Aporia

XXIII

Aporia

Undergraduate Journal of the St Andrews Philosophy Society

VOLUME XXIII

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

Letter from the the Editor

Dear Reader,

With the passing of another academic year in St. Andrews, it is my pleasure to introduce the latest edition of Aporia.

Additionally, this year I have had the great fortune of working with the Editor-in-Chief of the University of St. Andrews' very first feminist edition of Aporia, Mia-Belle Tierney.

I would like to thank the wonderful team of editors I had this year, my professors, and Lewis Muirhead for being a constant source of sanity and comic relief when it is most needed. Finally, this edition would not have come together so nicely without the help of Stepan Los and Mohit Agarwal.

I would also like to thank my loving parents, my grandparents, and Nigel Mika. As a friend and mentor, he has never led me astray. I am incredibly grateful for his kindness, support, and for gifting me his copy of Lemmon's Beginning Logic.

All the best,

Olivia Isabelle Griffin

Acknowledgements

Olivia Griffin
Editor-in-Chief

Stepan Los
Deputy Editor

Mohit Agarwal
Editor
Design Editor

Avery Cohen
Editor

Kirsty Graham
Editor

Christina Landys Herre
Editor

James Shearer
Editor

Arthur Zhang
Editor

BoYoung Zhou
Editor

Imogen Bowyer
Cover Art

CONTENTS

Kant on Space and Objectivity

David Zheng

Abstract. Gareth Evans attributes to Kant the following thesis: "space is a necessary condition for objective experience". However, Kant does not seem to argue directly for this thesis anywhere in the *Critique of Pure Reason*. In this essay, limiting myself to the *Transcendental Aesthetic* and assuming that Evans's attribution is correct, I attempt to reconstruct Kant's implicit argument for the thesis, finding it as a corollary of the main conclusion of the *Transcendental Aesthetic*, that space is a form of experience.

In 'Things without the mind', Gareth Evans famously attributes the Spatiality Thesis to Kant: "space is a necessary condition for objective experience".[1] The truth of the thesis bears upon the connection between the objective world and the idea of space, and the attribution to Kant indelibly shapes how one interprets the opening arguments of *Critique of Pure Reason*.

However, as Evans implies, Kant does not straightforwardly defend this thesis: the difficulty consists in locating or reconstructing Kant's argument for it. Thus, my question is this: *how does Kant argue for the Spatiality Thesis?*

With reference to Evans, Quassim Cassam locates Kant's utterance of the Spatiality Thesis near the end of the Transcendental Aesthetic: "Through space alone is it possible for things [*Dinge*] to be outer objects [*äußere Gegenstände*] for us" (A29).[2] I will take for granted that this is indeed an expression of the thesis.

There are at least three ways in which Kant could have implicitly defended the thesis in the *Transcendental Aesthetic*: (i) via the First Metaphysical Argument (A23/B37) where he argues for a connection between space and "outer experience", understood as objective experience; (ii) again via the First Metaphysical Argument but based on another interpretation by which "outer experience" is intimately connected to the individuation of objects (and thus objectivity); (iii) as a corollary of the main conclusion of the *Transcendental Aesthetic* that space is a form of experience.

Quassim Cassam would endorse (i) as Kant's way of arguing for the thesis, and Henry Allison would endorse (ii). However, I argue against (i) and (ii) by giving an alternative interpretation of the First Argument and propose (iii) as Kant's implicit argument for the thesis that space is necessary for objective

[1] Gareth Evans, "Things without the Mind—A Commentary upon Chapter Twp of Strawson's Individuals," in *Collected Papers: Gareth Evans* (New York: Oxford University Press, 1985) 250.

[2] Quassim Cassam, "Space and Objective Experience," in *Thought, Reference, and Experience,* ed. Jose Luis Bermudez (Oxford: Oxford University Press, 2005), 278.

experience.

After some remarks on the meaning of "objectivity" for Kant in relation to the thesis, and on some possible locations of Kant's argument for it (§1), I will be arguing that Cassam's reconstruction of the Kantian argument for the thesis is not implicitly made by Kant (§2). Similarly, I will argue that Henry Allison's interpretation of one of the passages of the Metaphysical Exposition, which appears modifiable into an argument for the thesis, cannot in fact be (§3). Rather, Kant's argument for the thesis is best conceived of as a corollary of his central claim that space is a form of experience; thus, his argument for it is best located where he explains that central claim (§4).

