analogy

To state the problem clearly, it will be useful to begin with an analysis of what it means for a person to be subjected to a state of slavery. We can start by providing a distinction between two forms of slavery: "captive slavery" and "bound slavery".[4] By "captive slavery", I mean to refer to instances where a person has *de facto* complete control over another person by virtue of the use of coercive power. This is the sort of coercive power exercised in cases of human trafficking: the relation of slavery is herein purely a matter of physical control. By "bound slavery", I mean to refer to the notion of a person who is *normatively* bound in a state of subjugation, the sort of normative relation presupposed by legal systems of chattel slavery whereby the slave is held to be owned by his master and therefore subject to his master's authority.[5]

Political authority, as I am using the concept, is also a normative relation. A legitimate political authority possesses (i) the power to command and obligate its citizens by fiat, and (ii) the right to coerce its citizens by sanctions or by physical force to ensure compliance with the law when needed. If political authority did not possess these powers, it would be unable to fulfil its function of social coordination.

It should be clear then at the outset that captive slavery will tell us little about the nature of political authority as a normative relation, for the state of being a captive slave is a purely descriptive relation. The mere fact that A can coerce B by itself establishes nothing about the legitimacy of A's coercion of B: all it establishes is A's *de facto* control over B.[6] Hence, if political authority is comparable to slavery, it must be comparable to bound slavery. Here however, the disanalogy is much less obvious, for bound slavery is putatively a normative relation which mirrors political authority. Just like the state, a slave-master claims (i) that he can obligate his slaves by command, and (ii) that he can coerce his slaves by sanctions or by physical force when needed to ensure compliance. So considered, political authority and slavery are both normative relations which stipulate the powers of command and coercion.

However, there is of course an intuitive difference between the two: political states may (*ceteris paribus)* be legitimate authorities, whereas slave-master authority is morally intolerable.[7] It follows that there must be some principled difference between slavery and political authority which explains this moral asymmetry, or else the anarchist will argue that the state's claim to authority is no better than that of a slave-master.

So what is the sort of distinction we are looking for? To see this, let us first consider a toy argument to establish the kind of solution we are after. The simplest kind of explanation sometimes offered for the slave/citizen disanalogy is an appeal to some contingent difference between slavery and political authority as they exist in the world. For example, one might appeal to the *de facto* presence or absence of consent. The thought is simple: legitimate

[4] My terminology is based loosely on Hobbes (1642/1998, §§8.2–8.5) and (1651/1996, p. 141).

[5] *Ibid.*

[6] A standard criticism of this is found in Rousseau (1762/2011, p. 159).

[7] Philosophical anarchists deny the former intuition is correct. However, anarchists would be hard-pressed to deny that (for most people at least) there is a strong pre-theoretical intuition that there is such an obligation to obey the state. See, for example Simmons 1996, pp. 249–250. Himself an anarchist, even Simmons seems ready to acknowledge that 'ordinary people do not experience political life as voluntary and [...] they do experience many of their duties (including their political duties) as nonvoluntary' (*ibid.*).

authority requires consent, and so, since people *do* agree to be citizens but do *not* agree to be slaves, it follows that the former kind of authority may be legitimate whereas the latter kind cannot.

Unfortunately, this thesis very quickly runs aground, for at least two reasons. First, as Hume was wont to remark, it is an unavoidable social fact that very few people have historically made an explicit act of consent to their governments.[8] I am a subject of the British Crown, but I have never sworn an oath of allegiance to the King. Therefore, the consent thesis would invalidate political authority for many (if not most) ordinary citizens. Second, it is historically the case that some people *have* consented to slavery. For example, in the practices of Ancient Rome and even the early Medieval period, people could voluntarily enlist into slavery as an alternative to other forms of punishment.[9] Therefore, the consent theory by itself would not be sufficient to delegitimate all forms of slavery.

What I take this argument to show is that the contingent facts of the matter are not what will explain the distinction for us. If the difference between slavery and political authority were simply a matter of some contingent difference between how they exist in the world, we could always envision instances where slavery and political authority shared the relevant properties and would therefore count as morally on a par.

Instead then, to defend the disanalogy between slaves and citizens, we will need an argument to the effect that the normative relation of slavery – as opposed to political authority – is in some way necessarily voided by the presence of some 'defeating conditions' inherent in such a relation.[10] In other words, it will need to be argued that slavery is *a priori* 'irrational' in a way that political authority is not.[11] This will explain the slave/citizen disanalogy, without the need to appeal to any contingent facts of the matter.

2. Replies

Having set up the problem of the slave/citizen analogy, in this section I will attempt to argue the difference between the two. Social contractarians of the early modern period often appealed to the notion of alienation to argue for the illegitimacy of slavery. The argument typically went that slavery as a normative relation is in principle morally unacceptable because it involved the alienation of the inviolable rights of the subject. I shall first argue against this view and from this critique then set out what I take to be the actual difference between slavery and political authority: the appeal to ownership.

2.1. Slavery as Alienation

Slavery has been standardly considered as the state of an agent being alienated from the totality of his rights. For our purposes, I propose to take the example of Locke. Locke describes that subjection to slavery is the complete subjection to 'Absolute, Arbitrary, Unlimited, and Unlimitable Power'.[12] What this means, on Locke's account, is that a slave is alienated of all his rights before his master, including his rights to life and liberty. A slave-master for Locke is thereby someone with total authority over his subjects, even the

[8] Hume 1748/1985, pp. 471–472.
[9] For historical discussion, see Rio (2012).
[10] Simmons 1993, p. 137.
[11] *Ibid.*
[12] Locke 1689/1988, p. 148 [I, §8].

authority to maim and kill his subjects at will.[13] By contrast, Locke considers that a legitimate political authority will be constrained to respect certain fundamental rights of its citizens. A legitimate state, for example, will never violate its citizens fundamental property rights or their rights to life.[14] Thus, it might be argued that the difference between slavery and political authority is as follows:

> **Alienation thesis**: Political authority is legitimate whereas slavery is illegitimate because slavery involves the total alienation of the person's rights, whereas political authority does not.

This captures not only the conceptual difference between slavery and political authority, but also the moral difference between the two, for Locke holds that the contract of slavery cannot be validly formed. In defence of this claim, Locke appeals to the principle of Roman law that '[n]o Man can, by agreement, pass over to another that which he hath not in himself'.[15] This is the simple rule that a person cannot contract away something which does not belong to him. For example, I cannot contract away your home, because it is not my possession and thus not mine to contract. Following this, Locke argues that the right over one's life is a right that we do not have the control to alienate. His claim rests on the assumption that all people are the 'Workmanship' and 'Property' of God, and so possess a duty of self-preservation in order to fulfil God's purposes.[16] God is the possessor of our right to life and thus it is a right we are not at liberty to yield. Therefore, if no one has the right of authority over his own life, it follows that no one has the authority to contract himself into slavery.[17]

In sum, Locke argues:

(1) A person cannot alienate his right to life.

(2) The contract of slavery requires alienating one's right to life.

(3) Therefore, a person cannot form a contract to become a slave.

By contrast, a person can authorise his participation in a political authority: political authorities do not claim to hold a citizen's right to life; they demand only the citizens' rights of self-governance. This latter right is something within our control and thus something we can alienate to political authority.[18] From this, it may be argued that slavery is unlike political authority due to the kind of alienation involved with one's rights.

Unfortunately, I am not convinced that this argument will work. First, setting aside Locke's theological defence of this claim, we might wonder if premise (1) is true. Consider that, for example, a captain in a war might command his soldiers to put their lives on the line. The fact that the captain can command his soldiers to place their lives at risk would suggest at least that the soldiers are ready to alienate their rights to life for their country.

[13] *Ibid.*
[14] *Ibid.*, pp. 357–358, 360–361 [II, §§135, 138].
[15] *Ibid.*, p. 285 [II, §24].
[16] *Ibid.*, p. 271 [II, §6].
[17] *Ibid.*, p. 284 [II, §23].
[18] *Ibid.*, pp. 352–353 [II, §§129–131].

Granted, the captain cannot simply command them to commit suicide, but the right is not, in principle, inalienable.[19]

Premise (2) is too weak as well, for subjection to slavery is compatible with one retaining (even if only minimally) a right to one's life. Grotius and Pufendorf had both considered that a person could contract himself into a state of slavery, and yet agreed with Locke that the right over one's life remained with God.[20] For Grotius and Pufendorf then, the slavery contract does not require the alienation of all the slave's rights. Rather, slavery is a kind of '*usufruct*', a property relation whereby the owner has the authority to use an item, but not to dispose of it at will. In the slave contract then, a slave-master has a right to control the services of the slave for his own benefit, even if the slave-master cannot damage or destroy the slave in his own person.[21]

It cannot be objected that the view of Grotius and Pufendorf is ahistorical. The slave codes governing the Atlantic slave trade could prohibit the murder of slaves by their masters, even if it otherwise granted the slave-master near-total authority over the slave's liberty.[22] Even given such prohibitions, this did not stop those slaves from being held *as slaves*. Therefore, the subjection to slavery is consistent with the retention of some minimal set of rights.

In any case, even if this were the correct analysis of slavery, the challenge of distinguishing political authority from slavery would persist. In principle, a political authority has the authority to alienate all the rights of its citizens: some constitutions, such as that of the United Kingdom, operate according to a doctrine of parliamentary supremacy, whereby Parliament in principle has the power to deprive any subject of the country of all his rights if Parliament so wishes. Of course, as a matter of political practice, modern governments grant various rights and privileges to their citizens – for the government to deprive its citizens of all their rights would be extraordinary. But this is not the point: as a matter of its sovereignty, if a government needs to exercise a power to deprive its citizens of various rights, a government has the authority to do so. This is, after all, a standard feature of emergency powers, which can be used by governments to enact extremely strong and extensive policies.

Therefore, slavery cannot consist simply in the feature that one is subjected to the near-absolute power over one's rights, for this feature is common to both slaves and citizens. Insofar as the alienation of rights is concerned, the only difference between slavery and political authority would appear to be the usual degree of power exercised. States do not

[19] Cf. Simmons (1993, pp. 137–145). One referee disagrees with the example, writing that '[i]t seems unintuitive to suggest a soldier alienates their right to life in following certain orders that lead to them putting their lives at risk, and that these orders have the same legitimacy as other orders which do not carry this risk'. Unfortunately, this reply seems to be mistaken. It *just is* part of the status of being a soldier that one can be ordered to do something that puts one's life at risk. Granted, such an order will need justification and this justification will need to be much stronger than that given for orders to perform non-life-threatening actions. But still, this threshold can be met and so a captain in war *can* legitimately issue such commands (e.g. in order to save civilian lives, for major strategic advantage against the enemy). By contrast, a captain cannot command civilians to place their lives at risk, no matter the justification. This suggests that there is a normative difference between citizenry and soldiery, and that the difference lies in the authority to use the life of another. This is not 'Absolute' or 'Arbitrary' authority in the Lockean sense of slavery, but it is an instance where one person delegates the authority over his life to another.

[20] Olsthoorn & Apeldoorn 2022, pp. 254–255, 258–262.

[21] *Ibid.*, pp. 258–262.

[22] Morris 1999, pp. 171–172.

usually deprive their citizens of their fundamental rights (such as liberty, speech, or property), whereas slave-masters do.

But it seems then that this view is subject to Hobbes' argument that the slave-contract and the citizen-contract are of one and the same kind. For Hobbes, the contract of a slave to a master and of a subject to a sovereign both involve the total alienation of one's rights to that ruler, who is therein endowed with the absolute powers of command. The difference between the slave- and citizen-contracts lies only in *how* these contracts have come into being. The slave, Hobbes takes it, accepts the contract of servitude due to the threat of a specific party. This is a specific individual who is able to inflict death upon the slave, typically as a matter of conquest. The citizen meanwhile agrees to the contract of servitude out of a general fear of the state of war. This is a fear not directed towards any particular conqueror, but the strategic fear of an unpredictable existence outside political society. To Hobbes therefore, both contracts of slavery and political authority are made out of the motive of fear, but the former is directed to a specific individual whereas the latter is not. Either motive is rational for the slave or citizen because they ensure the aid or survival of the individual. Thus, in either case, Hobbes considers that the motivations for the formation of a contract do not render the contract illegitimate.[23]

To this extent, Hobbes finds that political authority and slavery involve the same relations of power. He describes: [T]he Rights and Consequences of [...] Despoticall Dominion [i.e. slavery], are the very same with those of a Soveraign by Institution [i.e. political society] [...]. So that for a man that is Monarch of divers Nations, whereof he hath, in one the Soveraignty by Institution of the people assembled, and in another by Conquest, that is by the Submission of each particular, to avoyd death or bonds; to demand of one Nation more than of the other, from the title of Conquest, or as being a Conquered Nation, is an act of ignorance of the Rights of Soveraignty.[24]

Yet if Hobbes' analysis is correct, slavery and citizenry remain morally on a par and the anarchists' objection looms.

In light of this argument, one might argue that slavery is illegitimate whereas political authority is not, due to the empirical fact that political states are generally effective at procuring goods for individuals, whereas slavery is not. This is to accept R. M. Hare's claim that '[t]he wrongness of slavery, like the wrongness of anything else, has to be shown in the world as it actually is'.[25] Hare's view is utilitarian: there is nothing in principle wrong with slavery, but we will almost always judge instances of slavery to be wrong because, as practised in the world, 'the slave becomes [...] the most miserable of all creatures'.[26]

It will be hard to argue against this view wholesale without entering the more general debate over consequentialism and deontology, but for present purposes I shall note the following problem: the argument leaves us without any claim to a principled distinction between the status of being a slave and the status of being a citizen, *qua* their statuses as slaves and citizens. This is because the utilitarian argument defaults back to the original issue detected in §1, namely, that the distinction between slavery and political authority will not be a permanent feature of these relations, but a contingent fact. If slavery did somehow

[23] Hobbes 1996, pp. 138–142.
[24] *Ibid.*, p. 142.
[25] Hare 1979, p. 121.
[26] *Ibid.*

produce the greatest overall utility, then according to Hare's argument, it would become morally permissible, just like political authority. So again, the disanalogy breaks down. On this basis, we will need to go beyond the utilitarian reply.

2.2 Slavery as Ownership

We have seen then that the slave/citizen disanalogy is not explained by the kinds of powers involved in slavery and political authority or by their effects. Fortunately, this need not force us to accept the Hobbesian conclusion. Normative relations need not be defined solely by the powers within those relations, but may also be defined by the *justification* of those powers.[27] For example, a parent and a police officer may both have the authority to instruct a child to behave appropriately in public, but the parental right is not the same as the police officer's right due to the functional differences between these respective authorities. The parental right of instruction arises due to the parent's familial responsibility to the proper upbringing of the child. By contrast, the police officer has the right of instruction in order to maintain public peace, e.g. by preventing public nuisance. In this respect, the officer has no intrinsic duty to the child's long-term development. So described, the parental right contains the same kind of power as the police officer's right, but they are not the *same rights* since these powers have distinct forms of justification.

In this spirit, we can argue that the difference between slavery and political authority is not a matter of the kind of the powers exercised in these relations, but of the justification of these powers.[28] We can formulate this thesis as follows:

> **Ownership Thesis**: Political authority may be legitimate whereas slavery cannot be legitimate because slave-master authority depends upon the claim to *own* one's subjects, whereas political authority does not.

Such a thesis is suggested by its historicity. A persistent theme throughout the history of political thought is that slave-masters claim to 'own' their slaves and to hold their authority on this basis. For Aristotle, 'a slave is a living possession' who 'wholly belongs' to his master.[29] Similarly, for Grotius, the slave is a person who belongs to the 'private ownership' of another.[30]

What, then, does it mean for one person to own another?[31] I shall consider that powers of ownership are powers which exist in service of a person's interest in authority. For example, suppose I own a house. As the owner, I have rights to invite and exclude guests to and from my home as I wish, meaning that I am able to determine the duties of others with respect to my possession. This power of command is a result of my interest in possessing authority

[27] Owens 2022, pp. 176, 179.

[28] Owens ms, p. 16.

[29] Aristotle 2001, $1253^{b}30$–$1254^{a}15$.

[30] Grotius 1625/2012, p. 51.

[31] A point of clarification: I am trying to provide an analysis of property rights as *moral* rights, not legal rights. It is surely meaningful to speak of a person's moral rights of ownership, wholly apart from their legal rights. For instance, we can speak of a person's moral rights to his own body or to the fruits of his own labour. These are property rights, but there is conceptual space to speak of these rights independently of their establishment in a legal system. It is the nature of these rights that I am trying to determine.

over the permissions and prohibitions others have with respect to my home. In this sense, I take these powers to be a function of the right of ownership.This point can be generalised to other cases. A person may possess an interest in being able to determine who may or may not use one's body, and so the right to determine permissions with respect to one's body will count as property rights. Broadly construed then, rights of ownership are those rights which are held by virtue of the agent holding an interest in controlling the deontic properties of the item in question.[32]

So understood, ownership is key to the explanation of a slave-master's putative authority. Under chattel slavery, slaves undergo what Oswald Patterson has described as a 'social death'.[33] That is, the slave is considered an 'instrument' of the master without his own personality. When the slave acts in public, he acts as a representative for his master, meaning that his powers of action, promise, contract, etc. are permitted only to the extent that his master wills it so.[34] The slave-master uses the slave as an instrument, to work and act on his own behalf and so in this sense, the slave-master wishes to exercise rights over the slave based on his interests in dominating the slave. This normative relation is explicable if it is considered as a relation wherein the master has an interest in dominating the slave as his own possession. Slavery, so understood, is a form of ownership.[35]

By contrast, while it is the case that I am obligated to obey the authority of the British government, it is not the case that I am *owned* by the British government.[36] Nor does the government claim its sovereignty upon that basis, for it does not assert an interest in control *qua* its own intrinsic interest in dominating its subjects. Granted, a political authority may have an interest in controlling its citizens in the sense that it functions as a social coordinator, but this interest in control is not asserted as though the state has intrinsic interests in power separate from the interests of the good of the people. Rather, legitimate political authority typically asserts its interest in control because this facilitates its ability to advance justice

[32] Owens 2022, pp. 186-187.

[33] Patterson 1982, pp. 4–13.

[34] *Ibid.*

[35] One referee rightly notes that the appeal to ownership does not exhaust the justifications historically given for slavery. For example, Aristotle considers that slavery is justified by natural inequality (see *Politics* I.4–8). He argues that there is a natural inequality between the rational faculties of different racial groups: slaves lack deliberative faculties and therefore have an interest in being commanded by persons who are rationally superior. Hence, for Aristotle, slavery is justified by its utility. But while it is true that Aristotle thought the institution of slavery justified by natural inequality, it remains that he conceived of slave-master authority as an authority of ownership.Book I of the *Politics* also compares the slave/master relation to the husband/wife and parent/child relations. These latter relations are also justified by their putative utility, but Aristotle does not appear to claim that the wife is a 'slave' to her husband or a child a 'slave' to his parents. Thus, what is distinctive about slavery is that it involves a claim to own another (and hence that the slave is a 'living possession'). Here again, ownership appears to be understood as an authority which exists to serve the normative interests of the property-owner. This explains why Aristotle considers slavery the prototypical form of despotism. Aristotle describes the despot as a ruler who governs over a people for his own private interests and only incidentally for the benefit of others ($1278^{b}32$–7). Thus, slavery is a local form of despotism, for it is a form of governance where the master rules over the slave for his own benefit. The argument that slavery is justified by natural inequality seems to operate as a further argument for those who are enslaved as to why they should be bound to that authority.

[36] Owens ms, p. 17.

or overall public utility.[37] We need not be overly prescriptive about the exact details of what the state claims to accomplish: what matters is that the state does *not* consider the justification for its authority to be a result of an interest in the domination of its people. It therefore does not take its powers to be justified by ownership.

From this, the slave/citizen disanalogy comes into frame: the justification of slavery asserts a private interest in the domination of one person by another, whereas the justification of political authority does not. It is this authority-interest over people that is morally unacceptable: we are not entitled to dominate others simply due to an interest in domination for ourselves. Therefore, slavery is in principle a morally unacceptable form of normative authority.

By contrast, this fault does not inhere in the nature of political authority: modern states do not claim to own their citizens, and so contemporary forms of political authority cannot in principle be wrong for the same reason as slavery. In this regard, the state is instead more akin to a trust, a contractual relation whereby a trustee holds authority over a beneficiary for the good of that beneficiary, even if that beneficiary has no direct say in the control of the trust.[38] Trusts are ubiquitous: parents and guardians, for example, are trustors for their children. Yet it would be hard to claim that anything within this moral relation, in its capacity as a trust, is wrong. It seems plausible then, that trusteeship is not intrinsically morally dubious. To the extent then that the state is a trust and not a relation of ownership, political authority cannot be impugned as being a form of slavery. So understood, the slavery objection to political authority is unsuccessful.

Before concluding, I shall make two final points of clarification. First, I should not be taken to be claiming that political authority is morally permissible simply because it claims itself to be a trust. A tyrannical political authority will not escape moral criticism simply by claiming that it is a trust. Trusts can be mismanaged and abused, in such a way that the contract of a trust is voided. Likewise, even if a political authority claims to be acting on the behalf of the good of its citizens, it will not be legitimate if it does not *in fact* do so. It will presumably need to meet other ethical conditions as well in order to be normatively compelling. But note that this does not concede too much ground to the anarchist. Defenders of political obligation will only wish to argue that some states are morally legitimate authorities and the claim at issue here is that political authority is not inherently wrong for the reason that it is a form of slavery. This is not inconsistent with the further belief that a political authority may fail to be legitimate for other reasons.

Second, historically some states *have* claimed to own their citizens. For example, Tsarist authority involved the claim to the ownership of those living under serfdom.[39] In such a case, based on the argument given above, this political authority would be analogous to slave authority and therefore morally illegitimate. Such a conclusion is I think correct, but again need not concede too much to the anarchist. It is not essential that political authority claims its normative powers on this basis and indeed most modern states do not claim to own their citizens. For these states at least, the slavery objection cannot be applied. Understood this way, the slavery objection fails to support anything nearly as strong as a form of *a priori*

[37] For example, the Preamble of the Constitution of the United States of America.

[38] This was the view of some early modern political theorists, for instance, Edmund Burke. See Dreyer 1979, pp. 39, 72-73.

[39] For historical discussion, see Bartlett (2003).

anarchism which rules out in principle the legitimacy of any prospective political authority.

3. Conclusion

In this paper, I defended a reply to the slavery objection against political authority, arguing that the status of citizenry is not analogous to the status of slavery. In §1, I outlined that slavery and political authority are similar kinds of normative relations, involving the powers of command and coercion. In §2.1, I argued that the alienation of one's rights cannot be the distinction between slavery and political authority. Thus, in §2.2, I developed an argument to explain the difference between slavery and political authority based on their respective justifications. Slavery rests on the claim to ownership, whereas political authority does not, and this conceptual difference explains the salient moral differences between the two kinds of authority. Therefore, the comparison between slavery and subjection to political authority is illuminating to the extent that it reveals that the difference between slavery and political authority rests on the notion of ownership. But what it shows more crucially is that slavery and citizenry are not after all one and the same.

Bibliography

Aristotle (2001). *Politics*. In: McKeon, R. (ed.) & Jowett, B. (trans.). *The Basic Works of Aristotle*. New York, NY: Random House.

Bakunin, M. (1873/1990). *Statism and Anarchy*. Shatz, M. S. (trans.). Cambridge: Cambridge University Press.

Bartlett, R. (2003). "Serfdom and State Power in Imperial Russia". *European History Quarterly*, 33(1), pp. 29–64.

Dreyer, F. A. (1979). *Burke's Politics: A Study in Whig Orthodoxy*. Waterloo, Ontario: Wilfrid Laurier University Press.

Grotius, H. (1625/2012). *On the Law of War and Peace*. Neff, S. (ed.). Cambridge: Cambridge University Press.

Hare, R. M. (1979). "What is Wrong with Slavery". *Philosophy & Public Affairs*, 8(2), pp. 103–121.

Hobbes, T. (1642/1998). *On the Citizen*. In: Tuck, R. & Silverthorne, M. (eds.). *Cambridge Texts in the History of Political Thought*. Cambridge: Cambridge University Press.

——. (1651/1996). *Leviathan*. In: Tuck, R. (ed.). *Cambridge Texts in the History of Political Thought*. Cambridge: Cambridge University Press.

Hume, D. (1748/1985). *Of the Original Contract*. In: Miller, E. F. (ed.). *Essays: Moral, Political, and Literary*. Indianapolis, IN: Liberty Fund, Incorporated.

Locke, J. (1689/1988). *Two Treatises of Government*. Laslett, P. (ed.). Cambridge: Cambridge University Press.

Morris, T. D. (1999). *Southern Slavery and the Law, 1619–1860*. Chapel Hill, NC: University of North Carolina Press.

Olsthoorn, J. & Van Apeldoorn, L. (2022). " 'This man is my property': Slavery and political absolutism in Locke and the classical social contract tradition". *European Journal of Political Theory*, 21(2), pp. 253–275.

Owens, D. (2022). *Bound by Convention: Obligation and Social Rules*. Oxford: Oxford University Press.

——. (ms). *Property, Contract and Slavery*. [Unpublished manuscript.]

Patterson, O. (1982). *Slavery and social death: a comparative study*. Available from: https://hdl.handle.net\newline/2027/heb03237.0001.001.

Rio, A. (2012). "Self-sale and Voluntary Entry into Unfreedom, 300-1100". *Journal of Social History*, 45(3), pp. 661–685.

Rousseau, J. (1762/2011). "The Social Contract." In: Cress D. A. (ed. & trans.). *Jean-Jacques Rousseau: The Basic Political Writings*. Cambridge: Hackett Publishing Company, Inc.

Simmons, A. J. (1993). *On the Edge of Anarchy: Locke, Consent, and the Limits of Society*. Princeton, NJ: Princeton University Press.

——. (1996). "Associative Political Obligations". *Ethics*, 106(2), pp. 247–273.

Wellman, C., & Simmons, J. (2005). *Is There a Duty to Obey the Law?* Cambridge: Cambridge University Press.

Simone Weil and the Ethics of Attention: An Investigation Of Compassion, Love and Justice

Anna Maria Robles Fumarola

King's College London

In this essay, I will first define and explain Simone Weil's notion of attention. I will argue that her concept of attention is valuable in understanding our ethical life in interpersonal relations, for which principled, universalist ethics is unsuited. I will show this by discussing two important aspects of our moral lives – moral displacement, and compassion – in which, I will argue, the basic ethical framework of autonomous action is sidestepped by a feeling of *moral necessity*. I will then discuss the objection that attention may involve a level of "moral fatigue", as it might imply that by tending to others' needs and emotional demands, we displace our own. I will argue, however, that the notion of attention can, *contra* its seeming endorsement of passivity, allow us to gain awareness of both our own and others' fatigue, inviting an intersubjective understanding of human vulnerability. In the second part of my essay, I will argue that the value of Weil's concept of attention goes beyond its role in interpersonal relationships but can motivate the feeling and need for justice. This is because attention enables one to seek justice, where it would be reasonable not to. In this way, the notion of attention is valuable in its ability to bridge the traditional gap between the private and public realm, and between justice and care.

1. Attention as Acceptance, Patience and Openness

According to Simone Weil and Iris Murdoch, people have a tendency to run away from reality.[1] It is easier for us to act in this way: it is a form of protecting ourselves from facts that may upset, sadden, and disorient us. But the life of a selfish person, says Simone Weil, though surely more comfortable, is marked by a 'fear and unwillingness to encounter the real', which is made up of suffering and affliction that we may fail to make sense of.[2]

Attention, however, is the activity through which we come closer to grasping the reality we so frequently escape from. For Weil, it 'consists of suspending our thought, leaving it detached, empty and *ready to be penetrated by the object*'.[3] It requires that the subject set aside their own desires, wishes and ambitions. This process of selflessness is set by "decreation", in which we attempt to undo our sense of personal identity. This allows us, in turn, to be open and receptive to the world as it presents itself, so we are not afflicted with those features of our mental life which so usually can distract us. Involved, thus, is a

[1] Weil 1973, p. 159.
[2] Jesson 2017, p. 266.
[3] Weil 1959, p. 72.

level of an initial acceptance of the world. It is in some ways similar to beauty, in so far as 'responses to beauty [...] involve treating objects not as obstacles to be overcome, but as things whose own peculiar value is to be acknowledged, contemplated and respected'.[4]

Attention is patient and contemplative, and cannot be reduced to a particular action, moment or decision. Caprioglio Panizza describes Weil and Murdoch's concept of attention as a form of 'non-possessive desire'.[5] Underlying attention is a desire to seek contact with reality– with a level of truth– which may be hidden from our ordinary perception. But this desire can and never will be fulfilled or acquired, for what we may see in attending will never exhaust the possibilities of can be found. There are infinitely different perspectives from which I may look at the table now in front of me– many inaccessible to me– as well as thousands of wood carvings, which I would never be able to count. But this underlying 'defeat' in our knowledge of the world, only highlights the value of attention. In attending, multiple aspects of reality may reveal themselves to us, previously unknown, or ignored. A striking example of Weil's notion of attention can be found in Pauline Oliveros' idea and practice of 'sonic mediations'. Oliveros' idea consists in changing the way we listen. She proposes a (community-based) contemplative and patient engagement with ordinary sound. This allows us to perceive its nuances, tones, melodies, and pitches – so we come to hear, for example; subtly buzzing drones oozing from a building, or the whispers of a conversation, or the rhythms of another's breath. The purpose of this practice is not achieving a state of zen-like peace, but rather cultivating a sense of heightened awareness of the world. It thus stands in contrast to the notion of action, which comprises 'instances of moving things about in the public world', made possible by our will and the capacity to choose.[6]

The usefulness of such an approach for moral philosophy, however, seems limited. It cannot answer one of its founding questions: "how should I act?", or "what ought I to do?". At most, the notion of attention can seem valuable for the imaginative and aesthetic insight it might bring. But it might seem too contextual to be able to guide our intuitions of what is "right" or "wrong", yet alone determine them. It also seems counter to the key ethical discussions, about, for example, what constitutes a universalizable ethical principle, or how other people's actions can infringe our autonomy. At the forefront in these discussions, is the importance of autonomy, the role of rational decision-making and of (potentially) wrong-making action. Attention, however, due to its focus on 'decreation', seems to lie far away from these shared conceptual tenets.

However, I will argue that the focus on action, universality and the abstract individual is, in some ways, an inadequate framework from which to center our ethical discussions, both for its difficulty in accommodating certain of our ethical intuitions, as well as certain ethical and political values. Morality has generally been defined in terms of 'rationality, self control, strength of will, consistency, acting from universal principles and adherence to duty and obligation'.[7] But I will argue that this perspective follows the logic of a public sphere, concerned with administering justice and law. In the first section of this essay, I will argue that this approach is quite limited in understanding the ethics of interpersonal relations and show how the notion of attention can explain and account for many meaningful aspects

[4] Winch 1989, p. 174.
[5] Panozza 2022, p. 8.
[6] Murdoch 1970, p. 5.
[7] Nicholson 1992, p. 91.

in such relationships. In the second section, I will argue that the value of attention is not limited to this 'private' sphere –if it were, it could become a tool for tribalism and exclusion– but that it is a necessary tool for both thinking about and defending, the value of justice.

For this purpose, I will bring into conversation care ethicists, and Weil and Murdoch's philosophy. I believe all of these authors articulate an ethical position that responds to a neglect in philosophical discussion on questions of care, compassion and love. Though I will not delve into this issue in this essay, my assumption is that this neglect is due to a disregard for the ethical as it is conducted in the private realm, itself a result of the historical exclusion of women in the public sphere and a 'fundamental understanding of moral competence as masculine'.[8][9]

2. Attention and The Problem with Principles

A universalist ethics privileges evaluating our choices and actions according to a universal standard. There is a principle, or moral law, according to which we should behave. The good is defined as behaving according to a certain principle, of which one can make either a good or bad use: one can reason badly, by reasoning according to their own interests, or one can breach a certain right – depending on the theory. Goodness is both posited as a public, external virtue, and it is defined as the opposite of badness. On this view, it is the responsibility of the individual subject to choose to do "good", or "bad", as they are free to do either. The moment of choice and the autonomy of the individual are thus the most important evaluative notions for both punishment and reward. An example of this can be found in our cultural understanding of a "good" versus 'bad' immigrant. Their experience is flattened to fit a universal narrative. They will be rewarded or punished according to their decision to choose between being a good, "law-abiding" citizen, or pursuing a "criminal" life, flouting the norms and conventions of the society they are part of. But as Weil says: 'good considered on the level of evil and measured against it as one opposite against another is good of the penal order'.[10] The dichotomy between good and evil, the preservation of individual autonomy, and the underlying "sense of compensation" refer to a legalistic framework for morality, in which, according to the action one chooses to pursue, we can either commend or punish them.

Attention requires only a mode of reorientation from the part of the subject, in which action arises necessarily as a result of deepened engagement with the specific, individual reality in which we find ourselves. Attention does not require either concepts or principles – it is a form of *perceiving* and *seeing* someone or something, that is made possible by the suspension of our judgment. This ultimately changes the relationship we have to the world, as it gains concreteness. Attention is 'directed at reality and, specifically, at an individual reality: not "people", not "animals", not "nature", but this person, this animal, this blade of grass'.[11] Attending to another person, involves decreasing the distance between oneself

[8] Jaggar 1990, p. 82.

[9] I hope it is clear that I am not claiming that women are naturally more inclined towards questions of love and care, but rather, that women's (oppressive) historical confinement to the domestic sphere, has led to different forms of relational, ethical comportment. This sphere of moral deliberation has been historically lacking in philosophical discussion.

[10] Weil 1952, p. 70.

[11] Panozza 2022, p. 2.

and the other: so that their physicality, their mental state, and their needs gain a heightened form of presence. It is through this profound engagement, that we tend to their needs as a result of a feeling of *necessity*. This upsets the notion of autonomy, insomuch as it is not the case that one chooses to care about another – it is not a decision that we take, but rather a result of the acceptance of another specific individual with specific needs.[12] Caring about another is a result, therefore, of the other gaining a concrete, and urgent, sense of presence, from which our action will ultimately arise. As Weil writes, 'to get rid of the intolerable burden of the good-evil cycle [...], it is necessary either to confuse 'the essence of the necessary with that of the good' or to depart from this world'.[13] The action presents itself as morally necessary, a result of the specific and individual context to which we tend to, and from which it cannot be detached.

The concept of attention is particularly relevant in understanding compassion, in which psychical closeness to the Other is necessary for its expression. When we give attention to the Other, we displace ourselves into their mental state. By doing so, not only do we come closer to recognizing their own reality, but we also may make them feel understood and recognised in their pain. Imagine that Stan is suffering. His friends ask him how he is doing, but they are unwilling to acknowledge the possibility that, in fact, their friend is suffering – perhaps because it would be too painful. They try to reduce the issue or displace their attention onto other matters. They are unwilling to try to see Stan's pain from his own perspective and experience. But we can easily imagine the same sentence being asked to him, by another person. They have a different approach: they are interested in understanding the truth of Stan's suffering, despite the discomfort and pain it might provoke. They attempt to displace themselves into Stan's mental state: they contemplate his suffering in its manner of appearing to him, with him.[14] As a result, Stan may feel listened to, and perhaps, even, unburdened from his pain. Attention therefore shows the importance of the distribution of our mental life in understanding how people can meaningfully relate to one another.

Attention, however, with its account of moral necessity and mental displacement, may be problematic insomuch as it causes us to sacrifice our own wellbeing. It may seem to naturalise and reinforce a certain state of passivity. This seems particularly prescient, especially for women who have historically been burdened with the task of emotional labour, or care-workers who work in an oversaturated and precarious field. They continually respond to the other's emotional needs while displacing their own. As Hoagland writes, 'motivational displacement is one consequence of enslavement'.[15] But I will argue that, though attention may be fatiguing, it does not entail on a Weilian account, the uncritical acceptance of our passivity and subordination.

There is a sense in which fatigue is a necessary consequence of attention. Attention can be a tiring activity, for many reasons. Concentration is difficult and mental displacement emotionally burdening. But I would argue that, even on Weil's account, attention becomes problematic when we depend on it to define our sense of identity. As Weil remarks, 'we must not want to find: as in the case of an excessive devotion, we become dependent on

[12] Noddings 2010, p. 29.
[13] Weil 1952, p. 95.
[14] Jesson 2017, p. 272.
[15] Hoagland 1991, p. 258.

the object of our efforts'.[16] Necessary to attention, as I have remarked before, is patience and openness. If, despite our tiredness and fatigue, we tell ourselves to "keep going", we become blind to a real and material fact: our fatigue. The fixation on such a goal would disinvite the openness that is key to attention. We would be exercising a kind of attention contrary to Weil's, a type of attention marked by "strenuous concentration" and "muscular tension". Attention, Rozelle-Stone writes, 'must be cast upon our own fatigued being – not as a means to eradicate it [...] but so that we can acknowledge our own merciful limits as well as the oppressive nature of the institutions and people who dispense fatigue while claiming to cure it.[17] In practicing attention, we may come to accept that we have neither the energy nor competence in providing care. In fact, if done collectively, attention invites an interdependent understanding of human vulnerability, and the recognition of the ways in which oppressive structures can condition our wellbeing.

3. Attention as a Form of Love and Justice

Though attention may seem necessary in personal friendships, some object about its plausibility as a conceptual tool for the political sphere. When there are millions of people in need of care, it seems impossible for the state to attend to the individual need of its millions of citizens. Attention may be thus only a useful concept for our personal relationships. As a result, some have juxtaposed an ethics of care to an ethics of justice.