§1 Preliminary Remarks

(1) The thesis concerns objects of experience which are mind-independent.[3] It's useful here to distinguish between two senses of mind-independence: something is mind-independent if and only if it's independent of the existence of minds *or* independent of mental representation.[4] At the very least, then, the objects of experience in the thesis are *representation-independent*. As we will see, Kant distinguishes between two senses of "object", only one of which is mind-independent and thus relevant to the thesis.[5]

For Kant, "objects of experience" may refer to things perceived through the categories and concepts of understanding, usually denoted by "*Objekt*".[6] According to Kant, such objects result from syntheses of intuition and the applications of concepts, giving us well-formed, quotidian items. Since these objects are mind-independent in neither sense, the thesis cannot be about them.

"Objects of experience" may also refer to what's experienced *simpliciter*, untouched by understanding, usually denoted by "*Gegenstände*" (see A29 above). These "objects of intuition" (so I will call them) are distinct from objects of understanding, for "the manifold for intuition must already be given *prior to* the synthesis of understanding and *independently* from it" (emphasis mine; B145). Since Kant uses "*Gegenstände*" when stating the thesis, and such objects can be representation-independent (as per B145), it's likely that the thesis, as Kant states it, concerns objects of intuition.

(2) The aim of this essay is to explore *how* Kant argues for the thesis. A second, related question is what Kant means by it. Understood broadly, the thesis posits a conceptual link between space and objective experience, but

[3]Cassam, "Space and Objective Experience," 259.

[4]Pete Mandik, "Objectivity Without Space," *Electronic Journal of Analytical Philosophy,* vol. 6 (1998) para. 23.

[5]A third sense of 'object' for Kant is the things in themselves, but these cannot straightforwardly constitute objective experience.

[6]Dustin McWhether, *The Problem of Critical Ontology: Bhaskar Contra Kant* (Houndmills, Basingstoke: Palgrave Macmillan, 2013), 18.

the meanings of "space" and "objective experience" and the details of this connection need filling in. Since the thesis can be either trivial or captivating depending on how it's understood, and assuming that Kant intends for the thesis to be non-trivial, I will assess each interpretation (§2-4) according to the significance attributed to the thesis by each interpretation.

(3) Given the assumption that Kant states the Spatiality Thesis at A29, it's also reasonable to assume that he would argue for it somewhere before A29. I will consider two candidate locations: the paragraphs directly before A29 and the argument at A23/B37 (hereafter the "First Argument"). The First Argument is as follows:

> For in order for certain sensations to be related to *something outside me* (i.e., to something in another place in space from that in which I find myself), thus in order for me to represent them as outside <and next to> one another, thus not merely as different but as in different places, the representation of space must already be their ground. Thus, the representation of space cannot be obtained from the relations of outer appearance through experience, but *this outer experience* is itself first possible only through this representation. (A23/B37; emphasis mine)

The First Argument stands out because it argues for a connection between space and "outer experience", possibly referring to objective experience. The paragraphs before A29, where Kant states the thesis, also qualify as a candidate, for one may reasonably expect the argument for a conclusion to be located immediately before it. The first two interpretations (§2-3) which I consider will be based on the First Argument, whereas my own proposal will be based on the few paragraphs before A29 (§4).

§2 Cassam's Argument and "outer experience"

Cassam attributes to Kant the following argument for the Spatiality Thesis: "space is necessary for objective experience because we can only perceive objective particulars by perceiving their spatial properties" (hereafter "Cassam's Argument").[7] As above, the First Argument might be about objects in the mind-independent sense required by Cassam,[8] and it does make repeated mention of perceiving spatial properties (e.g., "represent them as outside"), something crucial to Cassam's Argument. So, the question becomes: *Can Kant be implicitly making Cassam's Argument within the First Argument?*

I argue that there appear to be two reasons to believe that Cassam's Argument occurs within the First Argument. The first is that the two Arguments

[7]Cassam, "Space and Objective Experience," 278.

[8]Ibid., 259.

have a structural affinity. The second is that they reach the same conclusion. However, as we will see, the two reasons taken together put us in a dilemma: the arguments cannot both have the stated structural affinity and reach the same conclusion.