There is, however, an important difference in the concept of attention between that in the thought of care ethicists and that of Simone Weil: the notion of impersonality. Noddings states that 'for care theorists it is the living other in all his or her personal, embodied distinctiveness who addresses us. The other is *not* "exactly like us"'.[18] But, as Hoagland points out, such analysis 'fears the proximate stranger, and it ignores the distant stranger'.[19] I believe Weil can overcome this difficulty. In *Human Personality*, Weil explains that the ethical value of attention cannot lie in perceiving certain aspects of another's personality – as this in fact would also involve our personal judgements of what is desirable of people, a supposition which Weilian attention disposes of. If, the fact that I am a philosophy student, or that I am Spanish, or that I am 170cm tall, was what would make me valuable, or worthy of attention, Weil claims then, you could easily put out my eyes, for these elements of my personality might well stay the same.[20] But there is something else, which, "stops my hand". It is that, according to Weil, you would be violating a sacred, *impersonal* condition of human life: our expectation that good, and not harm, will be done to us. Weil explains this in a striking passage:

> 'At the bottom of every human being, from earliest infancy until the tomb, there is something that goes indomitably expecting, in the teeth of all experience of crimes committed, suffered and witnessed, *that good and not evil will be done to him. It is this, above all, that is sacred in every human being*'.[21]

[16] Weil 1952, p. 117.
[17] Rozelle-Stone, 2021.
[18] Noddings 2010, p. 34.
[19] Hoagland 1991, p. 260.
[20] Weil 2005, p. 51.
[21] *Ibid.*, p. 1.

This is the source of injustice, which expresses itself in the cry of: 'why am I being hurt?', which may not even be expressed in language. But it is through a form of attentive silence that this cry can be heard. Attention, for Weil, thus arises from a deeper interest in an impersonal condition of human beings – in their expectation of good, and the depth of injustice when that expectation is violated– rather than in a specific interest in their behaviour or status. For this reason, the value of attention goes beyond its applicability to understanding interpersonal relations.

The concept of attention is valuable because it surpasses the oppressive division between the personal sphere (care) and the public sphere (justice), by bringing in love to the question of justice. According to Weil, justice is present when there is equal consent on both sides.[22] But where those relationships of power are unequal, there will be injustice. In such a scenario, the reasonable man will not seek the consent of the powerless, because 'reason does not allow that one might choose an uncertainty – consent being granted– over a certainty–consent being forced'.[23] But attention, because it involves the decreation of ourselves, will allow us to seek their consent and pursue justice, especially as it enables us to go beyond the mark of human language and personality. That is why Weil says that the 'spirit of justice and truth is nothing else but a certain kind of attention, which is pure love'.[24] Attention, through love, allows us to see injustice, and tend to needs, both basic and 'spiritual'.[25] The satisfaction and provision of these needs will necessarily be an intersubjective matter and a political question, as 'how one arrives at a need is a matter of social concern'.[26] But the ability to recognise such needs is only possible through love: through the ability to seek the impersonal aspect (the fundamental desire for goodness) within human beings.

4. Conclusion

Simone Weil's notion of attention is a practice that consists in decentring ourselves– our own desires and ambitions– so we can become open, and receptive to the world around us, including the suffering and pain which we usually avoid acknowledging. Attention, as I have argued, can explain important and meaningful aspects of interpersonal relations, such as the feeling of moral necessity and the ability to express compassion. It can be fatiguing and demanding, because it involves a deep awareness of other's (and our own) pain and suffering. But it does not normalize it either, for attention must be always grounded in a fundamental openness to the world, and thus to each other's fatigue. Attention also has a profoundly social and political dimension because it arises through a recognition of an impersonal aspect of the human condition, of humans' ultimate fragility: their expectation of goodness. Unequal power structures not only violate this expectation but make the cries of protest and injustice silent. But attention, can overcome these mechanisms of political recognition, and thus seek the desire for justice. In this way, fostering attention could allow for a justice that aims at securing not just the breadcrumbs of lessened exploitation, but the possibility for human flourishing.

[22] Thomas 2020, p. 156.
[23] *Ibid.*
[24] Weil 2005, p. 72.
[25] To give an account of Weil's notion of spiritual needs would go beyond the scope of this paper. But I hope it is clear that Weil's notion of need incorporates not just our most basic and primary physical needs, but also social and intellectual needs (such as the need for order, and the need for truth).
[26] Tronto 1993, p. 164.

Bibliography

Cameron, S. (2003). "The practice of attention: Simone Weil's performance of impersonality". *Critical Inquiry*, 29(2), pp. 216-252.

Hoagland, S. L. (1991). "Some concerns about Nel Noddings' caring". In: Card C. (ed.). *Feminist Ethics* Lawrence: University Press of Kansas, pp. 246-264.

Jaggar, A. (1990)."Feminist ethics: Projects, problems, prospects". In: Card C. (ed.). *Feminist Ethics*. Lawrence: University Press of Kansas, pp 78-104.

Jesson, S.(2017). "Compassion, consolation and the sharing of attention". In: Rozelle-Stone, R. (ed.). *S. Weil & Continental Philosophy*. Lanham: Rowman & Littlefield International, pp 273-320.

Murdoch, I. (1970). *The sovereignty of good*. London: Routledge & K. Paul.

Nicholson, L. (1993). "Women, Morality and History". In: Larrabee, M. (ed). *An Ethic of Care: Feminist and Interdisciplinary Perspectives*. London & New York: Routledge, pp. 87-102.

Noddings, N. (2010). *The maternal factor: Two paths to morality*. Berkeley: University of California Press.

Panozza, S. C. (2022). *The ethics of attention: Engaging the real with Iris Murdoch and Simone Weil*. New York: Routledge.

Rozelle-Stone, R. (2021). *How should we live our fatigue? Moral philosophy for an age of exhaustion. ABC Religion & Ethics*, May 13. Available from: https://www.abc.net.au/religion/living-our-fatigue-philosophy-for-an-exhausted-age/13343166#

Thomas, C. (2020). *Simone Weil: The ethics of affliction and the aesthetics of attention. International Journal of Philosophical Studies*, 28(2), pp. 145-167.

Tronto, J. C. (1993). *Moral boundaries: A political argument for an ethics of care*. London: Routledge.

Weil, S. (1952). *Gravity and grace*. London: Routledge & K. Paul.

———.(1959). *Waiting for God*. New York: Capricorn Books.

———.(1973). *Waiting for God*. New York: Harper & Row.

———.(2005). *Simone Weil: An anthology*. S. Miles, (ed.). London: Penguin.

Winch, P. (1989). *Simone Weil: The just balance*. Cambridge: Cambridge University Press.

Justification as an Evaluative Concept: A New Argument for Externalism

Guy Dinwiddy Smith

King's College London

In this essay I offer an argument to support the thesis that justification for belief depends on more than how things seem from the believer's own perspective. In arguing this, I endorse *externalism about justification (E)* rather than *internalism about justification (I)*. Srinivasan describes *(I)* as the thesis that 'epistemic justification is a matter of fit [. . .] with how things look from one's own perspective on the world'.[1] However, (I) can be more clearly described in terms of supervenience thus: (I) is the thesis that the justificatory status of an agent's beliefs supervenes on that agent's mental states. This means that if two agents have the same mental states, the justificatory status of their beliefs are necessarily the same. Like Srinivasan, I reject *(I)*, although my argument for rejecting it differs in method. Rather than presenting cases which are intuitively "vindicated" by *(E)*, I argue that *(E)* is entailed by the notion of justification itself. I present this argument in favour of *(E)* thus. First, I argue that justification is an evaluative concept. Therefore, I argue, epistemic justification is judged rather than merely described, and thus *depends* in an important way on an evaluator. I then argue that the evaluator's own beliefs affect their judgment or evaluation as to whether a belief is justified. Thus, I argue, justification for belief necessarily depends on more than the mental state of the believer and how things seem from the believer's own perspective.

1. Justification is an evaluative concept

I now expound and defend the thesis that justification is an evaluative concept. First, I show that this thesis is *prima facie* plausible, and acceptable to philosophers on both sides of the debate. Brent Kyle notes that epistemic justification is commonly taken to be 'conceptually equivalent to [. . . e.g.] "S has good reason to believe p [. . . or] S's belief in p is permissible'.[2] Concepts such as "good" and "permissible", he says, are "thin" evaluative concepts, and therefore 'it's plausible that *justification* is thin [$\therefore$ evaluative].'[3] Moreover, the evaluative content of "justification" seems widely acknowledged in the internalism / externalism debate (henceforth I-E debate), and is reflected in the evaluative language used in elucidations of justification across both internalist and externalist sources. Feldman and Conee, in their defence of *(I)*, describe cases in which 'one person's belief is *better* [my emphasis] justified than the other's.'[4] Moreover, Srinivasan, in her defence of a "radical"

[1] Srinivasan 2020, pp. 399-400.
[2] Kyle 2011, p. 43.
[3] *Ibid.*, p. 43.
[4] Feldman and Conee 2001, p. 3.

(E), talks about the agents in her cases potentially being 'epistemically better off' if they had access to an account which would explain their beliefs.[5] Elsewhere, she describes justification as an 'epistemic good'.[6] Therefore, the prevalence of evaluative language on both sides of the I-E debate demonstrates that an analysis of justification as an evaluative concept is *prima facie* plausible.

However, evidence of acknowledgement of a link between justification and evaluative content will not suffice. Rather, it is necessary to argue that justification *is* an evaluative concept. To this end, I propose that the term "justification" can be interpreted in two ways. I term these interpretations "justification as grounds" and "justification as (account of) reasonableness". In differentiating these interpretations, it is useful to compare the following sentences:

(A) "The war has *justification*".["justification as grounds"].

(B) "The defendant put forward a *justification* for his actions to the jury."["justification as (an account of) reasonableness"].

In A, justification is best understood as grounds or reasons for the war. In B, justification is best understood in terms of exculpation, as the defendant attempts to *put forward good reasons* for his actions. Thus, justification appears to be a property of something in A, and an act or account in B. I argue that internalists should prefer the interpretation of "justification as grounds". Indeed, Feldman and Conee endorse this analysis, saying 'we note that a theory requiring for justified belief that the believer... be able to articulate those reasons or persuade others is not an internalist theory by our standards.'[7]

"Justification as reasonableness" seems more clearly evaluative than "justification as grounds". I now offer two arguments in support of justification being an evaluative concept, using this distinction of interpretations. First, I offer a semantic argument that "justification as grounds" collapses into "justification as reasonableness". Second, I argue that, even if the first argument is unacceptable, "justification as reasonableness" is to be preferred in the epistemological context.

First, the semantic argument. I examine the active form of the verb "justify" and then show that my analysis extends to the passive form "is justified" and the noun "justification". "Justify" is often used as a ditransitive verb like "show" or "give". In these cases, the verb appears in sentences of the form "x justifies A to y". However, it can appear as a seemingly monotransitive verb in sentences of the form "x justifies A". Syntactically, therefore, "justify" can behave in the same way as "naturally" monotransitive verbs. Nevertheless, I argue that in the latter sort of case, an indirect object (y) is understood, even if not explicitly mentioned. For example, to appropriate sentences A and B above, one could *grammatically* say "I justify the war" or "I justify my actions" independent of an indirect object (y). However, these sentences in which "justify" appears to be a monotransitive verb seem disanalogous to sentences using more naturally monotransitive verbs such as "I bake the cake" and "I punch the wall" because the latter sentences express complete and satisfying pieces of information

[5] Srinivasan 2020, pp. 396, 398.
[6] *Ibid.*, p. 414.
[7] Feldman and Conee 2001, p. 15.

regardless of any additional pragmatic context. Conversely, the sentence "I justify my actions" provides incomplete information without an elucidation (or implicit understanding) of *to whom* the war was justified. This might possibly be demonstrated by comparing the effect of adding context to the sentences. In the sentence "I go into the kitchen and make the cake", the first half of the sentence (which provides context) does not imply an indirect object. Even the sentence "I go into my brother's kitchen and make the cake" does not imply that the cake was made *for* my brother. By contrast, "I go into the House of Commons and justify the war" or "I go into court and justify my actions" both imply that the direct object is *being justified to an indirect object*, namely the House of Commons or the court respectively.

It might be objected that, in focusing on the active form of the verb "justify" rather than "justification" as a noun, I am unfairly assuming the interpretation of "justification as reasonableness to x". However, on the basis of my argument that the verb "justify" can *always* have at least an implicit indirect object or audience, I suggest that the noun "justification" similarly can always have such an implied audience, which would be problematic for the independence of the interpretation of "justification as grounds". In particular, I argue that the act "to justify" is prior to the noun "justification" insofar as something can be justified to y without having a property of "justification".[8] However, the reverse is not true. A war *can* be justified to a small number of fanatics without necessarily having "justification" in the sense of [objective] grounds, yet it is more difficult to say that a war "has justification" in the sense of grounds without "being justified (at least implicitly) to someone." Thus, "justification as grounds" is not comprehensible without the prior notion of "justification as reasonableness", and the former ought to be explained in terms of the latter.

Second, I argue that, even if my argument above is not accepted, "justification as reasonableness (to (y))" is to be preferred in the epistemological context. Epistemic justification admits of qualitative degrees insofar as one belief can be "better justified" than another. I argue in this section that this makes more sense if justification involves an evaluator (y) to whom a belief is justified than if justification is merely a property of a belief. On this account, a belief is better justified if evaluated by (y) as being "better" or more reasonable in terms of its relationship to the truth.

An internalist might object that an account can be provided to make sense of "better justification" using the interpretation of "justification as grounds". Feldman and Conee, in their paper arguing on behalf of internalism about justification, give an example where "Bob" and "Ray", having read that it will be warm outside, believe that it is indeed warm. Bob then goes outside and feels the warmth, whereas Ray does not, continuing to believe that it is warm only on the basis of the newspaper. Feldman and Conee argue that 'Bob's belief is better justified... [because it is] enhanced by his experience of feeling the heat and thus undergoing a mental change.'[9] On this interpretation, the internalist might argue, "better justified" simply means "having better *internal* grounds / reasons to believe." Likewise, in the second case given by Feldman and Conee, the novice birdwatcher (by contrast with the

[8] This priority, I argue, is conceptual not causal. The idea that "justify" is causally prior to "justification" seems to create a causality dilemma. "Justify" being *conceptually* prior to justification is the point here insofar as I argue that a grasp on the idea of justification as "an account of reasonableness" is needed to understand the idea of "justification as grounds".

[9] Feldman and Conee 2001, p. 3.

expert) 'has no good reason to believe that it is a woodpecker and is not justified in believing that it is.'[10] In both cases, however, the evaluative language points to a problem for this account. If Bob and the expert are "better justified" than Ray and the novice respectively because the former have good / better reasons compared to the latter, how can these reasons be understood as good or better *except* because they are so evaluated by an evaluator? The internalist could seek to rid justification of this evaluative baggage, thus avoiding the counter-objection, by saying that Bob and the expert have *more* grounds than Ray and the novice rather than *better* grounds. However, this seems untenable. I might have more (faulty but manifold) reasons to believe that climate change is occurring than a climate expert, but her belief that it is could still be better justified than mine. This can only be explained by saying that the expert has *better* grounds than I do. Therefore, the evaluative language returns, and with it the problem for the internalist's account of better justification understood as "justification as (better) grounds". Thus, I conclude that an evaluator is essential to make sense of the idea of "better justification" and it makes more sense to understand epistemic justification as a type of "justification as (an account of) reasonableness (to$y$))". As this is an evaluative understanding of justification insofar as it essentially involves the idea of an evaluator to whom something is justified, I argue that epistemic justification is an evaluative concept.

2. Evaluation depends on the assessor

In the previous section I argued that justification is an evaluative concept, and that epistemic justification is better understood as a sort of "justification as (an account of) reasonableness (to$y$)" than as a sort of "justification as grounds". On this basis, I now elucidate the concept of epistemic justification and show that, like other forms of "justification as reasonableness (to$y$)", it importantly *depends on* the evaluator. The concept of epistemic justification can be elucidated by comparing it to another sort of "justification as reasonableness (to$y$)", the exculpatory concept of justification. On my account, a belief's being justified *is a thing judged by an evaluator rather than merely described by them*, in the same way that an act's being justified in the exculpatory sense is judged rather than described by a jury or judge. In the moral or exculpatory case, such as in the context of a court, whether an act is justified or not is *decided* by an evaluator or group of evaluators, based on their understanding of what is the case, and, critically, their understanding of what is right or good. In other words, moral evaluation involves an assessor evaluating a decision or action in terms of its (apparent) relationship to "the moral or legal good". Analogously, I argue, evaluation of epistemic justification involves evaluating belief in terms of its (apparent) relationship to the truth. In both cases, where the assessor decides that the act or belief *is* justified and says "$x$is justified", the truth conditions of that sentence depend on the assessor's evaluation, rather than being the result of something inherent in x itself.

I will illustrate this theory of justification with reference to Srinivasan's "bad ideology" cases, but first it is necessary to make two points.[11] First, it is important to note that, on this theory of justification, the evaluator need not know the *actual* truth, but merely the apparent truth to evaluate another's relationship to the (apparent) truth. Second, it is

[10] *Ibid.*, p. 3-4.

[11] Conditions of "bad ideology" are 'conditions in which pervasively false beliefs have the function of sustaining, and are in turn sustained by, systems of social oppression.' (Srinivasan 2020, 407.)

important to note that this theory does not commit me to relativism about truth. In the same way that acts are morally evaluated in terms of their relationship to the apparent good and this does not preclude there being an actual objective good, so beliefs are epistemically evaluated in terms of their relationship to the apparent truth and this does not preclude there being an actual objective truth. Evaluators can, in short, evaluate others' beliefs as justified (i.e. appropriate in terms of their relationship to the (apparent) truth) or not, regardless of their own knowledge of the actual truth. This independence of justification and truth might seem anathema to some, but, as Zagzebski has pointed out, it has the advantage that Gettier problems are avoided. This is because it introduces an 'element of luck [into justification and thus knowledge] ... so great that alleged counter-examples based on luck do not count against it.'[12]

Therefore, justification, on this theory, is:

> **Justification**: a belief x's reasonableness (to y) in terms of its apparent relationship to the truth.[13]

Justification is therefore located in (the judgment of) an evaluator of a belief rather than in the belief itself. I argue that this idea of justification makes sense of the 'externalistic intuitions' elicited by Srinivasan's "bad ideology" cases.[14] First, it is necessary to recall the salient features of three of Srinivasan's cases and her diagnoses. Her case CLASSIST COLLEGE involves working-class Charles, who has 'a dependable sensitivity to classism', yet is unable to persuade the Master of the College that he has faced the classism that he has in fact faced.[15] Likewise, in RACIST DINNER TABLE, Nour has a 'subconscious sensitivity' to racism and, using this reliable method, perceives that her host is racist, as he in fact is.[16] Finally, in a case titled ABORTION, an abortion-disapproving Thomas has a 'genuine sensitivity to the moral truth, namely, the truth that abortion really is wrong.'[17] Srinivasan describes these cases as "bad ideology" cases, and argues that 'externalism, in its insistence that justification can supervene on facts external to the agent's own ken – is poised to vindicate what we might think of as a *structural* rather than merely *individualistic* notion of justification.'[18] Srinivasan thus diagnoses that 'Nour and Charles's beliefs are justified [she seems to mean "grounded"] *because their group membership allows them to pierce through bad ideology.*'[19]

I, however, want to argue that the intuition that Nour and Charles's beliefs are justified (in my sense) is elicited *ex hypothesi*. Focusing on CLASSIST COLLEGE, I argue that the reason we intuitively think that Charles is justified in his belief is because Srinivasan tells us as "omniscient observers" to the world of the thought-experiment that Charles's belief is (a) true and (b) the product of Charles's 'dependable sensitivity to classism.'[20] Thus, as

[12] Zagzebski 1994, pp. 65-73, 72.
[13] Thus, "to give justification for a belief" would mean to "give an account of a belief's reasonableness (to y) in terms of its apparent relationship to the truth."
[14] Srinivasan 2020, p. 407.
[15] *Ibid.*, p. 397.
[16] *Ibid.*, p. 396.
[17] *Ibid.*, p. 423.
[18] *Ibid.*, p. 410.
[19] *Ibid.*, p. 411.
[20] *Ibid.*, p. 397.

omniscient observers, we automatically evaluate Charles's belief as appropriately connected to the truth, i.e. justified, *ex hypothesi*. On the other hand, Charles's true belief is *not* justified (in my sense) *to the Master* in the world of the thought-experiment. There is an important disconnect between the reasonableness of Charles's belief (and method of coming to that belief) to us as omniscient observers, and its reasonableness to the Master. It might be objected that I am committed to saying that Charles's belief does not (in my sense) have justification and does not amount to knowledge, despite being true and the product of a reliable method. However, this is not quite accurate. I am rather committed to saying that Charles's belief does not have justification *to the Master*, but *does* have justification in the eyes of us omniscient observers. We evaluate his belief as justified and therefore an item of knowledge, but the Master does not.

This alternative account, I argue, makes sense of Srinivasan's idea that Charles would be 'better off, epistemically speaking, if he had available to him the phenomenon of false consciousness as a debunking explanation of the Master's testimony'.[21] If Charles had this phenomenon available to him, and if the Master was amenable, he would be able to persuade the Master that he had an appropriate connection to the truth. Consequently, the Master would evaluate Charles's belief as justified. Furthermore, I am not committed to arguing that subjects who lack epistemic justification in the eyes of local observers in (for example) conditions of bad ideology *necessarily* lack epistemic justification entirely. Rather, they might have epistemic justification in the eyes of a different observer. I am also, crucially, not committed to arguing that such subjects lack true belief. Finally, I argue that the idea of "bad ideology" is only comprehensible from the point of view of an omniscient observer. It is only "bad" insofar as it detaches a "good" believer in the truth from justification in the world of the thought-experiment in a way that is "bad" in our eyes as omniscient observers.

Therefore, I argue that insofar as justification is located in the judgment of an evaluator, it depends on the evaluator in an important way.

3. Justification depends on more than how things seem from the believer's own perspective

Finally, I argue that the evaluator's judgment will essentially be affected by their own beliefs. If this were not true, and if justification could be determined by an evaluator looking only at internal facts such as the believer's mental state, my argument would not be an argument in favour of *(E)*. Srinivasan's "bad ideology" cases again provide examples in which (on my interpretation) beliefs which ought in our omniscient eyes to be afforded justification, are not. In these cases, the disconnect results from the difference between the evaluator's *own* lack of true belief *ex hypothesi* and our true belief *ex hypothesi*. This disconnect is a fact external to the believer. Nevertheless, the analogous internalist cases might suggest that beliefs can be evaluated by an assessor who looks at only the internal facts of the believer's mental state. Again, however, the problem is that we, as omniscient observers, have access to knowledge of justification-desiderata and the *actual* truth that force us to evaluate one belief as better justified than the other *ex hypothesi*. A variation on the first of Feldman and Conee's cases brings out this problem.[22]

[21] *Ibid.*, p. 398.
[22] Feldman and Conee 2001, p. 3.

Imagine that Bob, Ray and an additional man Paul are sitting in the air-conditioned hotel lobby. Bob and Ray have read in yesterday's newspaper that it will be very warm today and, thus, believe that it is very warm outside. At this point, Bob (who is from a cold climate country unbeknownst to Paul) goes outside and decides that it feels very warm (to him). Ray continues to believe that it is very warm only from the evidence of the newspaper. It is in fact not (relatively speaking) warm outside. Both Bob (now back inside) and Ray then tell Paul that it is very warm outside. Both Bob and Ray come to false beliefs through unreliable methods. However, Paul (reasonably) judges that Bob is better justified in his (apparently true yet actually false) belief than Ray. In this case, what is doing the work in making Bob's belief better justified (to Paul) is not Bob's 'internalising of the actual temperature', which ought in fact to be a reason *against* justification, but rather Paul's *own beliefs* that Bob has an accurate gauge for the temperature, and that therefore Bob's first-hand experience of the outside temperature affords Bob's belief better justification than Ray's belief.[23] Consequently, the extent to which Bob's belief is justified to Paul depends on more than factors internal to Bob's mental state. It *also* depends on Paul's own beliefs. Thus, aside from cases in which justification (in the eyes of omniscient observers) follows *ex hypothesi*, I argue that justification can be shown to depend on the beliefs of the evaluator. Therefore, justification depends on more than how things seem from the believer's own perspective.

4. Conclusion

I conclude that justification for belief *does* depend on more than how things seem from the believer's own perspective. This is because justification is an evaluative concept. Thus a belief's justification is a thing judged by an evaluator rather than described, and, since an evaluator's judgment *necessarily* depends on their own beliefs, justification for belief depends on more than internal facts about the believer's mental state alone because it also depends on (at least) the evaluator's own beliefs, and the evaluator, moreover, does not necessarily need to examine the internal facts of the believer's mental state when evaluating whether their belief is justified.

[23] *Ibid.*, p. 3.

Bibliography

Feldman, R., & Conee, E. (2001). "Internalism defended". *American Philosophical Quarterly*, 38(1), pp. 1-18.

Kyle, B. G. (2011).*Knowledge as a thick concept: New light on the Gettier and value problems*. (Doctoral dissertation, Cornell University).

Srinivasan, A. (2020). "Radical externalism". *The Philosophical Review,* 129(3), pp. 395-431.

Zagzebski, L. (1994). "The inescapability of Gettier problems". *The Philosophical Quarterly*, 44(174), pp. 65-73.

On the Logical Character of Emotive Language: Evaluating Williamson's Anti-luminosity Argument

Mette Andersen

University College London

Abstract

It is a deeply held intuition that we have privileged first-person epistemic access to our phenomenological experiences. For instance, we tend to think that whenever we feel cold, we are in a position to know that we feel cold. Let us call this the "luminist intuition". Timothy Williamson argues that the luminist intuition is false by deriving a contradiction from a set of inconsistent premises including the luminist intuition. Contrary to common objections to Williamson's argument, this paper will exploit the fact that Williamson's seemingly trivial premises are not trivial at all: Their triviality hinges upon the hidden assumption that they can be true or false (i.e. they are truth-apt), and the premises are truth-apt only if there is an independent fact justifying their truth-aptness. However, such a fact cannot be identified. Therefore, the most trivial premises are not truth-apt, and Williamson's argument is unsound.

I have reason to doubt my knowledge of other people's mental states, but it seems trivial that I know my own mental states. For instance, it seems that knowledge of my own sadness follows trivially from my phenomenological experience of tightness in the chest, unease, or some other sensation, which only I can access. Nonetheless, in *Knowledge and Its Limits* (2000), Timothy Williamson derives a contradiction from this intuition and some other supposedly trivial premises, concluding that there are no non-trivial mental states (e.g. being in pain and being cold) such that we can know that we are in them whenever we experience them first-hand.

The literature is torn between anti-luminists endorsing Williamson's argument and luminists objecting that its conclusion is false. This dissertation proposes an objection, which is not currently represented in the literature on Williamsons' argument. It does not fall into the luminist objection camp nor the anti-luminist endorsement camp. Rather, it is an objection to the truth-aptness of Williamson's seemingly trivial premises. I argue the following: Sensation expressions of the kind "I feel cold" are not truth-apt statements. Since Williamson must assume the truth-aptness of all of his premises in order for his argument to be sound, it follows by *modus tollens* that Williamson's argument is unsound.

I will proceed to this conclusion as follows: In section 1, I will outline the thought experiment on which Williamson's anti-luminosity argument is based. The argument trivially assumes "I feel cold in the morning" and "I do not feel cold at noon" as premises in his thought experiment. In section 2, I argue that the anti-luminosity argument is sound only

if its most trivial premises can be assumed true (i.e. they are truth-apt). I then reject the truth-aptness of the most trivial premises of Williamson's argument on the basis that "I feel cold", beyond appearance, does not report a sensation. Since no sensation would confirm its truth or falsehood, "I feel cold" is not truth-apt. Rather, it is a kind of sentence which is better accounted for as a non-truth-apt expression. This contradicts the necessary assumption that the most trivial premises of Williamson's argument are truth-apt, which undermines the soundness of Williamson's argument. In section 3, I respond to some pressing objections.

1. Williamson's Anti-luminosity Argument

The anti-luminosity argument reconstructs the following thought experiment due to Williamson:[1]

> **Cold Morning**: Consider a morning when one feels cold (P1). Over the course of the morning, one gradually warms up until one feels hot at noon (P2). Supposedly, the feeling of coldness is exactly simultaneous with one's being in a position to know it. According to Williamson's definition, one is in a position to know it if and only if 'no obstacle blocks one's path to knowing [it] [...] the fact is open to one's view, unhidden, even if one does not see it yet'.[2] So, whenever one feels cold, one is in a position to know that one feels cold (LUM). Throughout one's process of heating up, one focusses one's attention on determining whether one feels cold (FCS). For Williamson, knowing whether one feels cold is guaranteed provided that (1) one is in a position to know that one feels cold and (2) one is focusing one's attention on determining whether one feels cold (KNW).[3] So, the conditions are in place for us to infer that one actually knows that one feels cold whenever one feels cold. Suppose further that one is not able to discriminate between levels of coldness over milliseconds, so if one knows that one feels cold at a time *t*, then one millisecond later (*t+1*) one also feels cold (MAR).

From the premise that one feels cold in the morning (P1) in conjunction with LUM, FCS, KNW and MAR, Williamson derives that one feels cold at noon. However, we assumed that one does not feel cold at noon (P2). This is a contradiction. From this contradiction, Williamson concludes that whenever one feels cold, one is not in a position to know that one feels cold. In short, luminosity (LUM) fails.[4]

1.1 Common objections

Since everything follows from a contradiction, it is not necessarily the case that LUM is false. However, P1 (that one feels cold in the morning) and P2 (that one is not cold at noon) are considered to be statements that follow trivially from the set-up of the thought experiment (we will later disagree radically with this assumption), so luminists tend to respond that MAR (or assumptions on which MAR hinges) are false instead of LUM.[5] Such

1. Williamson 2000, pp. 96-97.
2. Williamson 2000, p. 95.
3. *ibid.*
4. See Appendix for an explication of Williamson's argument.
5. see Berker 2008; Cohen 2010; Ramachandran 2009; Zardini 2013.

responses are legitimate alternatives to Williamson's conclusion, since they preserve the logical form of the argument.

However, this dissertation will not evaluate these common objections. By objecting to MAR, these responses take the truth-aptness of all premises (including P1 and P2) at face-value. I will propose an alternative objection, which takes seriously a point that current objections tend to overlook: The trivial assertion of P1 and P2 is far from trivial.

2. Is "I feel cold" truth-apt?

The trivial assertion of P1 and P2 entails a commitment to their truth-aptness according to the following reasoning: Williamson assumes P1 "I feel cold in the morning" and P2 "I do not feel cold at Noon" as premises following trivially from the set-up of his thought experiment. It follows trivially that he also assumes that they are truth-apt, i.e. they can be true or false. However, '[. . .] according to a generally accepted definition of logical inferences, only sentences which are capable of being true or false [i.e. truth-apt] can function as premises or conclusion in an inference'.[6] So, if the premises are neither true nor false, this undermines their truth-aptness together with the possibility that they can function as true premises in Williamson's argument.

I will argue that "I feel cold" is not truth-apt in the following way: First, I will consider three accounts attempting to make sense of "I feel cold" as a truth-apt report of a factive (and hence, knowable) condition. On account 1, "I feel cold" is a truth-apt report of an internal sensation. On account 2, "I feel cold" is a truth-apt report of a physical neuron process on which the sensation supervenes or is identical. That is, the private sensation and the physical neuron process occur simultaneously, such that the sensation is or can be inferred from its physical manifestation. On account 3, "I feel cold" reports a list of external facts related to one's behaviour and environmental conditions. However, I reject accounts 1, 2 and 3. This leaves us with account 4, which renders "I feel cold" a non-truth-apt fact-independent sentence. From this argument, it follows that "I feel cold" is not truth-apt and cannot function as a true premise in Williamson's argument. I will begin with account 1.

2.1 The private sensation account

Firstly, we will contemplate whether "I feel cold" can be accounted for as a truth-apt report of a private sensation, which only the subject can access. We will consider this question through a thought experiment in Wittgenstein's *Philosophical Investigations*: The Diary of Private Sensations.[7] In the thought experiment, we conceive of a person attempting to define "I feel cold" by the internal process of ostensively picking out cold sensations, which specify the meaning and truth conditions of "I feel cold". So, whenever the subject writes "I feel cold" on a diary page, the sentence reports that she feels cold, and if she were to write "I do not feel cold", the sentence reports that she does not feel cold anymore. The question we will be pondering through the thought experiment is: Does the process of internally naming one's private sensations support the existence of privately accessible truth-conditions for "I feel cold", justifying a distinction between the truth and falsehood of the sentence?

Assume for the sake of thought experiment that one focusses on determining whether a particular sensation occurs. One associates this sensation with "I feel cold", and whenever

6. Jørgensen 1937, p. 290.
7. Wittgenstein 1953, §§244-271.

one experiences the sensation, one writes "I feel cold" on a diary page. Further, assume for a contradiction that "I feel cold" is a truth-apt report of a phenomenological sensation. For "I feel cold" to be a truth-apt report, it must be subject to the following safety condition: The subject must consistently designate the same phenomenological token sensations as the same type (i.e. as belonging to the set of "cold-sensations"). Without this consistency requirement, we would allow for the diffusion of the truth-conditions for "I feel cold", undermining its truth-aptness. Wittgenstein defines this safety condition in terms of a criterion of correctness: It must be impossible for the subject to make up new definitions as she goes, and any future ostensive definition must mirror the past. Candlish and Wrisley phrase the point thus: '[I]f this exercise is to be a genuine and successful ostensive definition, it must *establish* the connection between sign and sensation, and this connection must *persist*'.[8]

However, this safety requirement cannot be satisfied: Whatever is going to seem right to the subject is right,[9] but in order for her to be correct in her judgment, she must appeal to an independent justification beyond this seeming.[10] So, it does not make sense to assert that one is in a position to know whether "I feel cold" is true, because there is no fact by which one can judge truth and falsehood. It can, at most, constitute belief but is devoid of any stronger epistemic status.[11] Similarly, there is nothing for "I feel cold" to be false, because no fact could support such an assertion. This contradicts our assumption that "I feel cold" is a truth-apt report of a private sensation. Hence, "I feel cold" cannot be a truth-apt report of a private sensation.

Wittgenstein's thought experiment proves that we cannot conceive of "I feel cold" as a report of a private sensation, because no true nor false sensation report can be justified by an independent fact. Thus, in the case of Williamson's thought experiment, the subject's assertion that she feels cold in the morning (P1) and not at noon (P2) cannot be justified as true or false, because there is no fact of the matter open to her (nor our) view.

One may object that a criterion of correctness can be observed if the subject relies on her memory of the first sensation designated by "I feel cold". On Wittgenstein's account, the problem with this objection does not lie in the fallibility of memory.[12] Rather, the error consists in the presupposition that a proper designation has ever taken place. For the same reasons mentioned above, confirmation that future instances mirror the past instance is not open to our view. So, confirmation is impossible unless the sensation is public.

What are the implications of rejecting this account for Williamson's thought experiment? Note that we could have considered any other non-trivial sensation report e.g. "I feel pain". So, upon generalisation, Wittgenstein's argument rejects the widest possible scope of Williamson's argument (which Williamson is committed to): Sensation reports of the kind "I feel cold", "I feel pain" etc. are not truth-apt, because they do not report a private sensation fact allowing us (nor the subject) to distinguish between the truth and falsehood of the sentence.[13] So, under account 1, Williamson's commitment to the truth-aptness of sensation-reports generally cannot be maintained; nor can the limited commitment to the

8. Candlish and Wrisley 2019, §3.2.
9. Wittgenstein 1953, §258.
10. Wittgenstein 1953, §261.
11. Wittgenstein 1953, §260.
12. Kenny 1973, p. 194.
13. Williamson 2000, p. 106.

truth-aptness of coldness-reports be maintained.

In a sense, the failure to justify the truth-aptness of "I feel cold" under account 1 is good news for Williamson: If the meaning of "I feel cold" was luminous to the subject uttering the sentence, then this would be a direct way to motivate luminosity against Williamson's anti-luminosity argument. To see this, if Williamson were to assume that the subject is in a position to know the truth-conditions of her utterances of "I feel cold", then supposedly, the subject would also be in a position to know when the truth-conditions are satisfied, i.e. when she feels cold. Hence, in a sense, rejecting the private sensation account has only played in Williamson's favour. In a different sense, this result is bad news for Williamson, because he still has to justify the truth-aptness of "I feel cold". Since the private sensation account is perhaps the most intuitive account available, the task does not get easier. In the following, I will evaluate the other accounts available to Williamson, and ultimately argue that they all fail.

2.2 The physical sensation account

Secondly, we will consider whether "I feel cold" can be accounted for as a report of a physical entity on which the phenomenological properties supervene or are identical. If this is the case, then the private sensation could be inferred from some observable physical neural pattern, and there would thus be a fact justifying a distinction between true and false sensation-reports.

Let us consider this through the following thought experiment: Wittgenstein considers the idea of measuring brain activity and its correlation with our sensation reports.[14] We can expand the idea: if we imagine measuring the neural activity of the subject's brain using an advanced scanner, "I feel cold", when uttered by her, might then be found to be correlated with some pattern of neural activity, and taken to be a report of this observable, physical pattern. The idea is that if private sensations were to supervene on (or be identical with) some physical manifestation, then there *would* be a public measure as to whether the sensation occurs or not. We would thus avoid the problem of being unable to check that the subject is referring to the same thing when she utters "I feel cold". The truth-value of "I feel cold" would be independently justified by whether the neural pattern, when she utters "I feel cold", falls within the collectively decided spectrum for neural patterns correlated with utterances of the kind "I feel cold".