One may lay out Cassam's Argument as follows:

> (2.1) "[O]ne cannot perceive [or in Kantian terminology, "represent"] an objective particular without perceiving it as having spatial properties";[9]
>
> (2.2) If one cannot perceive X without perceiving X as being Y, then the perception of Y is necessary for the perception of X (suppressed premise);
>
> (2.3) The perception of spatial properties is just the perception of space;[10]
>
> (2.4) Therefore, "[t]he perception of space is necessary for the perception of objective particulars".[11]

The first reason to think that Cassam's Argument is contained in the First Argument is that the two Arguments take the same form, i.e., the same logical relation holds between their main terms. For comparison, one may lay out the First Argument as follows:

> (2.1K) "[I]n order for certain sensations to be related to something outside me, [...] the representation of space must already be their ground";
>
> (2.2K) If one cannot represent X without representing X as being Y, then the representation of Y is necessary for representing X (suppressed premise);
>
> (2.3K) The "outer" (or the "outside") is just the objective;
>
> (2.4K) Therefore, "this outer experience is itself first possible only through this representation."

Cassam's argument takes the following form: in order for one to represent X , one must represent X as Y; *therefore*, the representation of X is only possible through the representation of Y. Kant's First Argument can be read as having the same form: in order for one to represent "something outside me" ("X"), "the representation of space" ("the representation of Y") must already be their ground; *therefore*, the representation of these things as outside me, i.e., "this

[9]Ibid., 278.
[10]Ibid., 265.
[11]Ibid., 278.

outer experience" ("the representation of X"), is "possible only through" the representation of space ("the representation of Y").

Their premises and conclusions take the same form, and the same logical inference exists between the premises and conclusions, i.e., the necessity of representation of Y for the representation of X is *due to* the fact that one cannot represent X without representing it as Y.

The second reason is that the two Arguments seem to reach the same conclusion, namely a variation of the Spatiality Thesis. In addition to the conclusion that the origin of space is non-empirical, the First Argument concludes with "this outer experience [*diese äußere Erfahrung*] is itself first possible only through this representation [of space]". Whether this second conclusion amounts to the thesis, however, depends on what "this outer experience" refers to.

Cassam simply takes it for granted that the "outer" means "the objective" in the sense of being mind-independent,[12] and, so, that the First Argument concludes with the thesis. However, intuitively, "this outer experience" seems to refer back to something mentioned one sentence earlier, i.e., to the sensations related to "something outside me [*etwas außer mich*]". This intuition is based on the repetition of the root "*aus*". If so, the "outer" means what's outside me, but not necessarily the objective; this would make the First Argument trivial, as we will see.

Daniel Warren helpfully clarifies the meaning of "outside" in the First Argument.[13] On Warren's reading, "*außer*" at the beginning of the Argument has a "spatial" rather than a "metaphysical" sense, designating things spatially other than where I am, because the language in the First Argument is exclusively spatial (more on this in §3).[14] So, if "outer" refers back to "outside", the First Argument *does not* conclude with the Spatiality Thesis, which concerns objective experience. Rather, Kant should be read as saying something along the lines of "(the representation of) space is necessary for the experience of spatial things".

However, giving "outer" a spatial sense reduces the First Argument to triviality. The second conclusion of the First Argument would, at best, be read as being about the connection between the representation of space and what's spatial. But if something is already defined as *F*, then undoubtedly it follows that this thing entails *its being F*, i.e., *F* is necessary for it. If we define "outer" as "spatial", then it follows trivially that space is necessary for the experience of outer objects. In other words, Kant would be making an analytic claim, for the concept *spatial* is already contained in the concept *outer*. To be more charitable to Kant, it seems that Cassam is right to take the "outer" to denote the objective.

[12] Ibid., 280.

[13] Daniel Warren, "Kant and the Apriority of Space," *The Philosophical Review* 107, no. 2 (1998): 179.

[14] Warren, "Kant and the Apriority of Space," 184-7.

In response to this, one might argue that the main conclusion of the argument is that the origin of space is non-empirical ("space cannot be obtained [...] through experience"), and the clause following this (what one might think is the thesis) amounts to no more than a clarification. Moreover, one might suspect that even if the second clause is counted as a distinct conclusion, it does not matter for Kant that it's an analytic claim. For, at this early stage, Kant is concerned with the "metaphysical exposition" of concepts, i.e., with *a priori* conceptual analysis. In a sense, it's acceptable that his conclusion is trivial, as he is merely explicating our concepts.