However, this account can be eliminated on fairly trivial grounds. From the set-up of Williamson's thought experiment, we are not conceiving of a person with wires attached to her brain and a bunch of neuroscientists determining whether it is true or false that a special pattern of neural activity is present when she utters "I feel cold". Rather, we are conceiving of a person tapping into her phenomenological sensations and expressing those sensations (as we do as humans!). Hence, the additional non-trivial claim that these phenomenological sensations are identical with or supervene on some patterns of neural activity is not assumed. Indeed, as pointed out by Srinivasan, there is no need to explore the option further: Rather than being a thought experiment about a logical or metaphysical possibility, it is precisely the nature of *our* luminosity that the thought experiment is meant to address: 'the question of whether *we* [emphasis added] are luminous is the question we

14. Wittgenstein 1953, §§270-271.

should care about'.[15] Therefore, whilst account 2 may account for the meaning of "I feel cold" in some possible world, it cannot be satisfied in the ordinary case, which Williamson considers.

2.3 The behavioural/environmental sensation account

Thirdly, one could contemplate whether "I feel cold" could be reduced to a disjunction of behavioral and environmental facts. More precisely, one could contemplate whether the following disjunction could reduce the meaning of "I feel cold": "I am displaying cold-behaviour or I am in a sufficiently cold environment or (...)". For instance, the utterance "I feel cold" would be regarded as true if the subject is step-jumping, or heating herself with her hands, or uttering it in the Antarctic and regarded as false if uttered in Southern Spain around summertime. Williamson's anti-luminosity argument makes none of these additional assumptions about the state of the environmental conditions, but it seems that adding them to his thought experiment would be an easy fix. So, suppose we can specify a disjunction of sufficient bodily expressions and environmental conditions associated with the expression "I feel cold", qualified by an inference to the best explanation. There are three pressing problems with this account:

First, we do not have to look very far to find counterexamples of feelings that are not consistently identifiable by one or more sufficient disjuncts. Consider sadness. Does sadness have a bodily expression? Not necessarily. So, Williamson cannot assume the link between private sensation and bodily expression precisely because Williamson's argument is not limited to cold-sensations *per se*; it assumes 'nothing specific about the condition of feeling cold'.[16] This broad scope must be preserved to preserve the force of the argument, but it cannot be preserved to include all sensations if one maintains that they are reducible to their external bodily manifestation or some environmental conditions.

Second, reducing our sensation expressions ("I feel cold") to bodily expressions or environmental conditions does not sufficiently reduce what we mean when we express our sensations through language. When uttering "I feel cold", we are not ascertaining an empirical proposition concerning external facts. Williamson's thought experiment captures this intuition: "tapping into" the sensation internally is necessary for us to determine how we feel. On the contrary, if we accept that "I feel cold" can be reduced to the subject's bodily expressions or the environmental conditions, then everyone else's knowledge of the subject's private sensation would be on par at best or exceed her own. The necessary component of "tapping into" the sensation would be redundant, pretentious, and hence, not necessary at all. Hence, Williamson cannot adopt this account because "I feel cold" is not an assertion of external, publicly accessible facts.

Finally, if Williamson reduces the meaning of "I feel cold" to external conditions, his anti-luminosity conclusion would be unsurprising. To see this, if our sensations are external and hence public, then it would be a trivial conclusion that the subject herself is not in a position to know her sensations from 'tuning into' them internally. However, Williamson's argument is not trivial: it is surprising at best and counterintuitive at worst. Hence, if adopting this account would trivialise Williamson's conclusion as a self-evident result, something has gone

15. Srinivasan 2015, p. 24.
16. Williamson 2000, p. 106.

wrong: The intuitions we *do* have would stand in a conflicting relation with the intuitions we *should* have. On this basis, I reject account 3: Accepting it would trivialise a result, which is far from trivial. So, the truth-aptness of "I feel cold" cannot be defended by its reducibility to external facts, and account 3 fails.

2.4 The expression account

Where are we now? Under account 1, there does not exist a private entity, which is open to the subject's view, such that she can assert that "I feel cold" is true with reference to such a fact. Similarly, under account 2, there does not exist a physical entity such that she (or a scientist) can justify her assertion that "I feel cold" is true without Williamson making additional non-trivial assumptions about the link between the phenomenal and the physical realms (which lies beyond the scope of Williamson's argument). Finally, under account 3, "I feel cold" is not reducible to any bodily manifestations or environmental facts. Since accounts 1, 2 and 3 are exhaustive of how "I feel cold" can be accounted for as a truth-apt sentence in Williamson's thought experiment, I conclude the section by stating the following: We have eliminated the option that "I feel cold" is a truth-apt sentence. Hence, it is a non-truth-apt expression. I now turn to motivating the status of "I feel cold" as a non-truth-apt expression under account 4.

According to the expression account, the verbal expression of coldness ("I feel cold") *is* cold-behavior and not a report distinct from the sensation. It is an identical substitute for shaking behavior, attempts to heat oneself with one's hands, step-jumping, or the accompanying "brrrrr!", which one utters in conjunction with this cold behavior. Hence, when the subject utters "I feel cold", she is not reporting anything distinct from the utterance itself. The utterance is a verbal expression equivalent to bodily step-jumping.[17] This may seem puzzling at first sight. It is not if we conceive of the following motivations:

The expression account is motivated by an empirical observation about how we as children acquire a vocabulary through the ones who already possess it.[18] According to Wittgenstein, we acquire a language through repeated public ostensive definition (our parents showing us the meaning of words, e.g. by pointing at the object of reference), which grants us epistemic access to the objects of reference of their expressions.[19] If words are transmitted in this way, then it seems implausible that the meaning of the sentence "I feel cold" could be transmitted at all if it referred to a private sensation (account 1) or a pattern of neural activity (account 2): Our children do not have access to our inner sensations or brain activity. In this light, it seems more plausible that the meaning of "I feel cold" (for it to be apt for transmission to other generations) is something they do have epistemic access into. They do have access to the publicly observable bodily behaviour of other humans, which allows other humans to transmit "I feel cold" through their behavioural expressions. Of course, this also motivates account 3. However, one more reason can be mentioned to argue that account 4 exceeds account 3:

Account 4 preserves our private access intuition, which account 3 does not: under account 3, everyone else's access to the subject's coldness is on par or exceeds her own, because her coldness can be confirmed by external facts only. However, under account 4, "I feel cold"

17. Fogelin 1976, p. 157.
18. Fogelin 1976, p. 165.
19. Wittgenstein 1953, §245

is identical to its bodily expression, which preserves our intuition that our sensations are privately accessible: nobody can engage in my cold behaviour, because their bodies are trivially not mine.[20] Of course, our private access intuition is only preserved in a trivial sense and not in the full-fledged sense of private access considered (and rejected) in account 1. Still, the preservation of this intuition together with the empirical facts about how we acquire language motivate account 4 as an adequate account of sentences of the kind "I feel cold".

What are the consequences for Williamson if we accept account 4? Rather than being a report, "I feel cold" is an expression identical to the physical manifestation of chills (or other cold-behaviour). It follows that such statements cannot be known, because they do not express a doubt.[21] To see this, for there to be a doubt, there must be room for mistake, but there is no room for mistake if someone utters "Brrr!". Is it true? False? Both questions are nonsense; no evidence would support its falsehood and no evidence would support its truth.[22] So, by virtue of being expressions and not statements under the expression account, "I feel cold" is not truth-apt.[23]

This conclusion has further implications: It shows that Williamson's fundamental project of subjecting the epistemological status of our emotion expressions to logical scrutiny is misguided: Under Wittgenstein's expression account, the predicates "I know" or "I am in a position to know" preceding "I feel cold" are misleading signals, because there is no knowledge to be gained about the fact of the matter.[24] "I feel cold" and similar nonsensical statements should thus be exempt from Williamson's logical analysis.[25] In the following, I will elaborate on these consequences for Williamson's argument.

2.5 Consequences for Williamson's argument

The argument unfolded in the previous section is this: I initiated the section by arguing that the truth-aptness of P1 and P2 follows trivially from Williamson's assumption that they are true (premise 1*). Further, I assumed that accounts 1, 2 and 3 are exhaustive of the most plausible ways in which P1 and P2 could be truth-apt (premise 2*). However, one by one, I have rejected accounts 1, 2 and 3 (premise 3*). My argument shows that Williamson's assumption that P1 and P2 are true is a contradiction.

What are the implications of having premises in his argument, which are contradictions? Recall that an argument is sound if and only if all of its premises are true and the argument is valid. However, the first conjunct ("all premises are true") is false: P1 and P2 are not true, because they are not truth-apt. So, Williamson's argument is unsound.

2.6 Final remarks

If the reader should take anything away from this dissertation, it is this: Sensation expressions are not truth-apt statements. Based on the argument presented in this paper, we can explain the counterintuitive character of Williamson's conclusion in terms of a failure

20. Wittgenstein 1953, §253.
21. Wittgenstein 1975, §156.
22. Wittgenstein 1975, §504.
23. Kenny 1973, p. 215.
24. Wittgenstein 1969, p. 55.
25. Carnap 1959, p. 61.

to realise that the subject matter of his argument precisely is expressions of phenomenological experiences, which are not apt for such manipulation in the first place. This takeaway gives rise to a characterisation of Williamson's thought experiment as one, which illuminates that logic should not be applied, ignorant to the study matter. In particular, our logical instruments should not be used to illuminate truths about the nature of our luminosity when those instruments fall short of accounting for the phenomenon in question. Recognising this prompts us to take a position of epistemic humility when evaluating each other's and our own internal testimony: not in terms of truth or falsehood, but rather, acknowledging their role as expressions of whose epistemic status we can only be silent.

3 Conclusion

According to a naïve comprehension of Williamson's anti-luminosity argument, one takes the existence of a valid contradiction at face-value, and proceeds by arguing that Williamson derives the wrong conclusion from his contradiction (a classic luminist reply). This dissertation has evaluated Williamson's argument from an alternative angle: In a Wittgensteinian fashion, I have argued that Williamson's argument goes wrong exactly as it trivially assumes the truth-aptness of sensation expressions P1 and P2. In section 2, I have argued that the triviality of this assumption is unfounded: it calls for a justification that we are in a position to know what our words refer to when we utter "I feel cold". However, I have rejected that we are ever in a position to know the meaning of our sensation-reports by rejecting the most plausible accounts of what "I feel cold" could mean. So, I conclude that the premises P1 and P2 in Williamson's argument are non-truth-apt and as a result, Williamson's argument is unsound.

Bibliography

Berker, S. (2008). *Luminosity Regained*. Available from: www.philosophersimprint.org/

S. Candlish and G. Wrisley, "Private Language". In: Zalta, N. (ed.). *The Stanford Encyclopedia of Philosophy*, (Fall 2019 Edition). Available from: https://plato.stanford.edu/archives/fall2019/entries/private-language/.

Carnap, R. (1959). "Elimination of Metaphysics Through Logical Analysis of Language". In: Ayer, A. J. (ed.). *Logical Positivism*. Westport, Conn.: Greenwood Press, pp. 60-81.

Cohen, S. (2010). Luminosity, Reliability, and the Sorites. *Philosophy and Phenomenological Research*, 81(3), pp. 718–730.

Fogelin, R. J. (1976). *Wittgenstein*. Honderich, T. (ed.). London: Routledge and Kegan Paul Ltd.

Geach, P. T. (1965). "Assertion". *The Philosophical Review*, 74(4), pp. 449–465.

Jørgensen, J. (1937). Imperatives and logic. *Erkenntnis*, 7(1), pp. 288–296.

Kenny, A. (1973). *Wittgenstein*. London: The Penguin Press.

Ramachandran, M. (2009). Anti-Luminosity: Four Unsuccessful Strategies. *Australasian Journal of Philosophy*, 87(4), pp. 659–673.

Srinivasan, A. (2015). Are We Luminous? *Philosophy and Phenomenological Research*, 90(2), pp. 294–319.

Williamson, T. (2000). *Knowledge and its Limits*. New York: Oxford University Press.

Wittgenstein, L. (1953). *Philosophical Investigations*. Oxford: Basil Blackwell

———. (1969). *The Blue and Brown Books*. 2nd edition. Oxford UK & Cambridge USA: Blackwell Publishing Ltd.

———. (1975). *On Certainty*. Anscombe G. E. M. & Wright G. H. (Eds.). Oxford: Blackwell Publishing Ltd.

Zardini, E. (2013). Luminosity and determinacy. *Philosophical Studies*, 165(3), pp. 765–786.

Appendix

It is worth spelling out exactly what is going on in Williamson's argument. We can formalise the premises as follows:

P1: One feels cold at the very beginning of the thought experiment. If we let s_t be the proposition that one feels cold at any time t ($0 \leq t \leq n$, where $n \in \mathbb{N}$), we can formalise P1 as: s_0.

P2: One feels hot at noon. For our purposes, this is equivalent to 'one does not feel cold at noon'. Hence, we can formalise P2 as: $\neg s_n$

LUM: In the thought experiment, we assume that whenever one feels cold (s_t), one is in a position to know that one feels cold ($k(s_t)$). That is, one's feeling cold and one's being in a position to know that one feels cold are exactly simultaneous (Williamson, 2000, p. 97). We can formalise this as: For all s_t, if s_t, then $k(s_t)$.

FCS: Williamson assumes that one focusses one's attention on determining whether one feels cold throughout the thought experiment. So, at any time t ($0 \leq t \leq n$, where $n \in \mathbb{N}$), one focusses on determining whether one feels cold $f(s_t)$. We formalise this as: For all t, $f(s_t)$.

KNW: Beyond assuming that we are in a position to know that we feel cold at any time t ($k(s_t)$), Williamson assumes that we focus our attention on determining whether we feel cold at any time t ($f(s_t)$). According to Williamson, these conditions are sufficient for one's knowledge of one's coldness ($k'(s_t)$). In the thought experiment, this relation between $k(s_t)$, $f(s_t)$ and $k'(s_t)$ serves to bridge LUM, FCS and MAR. We formalise it as follows: For all $k(s_t)$, for all $f(s_t)$, if $k(s_t)$ and $f(s_t)$, then $k'(s_t)$. For clarity, k denotes the state of *being in a position to know* oneself to be in mental state as a function of the state s_t obtaining, and k' denotes the state of *knowing* oneself to be in a mental state as a function of the state s_t obtaining.

MAR: In order for knowledge to be reliably based, it must satisfy the following safety condition according to Williamson: If at a time $t + 1$, one millisecond after t, one does not feel cold anymore, then the original belief that one felt cold at t is not reliably based (Williamson, 2000, p. 97). So, for one's belief that one feels cold at t to be reliably based and count as knowledge, one must feel cold one millisecond later. A margin-for-error premise (MAR) follows trivially from this safety condition, and we can formalise it as follows: For all $k'(s_t)$, if $k'(s_t)$, then s_{t+1}.

From P1, P2, MAR, KNW, FCS and LUM, Williamson derives a contradiction. Since everything follows from a contradiction, Williamson proves by contradiction that LUM is false. The only inference rules he needs to derive a contradiction are *modus ponens* (MP) and *universal elimination* (UE). We can spell the derivation out as follows:

1	$\forall s_t(s_t \rightarrow k(s_t))$	Premise
2	$\forall t(f(s_t)$	Premise
3	$\forall k(s_t)\forall f(s_t)(k(s_t) \wedge f(s_t) \rightarrow k'(s_t))$	Premise
4	$\forall k'(s_t)(k'(s_t) \rightarrow s_{t+1})$	Premise
5	s_0	Premise
6	$\neg s_n$	Premise
7	s_n	UE,MP, 1, 2, 3, 4, 5
8	$\perp$	ABS, 6, 7
9	$\neg\forall s_t(s_t \rightarrow k(s_t))$	EX, 8

The Blind Roboticist: The Knowledge Argument Against Qualia

Dan Ayres

King's College London

Abstract

Exploring the enigmatic topic of qualia within the broader discourse of consciousness, this essay critically engages with Frank Jackson's Knowledge Argument for qualia. Jackson's thought experiments, most notably his "Mary" thought experiment, serve as the foundation for the polemic strength of his rejection of Physicalism, by questioning whether all physical knowledge includes the phenomenal aspects of experience. This essay seeks to unearth the flaws of the "Mary" thought experiment and introduces an alternative thought experiment, the Blind Roboticist, to attempt to turn the Knowledge Argument against qualia. Through examining the intuitions that this author believes most will share regarding this alternative thought experiment, this essay argues that the supposed intuitive gaps in physicalist accounts of experience can be reconciled. By presenting the flaws in Jackson's formulation and presenting a compelling alternative, this essay both denies the polemic strength of Jackson's rejection of Physicalism and reframes the Knowledge Argument, positioning it as an argument against the existence of qualia, rather than in support of it.

Explaining human consciousness is one of the most important ongoing philosophical and scientific projects. Great strides have taken place in explaining the brain and its functions physically, but the venture is not yet quite complete. Qualia are the final, unsolved piece of the puzzle of human consciousness that seemingly refuse to be explained, as Daniel Dennett writes, they are the 'last ditch defense of the inwardness and elusiveness of our minds'.[1]

Exacting a definition of qualia is challenging; it is a somewhat esoteric term, the use of which varies slightly between papers. In the first modern usage of the term, Clarence Lewis writes that qualia are the 'recognizable qualitative characters' of the 'direct apprehension of the immediate'.[2] Brian Farrell and Thomas Nagel use the terms 'raw feels' and 'what it is like to be' respectively to mean, seemingly, the same thing.[3] While Dennett describes qualia as the 'ineffable, intrinsic [and] private' qualities of a subject's mental states that are 'immediately apprehensible in consciousness'.[4] As this essay concerns qualia as a bastion against Physicalism, I will use qualia to mean the special, first-hand, phenomenal aspects of conscious experience that cannot be explained by purely physical means.

[1] Dennett 1988, p. 48.
[2] Lewis 1929, pp. 120-121.
[3] Farrell 1950, p. 174, Nagel 1974, pp. 437-438.
[4] Dennett 1988, p. 47.

In this essay, I will discuss Frank Jackson's Knowledge Argument for qualia. In *Epiphenomenal Qualia* Jackson begins by acknowledging that "qualia freaks", those who believe in qualia, in general, fall into two categories: those who 'say that their rejection of Physicalism is an unargued intuition' and those who use an argument similar to the following: 'Nothing you could tell of a physical sort captures the smell of a rose, for instance. Therefore, Physicalism is false.'[5] Jackson admits that this argument is 'weak from a polemical point of view' as there are 'many who do not find the premise intuitively obvious' and it is clear that an unargued intuition has no polemical strength at all.[6] Jackson therefore sets himself the task of 'present[ing] an argument whose premises are obvious to all, or at least to as many as possible.'[7] The Knowledge Argument for qualia is his answer to this task. This essay attempts to show that the Knowledge Argument for qualia fails to accomplish this goal and that an alternative version of the Knowledge Argument may be used to argue against the existence of qualia.

I shall begin by analysing the thought experiments that Jackson uses in the Knowledge Argument for qualia, which I shall argue are flawed. I will then offer a revised thought experiment, the Blind Roboticist, that avoids those flaws. I will use this thought experiment to argue that, when conceived of correctly, the polemic strength of Jackson's Knowledge Argument for qualia is undermined, it is returned to the assumption of the original qualophile argument which is, to most, unintuitive. Furthermore, I intend for this essay to show that the Knowledge Argument can instead be used to support the claim that there is nothing special about experiences that one cannot know from the complete physical information about that experience, and thus, that there are no qualia.

1. The Knowledge Argument

In *Epiphenomenal Qualia,* Jackson uses two thought experiments, which I shall call the "Fred" and "Mary" thought experiments, to raise what he calls the Knowledge Argument for qualia.[8] This essay will focus on the "Mary" thought experiment as it is the more discussed, although, *mutatis mutandis,* the arguments presented here can be applied to the "Fred" thought experiment. The "Mary" thought experiment, briefly, goes as follows:[9]

> **Mary**: From birth, Mary is confined to a room where everything she experiences is black-and-white: herself, the room, her books, etc. She knows everything there is to know about the physical nature of the world, including: everything in *completed* science; the causal and relational facts consequent upon this; and all the physical facts about humans and our environment.[10] If Physicalism is true, then she knows all there is to know. But it seems obvious that Mary does not

[5] Jackson 1982, p. 127. For the rest of this essay, I shall refer to those who Jackson calls 'qualia freaks' by using Dennett's term of 'qualophiles', as I believe it to be the more respectful of the two terms.

[6] *Ibid.*, p. 128.

[7] *Ibid.*

[8] *Ibid.*, pp. 128-130.

[9] Reduced from Jackson 1986, p. 291.

[10] Some may raise the concern that it is impossible for Mary to obtain this knowledge within the limited framework of our current language. I agree with this argument and I believe it is a mistake on Jackson's part to not address this in the Mary thought experiment, a mistake that I believe stems from his misunderstanding of just how complex all physical knowledge is. I will address this concern in Section 3.

> know all there is to know, for when she is let out of the black-and-white room and sees colour for the first time she will learn something about the experience of seeing that colour. Hence, Physicalism is false.

Jackson stresses that 'the knowledge Mary lacked [before she left the room] is *knowledge about the experiences of others*, not about her own.'[11] It is obvious that when Mary is let out she will have new experiences, for example, the experience of seeing the colour red, and that these new experiences will cause things to change, such as causing a new brain state in Mary. Thus, Jackson tells us, 'physicalist and nonphysicalist alike' can agree that Mary learns something upon being released, namely facts about her experience of red, for before she was let out 'there were no such facts to know'.[12]

This, however, is a trick, for if the physicalist agrees that Mary learns *anything* upon leaving her confinement, then they inevitably fall prey to Jackson's later conclusion that Mary will realise that her understanding of the experiences of others who had already seen colour lacked that thing. This would mean that even with all physical knowledge there are still things that Mary does not know, which in turn disproves Physicalism.

Fortunately, rejecting the idea that Mary learns *anything* whilst having these new experiences and undergoing these changes is possible. One must consider the fact that *ex hypothesi* Mary knows *everything* there is to know about the physical nature of the world, including the causal facts consequent upon it. From this, one may argue that even though after Mary is released some physical things change, she knows in advance *exactly* when these changes will occur, *exactly* how these changes will occur, and *exactly* what the effects of these changes will be. Thus, the idea that, despite her vast knowledge, Mary can only guess how the experience of seeing red would affect her is a misunderstanding of just how incomprehensibly vast her knowledge actually is. Instead, Mary would know for certain *exactly* how the experience would affect her and she would be able to imagine it precisely and concretely. Therefore, when Mary first sees red, she will not be surprised by her experience and thus will not learn something new from it despite its novelty, for she already knows *everything* about the experience of seeing red including how it will affect her.

Dennett suggests an alternative ending to the "Mary" thought experiment that, for the most part, illustrates this conclusion: Mary's captors attempt to trick her by making her first colour experience one of a blue banana, presumably expecting her to mislabel it as yellow, she is not fooled, with her justification for figuring out their ploy being 'I already knew exactly what *thoughts* I would have [when I experienced blue]'.[13]

This conclusion is correct, but I argue that Mary knows more than just what her thoughts will be; she also, importantly, knows exactly how a blue stimulus would cause her to have blue thoughts and exactly what these blue thoughts would be on a neurophysiological level. Paul Churchland arguably shares this intuition, stating that Mary 'may well be able to imagine being in the relevant cortical state [...] in advance of receiving the external stimuli that would actually produce it',[14] but I argue that his use of "imagine" is misleading, as it implies that she is speculating and thus that there is a possibility that she might be wrong

[11] Jackson 1986, p. 292.
[12] *Ibid.*
[13] Dennett 1991, p. 400.
[14] Churchland 1985, p. 26.

or miss something out. With her level of knowledge, I argue that Mary would be able to spontaneously *know* what relevant cortical state any given external stimuli would produce in her.

To expand upon this argument, I shall first define two brain states:

> B_1 is the brain state that one enters when one considers all of the changes in one's brain that would occur if one were to see object A.
>
> B_2 is the brain state that one enters when one has the visual experience of seeing object A.

Suppose that after Mary has learnt all physical knowledge, she is given the description of object A and is prompted to consider what would occur if she were to see it. She then enters brain state B_1. When Mary considers what the changes in her brain would be, her privileged position of having all physical knowledge enables her to know exactly what would occur in her brain were she to see object A. What one must consider is whether by entering B_1 Mary knows how she would be affected by brain state B_2. Suppose that she does not. What could she be missing? Certainly, she could not miss anything that could serve any function, nor could she miss anything that was generated by the physical stimuli that caused the experience. Anything that qualified for either of those two distinctions would be known by Mary in B_1. Therefore, for B_1 to be missing anything about B_2 it must be outside of the causal chain of having the visual experience of seeing object A. But B_2 just is having the visual experience of seeing object A. It seems unintuitive to suggest that for Mary to know how B_2 would affect her when she is in B_1, she must know more than what is in the causal chain of B_2. Furthermore, positing that there must be something that is outside of the causal chain of B_2 and has no effect on Mary but nonetheless exists and is an important part of her experience has the same polemic strength as the argument that Jackson wishes to improve. Thus, far from being a new argument with premises obvious to all, Jackon's argument relies on the same unintuitive assumption of the argument he wished to improve.

Howard Robinson objects to the above, arguing that '[Mary's] perfection as a scientist will not tell her [...] about current particulars', instead it will only provide her with 'general scientific knowledge' and thus she will not be able to 'recognize the stimulus just by looking at it'.[15] According to Robinson, all Mary can know is how one would be disposed to speak about and react to a certain colour stimulus.[16] To capture how the stimulus would affect her in its entirety, Mary must be informed of the nature of the physical stimulus, at which point she can 'work out (not know spontaneously) that this is the sort of stimulus that prompts "that's blue"', but, importantly, 'it will be a revelation that [that physical stimulus] *looked like that*'.[17]

I have two objections to Robinson's argument. Firstly, I believe Robinson misunderstands the extent of the general scientific knowledge that Mary possesses. Rather than only knowing how an individual would be disposed to talk about and react to physical stimulus, Mary knows exactly what effect their looking at that physical stimulus would have on their nervous

[15] Robinson 1993, p. 175.
[16] *Ibid.*, p. 176.
[17] *Ibid.*

system down to the sub-atomic level.[18] Thus, even if it is granted that Mary is unable to determine everything that is going on as it happens, the descriptions that she is capable of writing down in advance are a lot more complicated than Robinson suggests. It is not absurd to suggest that her descriptions would be so detailed that they would account for even the slightest effect in her potential goings-on. These descriptions, combined with her general scientific knowledge, would allow her to discern exactly which physical stimulus she was experiencing just by looking at them.

Secondly, there is no physical reason why Mary shouldn't be able to spontaneously know her current physical particulars. Given her knowledge of all physical information and acquaintance with visual stimuli of the black-and-white sort, how they affect her brain states and the experience of these brain states, it is not unreasonable to suggest that Mary might be able to equate the phenomenal content of her experiences to its physical counterparts. Even if it is granted that Mary cannot perfectly and spontaneously equate her experiences to the goings-on of her current particulars by herself, one can imagine that Mary could, given the required materials, create a purely physical spontaneous-current-physical-particulars-detector that would allow her to do so.

If what determines whether Mary knows what the "qualia" she is undergoing are is whether she has access to some purely physical, external device that monitors and informs her of her purely physical body states, then these "qualia" are not the qualia I defined in the introduction. Robinson's argument is fundamentally flawed as it makes a category mistake: its basis, a reliance on limitations that qualophiles and physicalists alike would agree are purely physical, is of the wrong sort to result in a conclusion that shows that there are qualia, which are supposed to be entirely non-physical. Therefore, I believe that Robinson's objection does not disprove the conclusion that in the "Mary" thought experiment Mary learns nothing when she first sees colour.

However, while this conclusion allows physicalists to answer the "Mary" thought experiment without having to concede non-physical qualia, it is not, per se, an argument against qualia. The qualophile can happily accept that, given the initial premises, Mary would not learn anything new when she first saw colour. Rather than this result being conclusive against qualia themselves, the qualophile can argue that this apparent ability to explain away qualia comes from a flaw in the original formulation of the thought experiment by Jackson; it gives the physicalist too much.

This can be seen both in Dennett's explanation of how Mary knows exactly what the colour in front of her is and in my refutation of Robinson's objection. The very first item Dennett lists in 'the things [Mary] obviously knows in advance' is that 'she knows black and white and shades of gray'.[19] But this 'knows' is not only intended to mean that she

[18] It may be argued that allowing Mary to know specifics of other people's brains is beyond the general scientific knowledge that Robinson limits Mary to. But this is also required to know how any individual would be disposed to speak and react to any given colour stimulus. She must know the specifics about the individual's life to know how their past experiences with that colour have shaped their dispositions towards that colour, or rather how those specifics have affected the individual's brain and thus their dispositions towards that colour. So that argument seems to limit Mary's abilities to an even greater degree than Robinson suggests they should be limited. Fortunately, Mary, at the very least, knows the specifics about her own life and how they have affected her brain, so she will be able to determine the effect of the physical stimulus on her own nervous system, which is all this argument requires.

[19] Dennett 1991, p. 400.

knows the neurophysiological effects that each will have upon her. Instead, Dennett includes the fact that Mary already knows what it is like to experience black and white and shades of grey, for she has experienced them in her contained life.[20] Similarly, my disproof of Robinson's objection relied on Mary's previous knowledge of visual experience. But these visual experiences themselves include what a qualophile would label qualia.[21]

It is by no means an incontrovertible blow against qualia that Mary can derive visual qualia from all physical knowledge *in addition to* direct experience of other visual qualia.[22] To be able to argue against qualia, the physicalist must start from a position where the lack of *any* visual qualia is undeniable. Thus, the thought experiment must be reimagined.

2. The Blind Roboticist

Alan, through some means, be it genetic alteration, surgical intrusion, or otherwise, has his optic nerves deliberately and reversibly blocked by some experimenters before he is born. He is then taught, through a combination of auditory and tactile delivery, to be both a brilliant scientist and a history expert, learning all physical information, the entirety of anthropologic history, and the process of humanity's evolution from single-celled organisms. Included in this knowledge is all the physical information about what occurs when a normal-sighted person has visual experiences, including but not limited to everything about the physical structure of the brain and the nerves involved in the visual system, and the evolutionary pressures upon humanity and its ancestors and how visual experiences helped to overcome them.[23]

Alan uses this information to construct a robot which can examine and inform him of his visual surroundings in exquisite detail; specifying the luminance and wavelength of the light received from them, to the limits of what is detectable to the human eye, and shapes, including textures, to a microscopic level of detail, of any object that he instructs it to.[24]

One day, Alan is placed in a device so that his entire visual field, were he able to see, would be taken up by Vincent van Gogh's *The Starry Night*. He then instructs his robot to examine the painting and inform him about the physical stimuli that he would receive from it. Alan memorises this information and considers it all simultaneously, noting exactly how he would experience these physical stimuli if he were able to see. The experimenters then remove the blockages from his optic nerves, allowing him to see for the first time. He immediately has the visual experience of *The Starry Night*. Does Alan learn something new when this occurs?

[20] Dennett continues to list the other experiences that she knows, surface properties such as glossiness and matte, and the luminance boundaries that she experiences from her TV.

[21] Some might argue against the possibility of meaningfully labelling and retaining sensations as objects of knowledge (See Wittgenstein 2009, § 258.). I have addressed a related issue in section three, this could be taken as a preliminary suggestion on this problem, but more to follow.

[22] See Morreall's discussion of Hume's missing shade of blue for why such derivations are possible and unproblematic to qualophiles. (Morreall 1982).

[23] Some may object that the last element on this list is not physical information. This I concede, but respond that if the existence of qualia relies on the absence of knowledge of history external to oneself, then they are far from the qualia present in usual discourse.

[24] According to Edwin Land 1997, p. 128, every possible colour that a human can see can be uniquely specified by its position according to the three axial values of lightness; the robot provides Alan with these values.

This is the question to which the physicalist must answer no and the qualophile must answer yes, for if the entirety of visual experience can be explained by physical information alone without the need for any prior visual experiences, then non-physical qualia do not exist.

To motivate the physicalist intuition, I will now posit an alternative thought experiment. Imagine that the device that encompasses Alan's visual field allows for the insertion and removal of different pictures. Suppose that after the robot had described *The Starry Night* to Alan, the experimenters removed it from the device and replaced it with Raphael's *The School of Athens* before returning his sight to him. Would Alan be able to tell that his first visual experience did not correspond to the description he had been given by his robot? I think the answer to this, irrespective of whether one is a physicalist, is intuitively yes.

Regardless of the finer details, Alan would, at the very least, be able to tell that the areas of most intense lightness in the information he had been given did not align with those of his experience. It also seems obvious that Alan would be able to tell that the immediate differences in his experience of the colours of two proximal areas where a colour changes to an entirely different colour, for example from blue to red, would be too great to correspond to information that indicates that the colours should be almost identical, for example between two close shades of blue. To give a specific example, the information for the dark green trees that dominate the left half of *The Starry Night* would not correlate at all with the same area in *The School of Athens,* which at the bottom contains a myriad of different coloured robes and at the top has almost white marble walls. Thus Alan would be able to tell that that area in *The School of Athens* must have changed from when his robot examined it.

It seems intuitive that this should also be the case for any two areas taken from any two images where Alan receives the information for one and has a visual experience of the other regardless of how subtle the differences between them are, for why should the scale of the difference, given that it is perceptible, render Alan unable to notice it?

If one does not at first share this intuition, then one must remember the level of detail of the information that Alan receives, it is microscopic and to the limits of the human eye. If this is where one's doubt stems from, for example one does not believe that Alan would notice if a single μm^2 section of an image were changed by small degree, then one must question whether one would notice such a change if one were to see both images. If the answer to the second question is no, then it is unreasonable to expect Alan to be able to notice such a change. The intuition my position relies on is not that Alan could notice any change, but only that he would notice any change that someone who could see would notice.

One might have the intuition that Alan would not be able to tell if the original image was replaced with an identical image except for the fact that all the shades of blue were replaced with shades of red of corresponding intensity.[25] This intuition only comes from the way we typically talk about colour. Churchland suggests an alternative way to identify colours, as the 'various spiking frequencies in the nth layer of the occipital cortex (or whatever)'.[26] If

[25] Note that if one makes this argument then one still allows Alan to know vision monochromatically. But, from the "'Mary"' thought experiment, if Alan knows everything about monochromatic vision, then it is possible for him to derive everything about colour vision.

[26] Churchland discusses Mary, but his reasoning can equally be applied to Alan, with the caveat that he does not have any visual sensations of black, grey or white, but I do not believe that this is relevant to the point (Churchland 1985, pp. 25-26).

we consider colours in this way then it becomes clear that global colour changes would be as obvious to Alan as any other change, for they would cause different spiking effects in his brain to the ones the information told him would occur.

Another possible intuition is that regardless of the knowledge that Alan possesses, he simply could not know what visual experience would be like, as it consists of the sensing of something *out there* that, prior to the restoration of his sight, Alan simply has no access to. This intuition is incorrect, for it is a mere historical accident that the human visual apparatus works as it currently does. As Dennett writes, 'It is a mistake to think that first there were colors [. . .] it is rather that first there were various reflective properties of surfaces, reactive properties of photopigments, and so forth, and Mother Nature developed [colour vision] out of these raw materials.'[27] There are no colours *out there,* instead there are only raw physical facts and how our brain interprets them. This is similarly true of the entirety of visual experience, there is nothing visual *out there*; one can imagine an alien creature that uses a system that is entirely different to our eyes to detect the same raw physical facts and thus has a completely different experience for identical stimuli.[28] But *ex hypothesi* Alan already knows the properties of the raw materials and knows how the human visual system evolved relative to them. Thus, he must also know exactly what the human visual experience of any given physical stimuli would be like.

Thus, I believe the original intuition, that Alan would be able to tell if *The Starry Night* was swapped to *The School of Athens* after his robot examined it, both remains strong upon further inspection and extends to all possible changes in the picture regardless of how subtle they are, given that they would be perceptible to a normally sighted human.

But if it is agreed that Alan would be able to tell if there were any differences between the visual experience that he expected based on the information he received from his robot and what his first visual experience actually was, then *ipso facto* it must also be agreed that Alan learns nothing about visual experience when he has his first visual experience. For if Alan were to have learned something when he had his first visual experience, then there must have been something about the visual experience that he did not know before it. But if there was something about the visual experience that he did not know before he had it, then that thing could have been changed without him noticing that it was different to the visual experience he expected from the information he had been given by his robot. But we have already agreed that there is nothing about the visual experience that can be changed without Alan noticing.

So, provided that one agrees with the intuitive answer to the alternative experiment, one must conclude that the answer to the original Blind Roboticist thought experiment is that Alan does not learn anything when his sight is returned to him. But this conclusion results in the physical explainability of qualia. Thus, to maintain the elusive, non-physical nature of qualia, the qualophile must dispute the above intuition and argue that Alan would be unable to tell if the experimenters swapped the paintings.

[27] Dennett 1991, p. 378.

[28] Indeed, one's imagination need not stretch far. In *Other Minds* (pp. 119-123) Godfrey-Smith (2016), discusses octopi who have two 'visual' systems, one in their eyes and the other in their skin, which can detect and respond to light waves independently. How octopi experience 'seeing' with their skin is not yet understood, but one can imagine that it would be an experience of an entirely different nature to the visual experience we, and they, receive from eyesight.

3. Is It Possible To Obtain All Physical Knowledge?

Having put forth the physicalist argument, it is necessary to address a possible objection that some may have concerning both thought experiments: that it is impossible for an individual to learn all physical knowledge. It is clear that no current language is sufficient to express and explain all of this knowledge and some might doubt whether it is possible for there to be a language that is capable of doing so. I do not share this doubt, as I believe that it is arbitrary to draw a line of sophistication beyond which we cannot communicate in the physical world, and, even if one believes that there is more to the world than just the physical, the physical world is all that is relevant for all physical knowledge to be obtained. Additionally, some may argue that even if such a language were to exist it would be impossible for any individual to know that much information and be able to readily access it. While I agree that both individuals would have to possess exceptional intelligence and memories, beyond anyone who has actually existed, I stipulate that they remain within the realm of what is humanly possible, but I concede that this is disputable.

However, even if these concerns are granted, and it is thus impossible for any human to know all physical knowledge, this would apply to both versions of the Knowledge Argument, and thus render it mute on the topic of whether there is something about sensation that is ineffable to humans, therefore returning the argument for qualia to its initial qualophile assumption based form. But the arguments presented in Section 2 would still stand against the existence of metaphysically ineffable qualia, even if it remains true that some aspects of sensation are ineffable to humans. Therefore, the argument raised in Section 2 still presents an intuitive argument against the existence of metaphysical qualia, while the above concern questions whether the argument addresses only the metaphysical possibility for qualia themselves or includes whether there is something ineffable about human sensory experience.

4.Conclusion

The purpose of Jackson's Knowledge Argument is to provide polemic strength to the qualophile position by suggesting that it is intuitively the case that something is missing from the physicalist account of experience. Jackson writes that 'the polemical strength of the Knowledge Argument is that it is so hard to deny the central claim that one can have all the physical information without having all the information there is to have.'[29] In this essay, I discussed the flaws of the "Mary" thought experiment Jackson used to raise the intuitions that support this claim and suggested an alternative thought experiment, the Blind Roboticist, that instead shows that it is intuitively the case that nothing is missing from the physicalist account of experience. If the arguments raised here are agreed with then, rather than being problematic for the physicalist, Jackson's Knowledge Argument for qualia does not provide any polemical strength to the qualophile position, and, by using the Blind Roboticist thought experiment, an alternative Knowledge Argument can be used to support the conclusion that there is nothing about experience that cannot be explained physically, and hence that there are no qualia. Whether this conclusion extends to the rejection of the ineffability of sensations for humans or remains contained only to the rejection of the existence of metaphysically ineffable qualia remains in the eye of the reader. But, in either

[29] Jackson 1982, p. 130.

case, I believe this essay shows that the Knowledge Argument should be used against qualia rather than for them.

Bibliography

Churchland, P. M. (1985). "Reduction, Qualia, and the Direct Introspection of Brain States". *The Journal of Philosophy,* 82(1), pp. 8–28.

Dennett, D. C. (1979). "On the Absence of Phenomenology". In: Gustafson, D. F. & Tapscott, B. L. (eds.). *Body, Mind, and Method.* Synthese Library, vol 138. Dordrecht: Springer.

Dennett, D. C. (1988). "Quining Qualia". In: Marcel, A. J. & Bisiach, E. (eds.). *Consciousness in Contemporary Science.* Oxford: Oxford University Press.

———. (1991). *Consciousness Explained.* Boston: Little, Brown and Company.

Farrell, B. A. (1950). "Experience". *Mind*, 59(234), pp. 170–198.

Fogelin, R. J. (1984). "Hume and the Missing Shade of Blue". *Philosophy and Phenomenological Research*, 45(2), pp. 263–271.

Godfrey-Smith, p. (2016). *Other Minds: The Octopus, the Sea, and the Deep Origins of Consciousness.* New York: Farrar, Straus and Giroux.

Hume, D. (1739/2003). *A Treatise of Human Nature.* Project Gutenberg. Available from: https://archive.org/details/mindtheworldorde007547mbp.

Jackson, F. (1982). "Epiphenomenal Qualia". *The Philosophical Quarterly,* 32(127), pp. 127–136.

———. (1986). "What Mary Didn't Know". *The Journal of Philosophy*, 83(5), pp. 291–295.

Land, E. H. (1977). "The Retinex Theory of Color Vision". *Scientific American*, 237(6), pp. 108–129.

Lewis, C. I. (1929). *Mind and the World Order: Outline of a Theory of Knowledge.* Toronto: Charles Scribner's Sons. Available from: https://archive.org/details/mindtheworldordinee007547mbp.

Lycan, W. G. (1993). "Review". *The Philosophical Review,* 102(3), pp. 424–429.

Morreall, J. (1982). "Hume's Missing Shade of Blue". *Philosophy and Phenomenological Research*, 42(3), pp. 407–415.

Nagel, T. (1974). "What Is It Like to Be a Bat?". *The Philosophical Review*, 83(4), pp. 435–450.

Robinson, H. (1993). "Dennett on the Knowledge Argument". *Analysis*, 53(3), pp. 174–177.

Shoemaker, S. (1991). "Qualia and Consciousness". *Mind*, 100(4), pp. 507–524.

Wittgenstein, L. (2009). *Philosophical Investigations.* 4th edition. Anscombe, G.E.M.; Hacker, P. M.S.; Schulte, J.(ed. & trans.). Oxford: Wiley-Blackwell.

Exploring the Boundaries of Word Identity through a Nominalist Ontology of Words

Sophie Keay Rodger

University of Durham

Abstract

This paper explores the underdeveloped topic of the ontology of words within metaphysics, specifically focusing on the neglected area of handwriting. The argument presented here is based on David Kaplan's critique of John Hawthorne and Ernie Lepore's account of the nature of words. According to Kaplan, Hawthorne and Lepore's view contradicts our intuitive understanding, as it implies that illegible handwriting does not satisfy the necessary criteria for word identity, despite the common perception that illegible handwriting should still be considered to meet such standards. To address this issue, it is proposed that we adopt a nominalist trope bundle theory of words (NTBT). NTBT provides a comprehensive framework that can accommodate all forms of handwriting. Despite Kaplan's theory having the capacity to incorporate illegible handwriting in his ontology, it faces larger shortcomings and therefore should not be accepted. NTBT not only accounts for illegible handwriting as a legitimate instance of a word, in line with our intuitive understanding, but it also overcomes other criticisms directed at both type realism (proposed by Hawthorne and Lepore) and the stage-continuant view (proposed by Kaplan). By employing NTBT, it is possible to conclude that the example of illegible handwriting used in this paper is an instance of a word. This is because each written word can be identified by a bundle of different properties, and therefore can be associated with various sets or collections based on these properties. The example has a membership to its relevant set of semantic property tropes despite the fact it fails to be recognised by standard conventions of calligraphic property tropes generally associated with the word. By recognising illegible handwriting as a valid manifestation of words, we can acknowledge the nuanced nature of linguistic expression and broaden our understanding of how language operates in different contexts. This conclusion not only satisfies our linguistic intuitions but also has the foundation of a very inclusive theory of words that reflects how we use language and avoids discrimination based on ability.

The ontology of words has received limited attention in metaphysics. This is surprising, given that words are commonly (although not exclusively) relied upon as a means to communicate ideas. Within this under-explored area is the neglected topic of handwriting. Typically, when investigating the ontology of words, researchers discuss different methods of producing words, such as, digital pixels on a screen, spoken utterances, sign language, braille or ink markings. Researchers have extensively studied the ways in which individuals

understand each other through spoken utterances, including the impact of interruptions to standard word performances such as slips of the tongue, stutters, accents, and spoonerisms. Contrary to spoken word, in written instances of a word, it has been largely assumed that the arrangements composing the word (or nonword) are legible.[1] Handwriting, in this way, is worthy of philosophical analysis because it can act as a unique vehicle to investigate the boundaries of word identification in a way that other forms cannot. This is because the calligraphic and orthographic properties of handwritten words are unique to the individual writer and can vary based on factors such as mood, health, environment, and other influences; resulting in a fluctuating calligraphic representation from one instance to another.[2] This creates a dynamic performance of word creation which can be manipulated in a way that typography cannot.[3]

While there has been limited exploration of this topic in existing literature, one exception can be found in Kaplan's *Words on Words* (2011), in which the idea that the non-inclusion of "bad handwriting" is a reason to doubt an ontology of words is briefly introduced. However, the lack of comprehensive discussion warrants the need for further analysis. Illegible handwriting as a failed performance of a word contradicts our intuition as it suggests that hard to read handwriting does not meet specific orthographic conditions for word identity. Kaplan's critique against Hawthorne and Lepore's account is in favour of his stage-continuant (S.C.) model, but significant criticisms can be brought against Kaplan's model which NTBT avoids. Therefore it is suggested that NTBT offers a more comprehensive ontology of words that better accounts for word creation and usage and is therefore favoured.[4]

Section I discusses shape-theoretic nominalism (STN), showing why relying on the orthographic properties of a word is not advisable in word identification. This section will introduce and overcome a significant criticism that is traditionally brought against nominalist accounts of words, namely as an approach that cannot provide a sufficient answer to incorrect spelling or mispronunciations of words. Section II introduces Miller's NTBT to demonstrate that a nominalist approach to word ontology can be very flexible and accommodate an extensive range of linguistic phenomena. This section will consider some ways in which Miller's version is superior to STN, type-realist (T.R.) accounts, and the S.C.model.

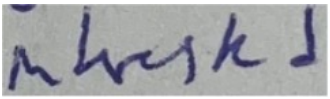
[1] Legible in this sense can be thought of as the physical appearance of each of the letters contained within the word being clear and distinguishable from each other, not variations in spelling which have been discussed in the literature and which will be discussed throughout the paper.

[2] "Calligraphic" and "orthographic" represent two distinct aspects of linguistic expression. "Calligraphic" refers specifically to the visual and stylistic dimensions of handwriting, emphasising the physical form and visual presentation of handwritten words. On the other hand, "orthographic" is a broader term encompassing the shape properties of any written form. As such, calligraphic practices can be viewed as a subset of the broader category of orthographic properties.

[3] Here, we can caveat that while a different typeface or stylistic variation of a font can alter the appearance of a word, particularly in terms of its shape, the individual letters that compose the word typically remain distinguishable. The impact of typeface and font variations on word shape will be discussed in Part IV. However, despite these potential changes, the core orthographic properties of the word are generally preserved, ensuring that the word remains legible and recognisable.

[4] A comprehensive account of the stage-continuant model and type-realist accounts are beyond the scope of this paper. However, they are a less viable approach to reaching a comprehensive ontology of words because the stage-continuant model is too reliant on intention and word-stages result in the necessary creation of a word (Hawthorne and Lepore, 2011; Miller, M.S) and type realist models have an unnecessary ontological commitment to abstract entities.

Namely that it i) can accommodate difficult word instances, ii) avoids invoking substances and abstract entities, and iii) does not overstate intention. Sections III and IV discuss a particular instance of illegible handwriting to investigate when some φ constitutes an instance of the same word under NTBT. The following insert is the example that will be referenced throughout this paper:

 [5]

To avoid presupposing the existence of a specific word and to demonstrate that the identification of this particular word is not important for the purposes of this discussion, this paper will refer to the above shape as ink marking (/IM/). This paper seeks to persuade that despite its illegibility, /IM/ is an instance of a word. This conclusion is reinforced by the tolerance principle,[6] drawing similarities to mispronunciations and spelling errors, and word-token collections. Building upon this analysis, Section V concludes by defending the superiority of NTBT over the other aforementioned ontologies. This conclusion suggests that NTBT provides a more compelling ontology of words, capable of accommodating complex linguistic breakdowns and the changing dynamic and social nature of language, whilst avoiding unnecessary ontological commitments to abstract entities.

1. Shape-theoretic Nominalism (STN)

Nominalism is a philosophical view that argues that only particular or concrete instances of things exist, and general or universal concepts are merely useful ways to categorise and describe these instances.[7] In the context of words, nominalism holds that their meaning is not inherent or universal but depends on specific usage, context, and conventions.[8] One drawback of adopting a nominalist stance on words is its apparent inability to accommodate misspellings. Wetzel (2000) discusses this particular concern regarding Goodman's perspective on the nature of words. Goodman suggests that:

> "'Paris' consists of
>
> five letters is short
>
> for any
>
> Every 'Paris'-inscription consists of five letter-inscriptions"[9]

At first glance, the characterisation of the word inscription "Paris" may seem appropriate. Paris, or P-a-r-i-s is a five letter-inscription. This view is attributed to Goodman and Quine (1947) and Bloomfield (1933, 1936), and is identified by Miller (2020) as a specific form

[5] This (potential) word token has been taken from handwritten feedback that I received during my studies at Durham. The creator of this shape has been informed about their contribution to the paper and has no objections to its inclusion

[6] Defined and analysed in Section III.

[7] Panaccio 2022.

[8] Miller 2020.

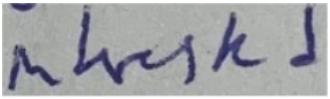
[9] Goodman 1972, p. 262

of nominalism known as Shape-theoretic Nominalism (STN). STN asserts that the visual shape of written marks or the distinctive sound characteristics of spoken words are what make them recognisable as individual word units. This idea may seem intuitive, as naturally we associate the word 'apple' with its spelling, a-p-p-l-e, and its pronunciation,/ˈæp.əl/. However, upon closer examination, this seemingly straightforward criterion exposes its limitations in accounting for the broadly acknowledged variations found in everyday language. Wetzel persuasively demonstrates the flexibility of language by citing "Paris" variations like "Parrys" and "Pareiss," supporting the recognition of spelling and pronunciation differences. Moreover, Wetzel writes, "extraordinary" consists of six syllables, but not every utterance of it does. The word is variously pronounced with 'six, five, four, three or even two syllables'.[10] In recent years there have been studies investigating the generational relationship between anatomy, geography, and the production of vocalised speech. Geographic location in this way impacts the way we move our mouth to produce sounds.[11] For example, "Schedule" is pronounced /ˈʃɛdjuːl/ in the UK and /ˈskɛdʒuːl/ in the US and yet are the same word. Miller (2019, 2020a, 2022) argues that the existence of diverse accents challenges the adoption of an ontology that strictly defines words. Such examples demonstrate the challenge faced by STN in differentiating words.

The distinction between homonyms, as discussed by Wade Munroe (2022) and Miller (2020a) demonstrates STN's limitations. We frequently encounter words that sound alike but carry different meanings. These instances are widely recognised and used within common language practices. For example, hear (listening to something) and here (in, at, or to this place or position) have distinct meanings but sound the same. Another way in which specifying words by their phonetic properties can fail is through looking at false friends. The English word for "magazine" and the French word "magasin". The words sound the same but do not mean the same.[12] Moreover, Cappelen (1999), Wetzel, (2000), Alward (2005) and Miller (2019, 2020a, 2022), persuasively argue that different spellings of the same word appear in various geographically specified lexicons. For example, color and colour. Despite linguistic variations, the provided examples are typically identified as instances of the same words within conventional linguistic frameworks. Nevertheless, STN faces challenges in adequately accounting for this phenomena. It is evident that a particular word can be conveyed through many different linguistic mediums and may possess a range of acceptable spellings. Such words may simultaneously possess multiple connotations or interpretations and have more than one acceptable pronunciation. Therefore adherence to specific spelling or pronunciation patterns is neither necessary nor sufficient for word identity. A theory of words must recognise that language evolves over time and that variations can coexist without undermining the integrity of a word. STN cannot provide a satisfactory account for these common language puzzles and cannot be accepted as a sufficient theory of words.

Before proceeding to the subsequent section, it is worthwhile revisiting /IM/ through the perspective of STN to demonstrate its inflexibility.

[10] Wetzel 2000, p. 363.
[11] Everett 2013.
[12] Munroe, 2022.



/IM/ challenges a STN approach's ability to provide a precise ontology of words, as it struggles to accommodate cases like it within its framework. Miller persuasively argues that, even if we modify STN so that it can account for "non-standard" instances of words, the challenge of determining the basis for distinguishing between significant and insignificant variations of spellings or pronunciations in identification[13], and which versions (if any) to privilege, remains. Miller also highlights that despite this accommodation, there is still the limitation of being unable to make sense of cases where word tokens share the same spelling but are intuitively perceived as different words.[14] For instance, the importance of this issue and the need for a precise ontology is evidenced by "rock", "rock", and "rock" which have different semantic meanings but share the same orthographic and phonetic properties. To be clear:

(**1**) "Rock" refers to the solid mineral material;

(**2**) "Rock" refers to the genre of music;

(**3**) "Rock" refers to the illegal smokable substance.

According to STN these word-tokens are instances of the same word despite having vastly different meanings and applications. Miller is persuasive when he argues that STN is unsatisfactory due to our strong intuitions in such similar cases.[15] In this context, "rock", "rock", and "rock" are surely not the same word. It is important to note that STN's inability to account for /IM/ does not necessarily invalidate nominalism as an ontological framework. Rather, it is this specific form of nominalism that cannot accommodate /IM/ and (as a result of other linguistic cases) has been largely discredited in the existing literature. Considering this critique, Miller proposes an improved nominalist approach that addresses Wetzel's critique. The proposed approach can overcome the problem of illegible handwriting and avoids problems associated with T.R. theories and the S.C. model. As a result, it offers an inclusive and ontologically less demanding solution to the problem of whether /IM/ is an instance of a word token.

It is important to note here that the T.R. and S.C. advocate can provide an explanation for various linguistic phenomena, including differences in spelling, accents, and mispronunciation by invoking intentionalist accounts of stages of words (S.C.) or with reference to tokens and types (T.R.). A thorough analysis might also reveal the adaptability of S.C. and T.R. models to illegible handwriting. However, the S.C.model should not be accepted on the basis that it overemphasises intentions,[16] and as Hawthorne and Lepore (2011) and Bromberger

[13] Miller M.S.
[14] *Ibid.*
[15] *Ibid.*
[16] Discussed in Section III.A

(2020) note, the S.C.theory forces us to accept that words could not be otherwise.[17] Moreover, a T.R.model cannot serve as the foundation for word ontology due to its difficulties in accounting for word creations and the concept of word sameness, while also entailing an additional ontological commitment to abstract word-types and substances.[18] Hawthorne and Lepore (2011) acknowledge the uncertainty surrounding borderline cases in their type-realism analysis of words, proposing the concept of sloppy realism to reconcile language imprecision with an objective worldview.[19] However, perhaps this serves as a convenient justification to promote their particular viewpoint, while overlooking alternative accounts that are better equipped to handle the challenges posed by borderline cases. By embracing the NTBT , it is possible to explain these borderline cases and develop a more flexible and comprehensive account of words that avoids leaving word-tokens as unexplained or unexplainable.[20]

2. Nominalist Trope Bundle Theory (NTBT)

The previous section challenged the notion that a word's orthographic or phonetic property is a fixed or definitive aspect of a word. This challenge exposes the limitations of criteria solely based on these properties in constructing a comprehensive ontology of words. This is because we cannot suggest that the standard conventional "physical" properties of a word results in *ipso facto*, a token of that word. This fact holds great promise for the potential incorporation of /IM/ into our lexicon using a NTBT because /IM/'s orthographic dissimilarities to other word-tokens. This improves upon previous nominalist word conceptions.

Miller presents a NTBT of words that builds upon a bundle theory of objects, seeking to provide an ontology that avoids invoking any ontologically committed talk of substances or abstract entities and avoids the criticism that can be levelled at STN.[21] Unlike T.R.accounts which argue that words are abstract eternal types, or Kaplan's S.C.view that argues that words are "interpersonal stages along with some more mysterious intrapersonal stages",[22] Miller's view instead posits that word-tokens are "nothing but a bundle of properties".[23] In this perspective, words are not abstract entities; instead of types, they are collections of tokens that have certain suitably resembling properties, "most notably semantic, phonetic, orthographic, and grammatical".[24] This approach neatly accounts for both misspellings and different pronunciations because of its flexible framework and its sympathetic disposition towards the messy, socially dynamic features of linguistic occurrences.

Using NTBT, the identification of /IM/ as a word is not dependent on its categorisation as a token of a specified word-type, as the concept of type is approached differently compared to the T.R.model. In NTBT, "type-talk" can be employed to communicate ideas succinctly without ontological commitment.[25] This is because word identification revolves around word-

[17] For instance, disregarding the metaphorical and figurative usage of words
[18] Miller M.S.
[19] The term "sloppy" refers to the idea that language and its meanings are not always precise or well-defined, and that there may be borderline cases where the application of words is uncertain
[20] Admittedly a harder feat for Hawthorne and Lepore than Kaplan who is ambivalent towards abstract words.
[21] Miller M.S.
[22] Kaplan 1990, p.98
[23] Miller M.S., p. 49
[24] Miller 2021, p. 5737
[25] Miller 2020b

tokens sharing partly composing property tropes, which can be exact or non-exact, depending on the context and degree of resemblance between the tropes in question. In this way, there are no talks of type in the same way as the T.R.model within the context of NTBT.[26] This is undoubtedly a strength of this approach. Using Ockham's razor as a guiding principle, it is possible to see how positing fewer entities is a strength here. It is not abundantly clear why we need to elicit talks of types. What does the application of this look like? In the case of the homonyms "rock," "rock" and "rock" each word can be identified as having members being partly composed of numerically distinct exactly resembling phonetic (P) and orthographic (O) property tropes, and numerically distinct non-exactly resembling semantic (S) property tropes (although this is not exhaustive):

(I) Rock {S*, O1, P1}

(II) Rock {S**, O2, P2}

(III) Rock {S***, O3, P3}

This application can be substituted for the aforementioned examples in Section II, of spelling mistakes, false friends and pronunciations. In this way, the same-ness of a word is a matter of exactly (or non-exactly) resembling collections of orthographic property tropes, semantic property tropes and phonetic property tropes, respectively. The question of whether Paris, Parrys and Pareis are the same words is dependent on which exactly (or non-exactly) resembling property tropes are deemed contextually relevant.[27] A spelling test (Miller, M.S) or a handwritten postcard will have different answers.

Munroe questions whether Miller's ontology of words is sufficient for false friends. Munroe (2021), using the example of the French word "auge", the Spanish word "auge" and the German word "auge" highlights his scepticism of a complete ontology of words using NTBT. Is the digital inscription "auge" the French, Spanish or German word-token? According to Munroe it is not immediately discernible to us. Therefore, a NTBT is insufficient in giving us an indication of which semantic property trope bundle "auge" belongs to. In Munroe's words, 'there must be some further fact'[28] to instantiate the fact that auge 'possesses certain language specific properties and, thus bears the membership or parthood relation to the set or collection/plurality of Spanish "auge" tokens'.[29]

Under Munroe's analysis, NTBT commits us to the claim that the word-token "auge" has a semantic property membership to exactly one of the languages. On Munroe's (2022) interpretation, we must commit to the brute fact that "auge" is a word-token whose members are exactly resembling one of the language's semantic properties and not the others. The decision for Munroe appears arbitrary. However, it is not clear that there needs to be some further fact. The French word "auge," the Spanish word "auge" and the German word "auge" have distinct pronunciations, meanings, grammatical behaviours, uses and applications. In French it is pronounced, /oʒ/ and is a feminine noun. In Spanish, /ˈauxe/ and it is a masculine noun. In German, /ˈaʊ̯gə/ and it is a neuter noun. The only commonality among

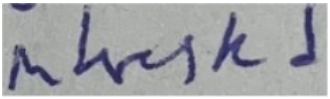
[26] *Ibid.*
[27] Miller M.S.
[28] Munroe 2022, p. 10
[29] *Ibid.*, p. 7

them lies in their orthographic properties, although even this observation is not entirely accurate. For example, in French "auge" is preceded by "une" or (because "auge" begins with a vowel) the definite article "l" partnered with an apostrophe. In Spanish it is "el" or "un." In each, the gender of the nouns impacts the rest of the sentence formation. Does the word-token "el auge" bear a membership relation to the collection of Spanish, French, or German warrants a shorter investigation. It is an unusual case for there to be a need to identify the specificities of language out with the parameters of language context. Thus, it is perhaps asking too much for an ontology of words to be able to distinguish between these words in isolation, notwithstanding the contextual relevance of the person who performed the word-token.

However, to meet Munroe halfway, there is no reason why NTBT cannot accommodate this phenomenon. Visually it is both possible and not possible to discern which language "auge" belongs to depending on the context and what is determined to be important. "Auge" "Auge" and "Auge" each belong to corresponding language specific collections. Collections that have overlapping members is not a problem for NTBT. Munroe's linguistic problem appears to us similarly to Schrödinger's cat. The word token "auge" is simultaneously a member and non-member of the French, Spanish and German lexicon. In this way, the "further fact" already exists, and we can determine the semantic property in the presence of important social contextual properties. In this sense, the word "auge" exists in a state of linguistic superposition. It simultaneously occupies multiple overlapping metaphysical categories. All instances of the different words shaped like "auge" are members of the collection $\text{Auge}_{\text{ORTH}}$. The word tokens are both French and not French, German, and not German, and Spanish and not Spanish, until a relevant context (like its semantic property) can clarify its true linguistic belonging.

Collections of words have specific membership conditions. Each language collection has word tokens in that collection because they meet the (non-exact) criteria for belonging to that particular language. Therefore, $\text{Auge}_{\text{ORTH}}$, is part of a larger conjunctive collection of collections. The word tokens shaped "auge", are members of various overlapping collections because it can be found within the larger, distinct particular language collections, such as $\text{Auge}_{\text{FRENCH}}$, $\text{Auge}_{\text{GERMAN}}$, and $\text{Auge}_{\text{SPANISH}}$. These language-specific collections, in turn, contribute to the even larger collection known as $\text{Auge}_{\text{LANGUAGE}}$. The presence of overlapping word tokens in these collections is not a problem but rather a reflection of how language has evolved over time. It signifies the shared characteristics between different linguistic systems. The fact that "auge" exists as a member in multiple collections does not undermine the coherence of a NTBT or the understanding of language as a social phenomenon. Trope bundle collections show the complex social dynamic nature of language influenced by historical, cultural, and social factors. A theory that can neatly categorise this evolution of language in a way that shows how words can span multiple linguistic categories is a key strength of this approach.

3./IM/



From the preceding sections of this paper, it is clear that word identity is far from straightforward. Our intuition strongly suggests that a lecturer using a pen in hand as the mechanism to produce a word should result in the performance of that word. As earlier sections have highlighted, relying solely on orthographic shape to determine a word token is an insufficient approach to word ontology. Consequently, the identification of /IM/ as a word does not depend on satisfying any regimented orthographic property conditions. Considering the context in which /IM/ was encountered, it is reasonable to assume that if it is indeed a word, it belongs to the English language's lexical inventory.[30] However, this assumption is not necessarily the case. It is possible that /IM/ is a new word or written in a different language. The objective of this paper is not to decipher the specific word but rather to establish the viability of recognising it as, i) an acceptable performance of a word within this context and ii) having exact (or non-exact) resembling properties to another word already existing in our language lexicon despite its orthographic illegibility. This section explores the relation between ambiguity, intention, and tolerance in the context of the NTBT and /IM/.

3.1 Intention

There is a strong intuition that (presumably) whatever /IM/ is, it was intended to be an instance of a word because of the social context in which it was presented. Miller does not explicitly include intention in their list of properties (in fact they are sceptical). However, the property list is not meant to be exhaustive, and Miller remains open to its inclusion. Therefore, one could argue that / IM/, is a member of its relevant word_{INT} if intention is deemed to be an appropriate property. Kaplan (1990/2011) proposes that two tokens are tokens of the same type if the speaker intends them to be so. Kaplan's ontology includes a token/type formalisation of words but to qualify as a nominalist perspective we might say:

> two word-tokens, *x*, and *y*, are instances of the same token collection *w*, iff their members are partly composed of (exactly or non-exactly) resembling intentional properties.

However, we should remain sceptical regarding any emphasis on the inclusion of intentional properties. As Cappelen (1999) argues we cannot reliably assume the intention of the mark maker. In the context of academic written feedback, inferring the intention behind word usage may be less challenging than interpreting other types of markings. Although the intent to communicate is generally implied in this context, the occurrence of instances like /IM/ still raises the possibility of it being a mistake or an unintentional marking. It

[30] This assumes the person who made /IM/ did not expect me to understand their communication in a different language.

could be the result of a pen's ink ball getting stuck, leading the writer to scribble on the paper to create friction and resolve the issue. Alternatively, someone else might have made the mark, without any intention of conveying a message. Despite our linguistic intuition derived from the norms of feedback commentary, it is impossible to achieve complete certainty about a person's intention without further communication. Thus, it becomes more valuable to explore the concept of tolerance to decipher the meaning and significance of /IM/.

3.2 Tolerance

In word identification and criteria, the concept of tolerance is widely discussed. The criteria was originally introduced by Hawthorne and Lepore[31] using a type/token distinction:

> " **Tolerance**: Performance p is of a word w only if it meets relevant performance standard."

Miller, further develops this idea to satisfy a nominalist framework:[32]

> **"Tolerance***: Performance p is a member of a collection of word-tokens, w, only if p meets relevant local performance standards on being a member of w."

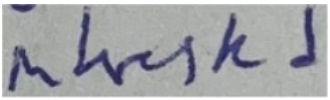
Tolerance*, offers greater flexibility compared to Tolerance because it can include illegible handwriting. The legibility of handwriting depends on a range of factors that can differ person to person and individually. Clear handwriting can be affected by many factors, making it difficult to adhere to a strict standard. These factors include tiredness, sore hands, time constraints, or distractions. Kaplan criticises Hawthorne and Lepore's concept of 'tolerance' in word identification, as it relies on the idea of being 'locally intelligible'.[33] The problem with local intelligibility is twofold. Firstly, if an individual travels to a different community and their articulation of a word does not meet the local standards, Hawthorne and Lepore suggest that they have not effectively used the word in their performance.[34] Despite the individual's ability to communicate using the same word in another community without any issues. In NTBT, mistakes in language use, such as mispronunciations and spelling errors, can be seen as words not having exactly resembling phonetic (mispronunciations) and orthographic (spelling errors) property tropes but their resembling semantic property trope remain intact. Therefore, context dependent word usage can be caveated with reference to specific properties of that spoken or written word.

According to Miller (M.S), mistakes occur when there is a failure to express or recognise the membership of a relevant collection instance. In the case of /IM/, it represents a combination of the writer's inability to express a word 'perfectly' and the reader's failure to recognise the expression. For instance, the Shetland Islands accent, with its Nordic stress patterns, might make English sound unfamiliar to those unfamiliar with the dialect. Can a person from Canterbury claim that a Shetlander is not using words to communicate? A person's lack of understanding should not equate to non-performance of a word. Instead, we might introduce that "spoken with a Shetland accent" is a contextually relevant property.

[31] Hawthorne and Lepore 2011, p. 18
[32] Miller M.S., p. 71
[33] Kaplan 2011, p. 519
[34] *Ibid.*

Similarly, in the written domain, 'pneumonoultramicroscopicsilicovolcanoconiosis' remains a word even if a reader fails to understand it due to its complexity or unfamiliarity. It is simply a string of letters coming together to form a word the reader might not have seen before. Furthermore, Hawthorne and Lepore's Tolerance principle raises concerns about potential ableism in word construction.[35] For instance, individuals with speech impediments may pronounce words differently. It would be unjust to claim that they are not using words based on such differences. NTBT avoids this.

4.Revisting /IM/



How far can we manipulate a word before it stops being a member of relevant-word-token$_{\text{ALL}}$? When reading the dissertation notes, I misidentified the relevant collection of /IM/. In this way, I have failed to recognise a word instance that is a member of the orthographically and semantically determined collection /IM/$_{\text{ORTH}}$ and /IM/$_{\text{SEM}}$ due to the unique features of this particular handwriting. This metaphysical organisation could also be taken further to suggest that /IM/ is a member of the collection /IM/$_{\text{HAND}}$, because it is partly composed of a handwriting trope. This allows us to explain why it looks different to other instances of the same word, whilst still being the same (numerically distinct) word. This collection can be broad depending on our requirements. We might posit a collection whose members are partly composed of either the property of 'handwritten with a time constraint' or 'handwritten while distracted.' These collections are important for this investigation but are not necessary in everyday language. Additionally, the reader's comprehension or lack thereof does not guarantee the success or failure of the written word. The writer and some readers might identify the collection to which /IM/ belongs. A reader's inability to understand does not invalidate the words performance.

NTBT can accommodate /IM/ by referencing its semantic property trope or through analysis of its individual letters. To Hawthorne and Lepore, /IM/ cannot satisfy local performance standards of a word. Similarly, to a grunt in a place of a word, /IM/ (for Hawthorne and Lepore) is too far removed from what it was originally supposed to represent. This is not a fair characterisation of / IM/. To illustrate this, it is suggested that it might be easier to comprehend /IM/ as individual strings of letters. Here, we can begin to reveal how NTBT accommodates for a variety of linguistic phenomena. Similarly to words, letters in NTBT can be understood as collections of letter tokens that have certain suitably resembling properties.[36] In line with our everyday usage of letters, "a","a", "a", "a", and "a" are instances of the letter "A" even though they do not have exactly resembling orthographic properties. Although /IM/ is not immediately accessible to the reader (although it might be) it does not mean that it is a non- performance of a string of letters. If the writer was to re-write this word in capital letters so that each letter becomes more legible let's call this /IM*/, we might then say that /IM/ is partly composed of semantic trope property s1 and /IM*/

[35] *Ibid.*
[36] Miller, M.S.

is partly composed of semantic trope property s2, where s1 and s2 have exactly resembling semantic property tropes but non-exact orthographic property tropes (o* and o**). Looking at phonetics, Miller (2020a) argues that the Great Vowel Shift, which occurred during the 1400s to 1700s and resulted in pronunciation changes that made certain words challenging to comprehend, does not necessarily signify a transformation in the underlying language or the identity of words undergoing this phonetic shift. Similarly, in orthographic structure, individualised interpretations of handwritten letters do not constitute different letter tokens.

5. Conclusion

By discussing Kaplan's critique of Hawthorne and Lepore's inability to accommodate illegible handwriting as instances of words, it has been revealed that nominalism can provide a solution to this problem. An STN account of handwriting suggests that /IM/ fails to meet certain conditions of word identity, challenging our intuitions. However, it is suggested that if we adopt NTBT, we can circumvent this problem. /IM/ is a valid instance of a word, aligning with our intuitions, and effectively addressing criticisms directed towards T.R., the S.C. model and STN. Within the perspective of a NTBT, although the example of handwriting examined in this paper may not qualify as a member of its relevant- $\text{word}_{\text{ORTH}}$ (or $\text{letter}_{\text{ORTH}}$), it belongs to the relevant collection of semantic property tropes associated with the word and furthermore relevant-$\text{word}_{\text{HAND}}$. Thus, falling within the ambit of its relevant- word.[37] This conclusion aligns with our linguistic intuitions, offering a more viable perspective on word identification. Although Miller does not address handwriting, it has been extrapolated through similar linguistic phenomena [38] that a NTBT can neatly and persuasively accommodate illegible handwriting. The implications of studying illegible handwriting extend beyond mere theoretical exercise. Such investigation can contribute to expanding our comprehension of word identification and by embracing a broader understanding of what constitutes a word, invaluable insights into language dynamics and the complexities of communication can be revealed.

[37] It has also been discussed that /IM/ could be a member of its relevant wordINT. However, due to the difficulties addressed with knowing someone's intention, it has not been included as an important member set (although this might change in the future)

[38] Misspelling, mispronunciation, accents etc.

Bibliography

Alward, P. (2005). "Between the lines of age: Reflections on the metaphysics of words". *Pacific Philosophical Quarterly*, 86(2), pp. 172–187.

Bromberger, S. (2020). "The Ontology of Words". *Compass*, 14(7), pp. 1–9.

Cappelen, H. (1999). "The shape of the word: The semantics of quotation and the nature of language". In: Goldstein, S. & Martin, R. (Eds.). *The philosophies of Quine and Wittgenstein*. Cambridge, MA: Harvard University Press, pp. 192–213.

Hawthorne, John & Lepore, Ernie (2011)."On Words". *Journal of Philosophy*. 108(9), pp. 447-485.

Kaplan, D. (2011). "Words on words". *The Journal of Philosophy*, 108(9), pp. 504–529.

Miller, J. T. M. (2019). "A bundle theory of words". *Synthese*. 198(6), pp. 5731–5748.

——. (2020a) "On the individuation of words". *Inquiry*. 63(8), pp. 875–884.

——. (2020b). "The ontology of words: Realism, nominalism, and eliminativism". *Philosophy Compass*, 15(7), pp. 1–13.

——. 2021. "Words, Species, and Kinds". *Metaphysics*, 4(1), pp. 18–31.

——. (2022). "Sameness of word". *European Journal of Analytic Philosophy*, 18(2), A2–A26.

——. (M.S.). "Words and Other Linguistic Entities". [Unpublished manuscript].

Panaccio, C. (2022). "Nominalism". In: Panaccio, C. *Ockham's Nominalism: A Philosophical Introduction*. New York: Oxford University press, pp. 27–45.

Munroe, W. (2022). "What it takes to make a word (token)". *Synthese*, 200(4), pp. 1–30.

Wetzel, L. (2000). "The trouble with nominalism". *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 98(3), pp. 361–370.

Apophansis

Contents of Vol. 1 No. 1, June 2024

History of Philosophy

Aquinas on Angelic Difference
– Brandon Smith 2

Pleonexia and Political Decline in Plato's Republic
– Arthur Kleinman 12

Moral and Political Philosophy

On the Slavery Objection to Political Authority
-Jonathan Anderson 24

Simone Weil and the Ethics of Attention: An Investigation on Compassion, Love and Justice
-Anna Maria Robles Fumarola 36

Philosophy of Mind and Epistemology

Justification as an Evaluative Concept: A New Argument for Externalism
– Guy Dinwiddy Smith 44

On the Logical Character of Emotive Language:Evaluating Williamson's Antiluminosity Argument
– Mette Andersen 54

The Blind Roboticist: The Knowledge Argument Against Qualia
– Dan Ayres 68

Ontology

Exploring the Boundaries of Word Identity through a Nominalist Ontology of Words
– Sophie Keay Rodger 80

Contact Details
https://apophansis.co.uk
apophansis.excc@gmail.com