I believe, however, such a response is mistaken. Kant's main conclusion of the *Transcendental Aesthetic* is that space is the "form of all appearances of outer [*äußere*] sense" and that it alone makes possible "outer [*äußere*] intuitions" (A26/B42). If we insist that "outer" be given a merely spatial sense, one of his primary claims in the *Critique* would be rendered trivial: "*space* is the sole necessary condition for *what's spatial*". We cannot accept this.

In response, one might suggest that it's still possible to attribute some significance to Kant's claim above. Cassam, for instance, would probably suggest that "what's spatial" has a *metaphysical* sense and that "space" is a *subjective* form.[15] So, on Cassam's view, Kant would be establishing an epistemological link between the subjective and the metaphysical. However, I contend that even Kant's second conclusion in the First Argument could be given a strong and interesting reading.

To that end, one may first observe that Kant distinguishes between "outer" and "outside": "by means of outer sense [...] we represent to ourselves objects [*Gegenstände*] as outside, and all as in space" (A22/B37). That is, it is by virtue of the outer sense that we can represent things spatially other than where we are. So, to give "outer" a spatial reading would be to conflate the means and the ends, i.e., to conflate that by which we represent and the represented.

To the same end, we need to elucidate what Kant means in talking about "objects" of the outer sense at A22/B37. At this early stage, it is most plausible to read "objects" ("*Gegenstände*") as referring to those which are given immediately in intuition prior to the application of the categories. (These intuitive objects, as in §1, are objective in the sense that their existence is not dependent upon mental representation.) If it's *by means of* outer sense that we represent intuitive objects, it's natural to assume that "outer experience", i.e., the experience had via outer sense, is the experience of these objects of intuition.

On my interpretation, then, the second conclusion of the First Argument is this: the representation of space is necessary for the experience of intuitive objects. Note that, given my definition of "intuitive objects", this attributes to Kant a significant claim: that, for the things which are characterised as objective in the sense that their existence is not dependent upon mental representation,

[15] *cf.* Cassam, "Space and Objective Experience," 281.

if they *are represented* by the mind (i.e., if they become objects of outer experience), they are *necessarily represented as spatial*.

My version of the conclusion is preferable to ones that take "outer" to mean "spatial", for my interpretation attributes to Kant a non-trivial conclusion. If I am right to interpret Kant this way, he does reach a version of the Spatiality Thesis by the end of the First Argument. Combined with the first reason, viz., that the First Argument takes the same form as Cassam's argument, it appears that Cassam has correctly reconstructed Kant's Argument for the thesis.

However, these two reasons are, in fact, in tension: if the two arguments have the same form, the First Argument *does not* conclude with the Spatiality Thesis; but if the First Argument does, they *do not* share the same form.

If the two arguments take the same form, as per the first reason, the conclusion of the First Argument must be expressed like this: X is only possible through Y. Filling in "X" and "Y", we get: the representation of *these things outside me* (which has a spatial sense) is only possible through the representation of space. This is because "X", when the argument takes this form, refers back to the things spatially other than where I am. However, the second reason *denies* that the First Argument concludes with a claim about spatial things.

If, however, the First Argument concludes with the claim that space is necessary for experience of intuitive objects (as per the second reason), it does not conclude with the claim that space is necessary experience of spatial things (denoted by "something outside me"). But the latter claim would be the conclusion if the two Arguments take the same form.

The two reasons put us in a dilemma. Neither reason is individually sufficient to ground the containment of Cassam's Argument in the First Argument. Therefore, I conclude that Cassam has failed to reconstruct Kant's argument for the Spatiality Thesis.

§3 Allison's interpretation and "outside me"

Assuming that my argument above, that the First Argument does conclude with a Spatiality Thesis, is correct, it's reasonable to examine other interpretations of the First Argument, with hopes to reconstruct Kant's argument for the thesis, before looking elsewhere. One influential interpretation is offered by Allison who maintains that the First Argument concerns the necessity of space in grounding the *individuation* of objects.[16]

Allison himself does not intend for his interpretation to concern the thesis. But it appears that if we supply a suppressed premise and accept the claim which Allison attributes to Kant, the result is an elegant argument for the thesis, contained in the First Argument. However, as I will argue, we cannot interpret Kant as arguing for the thesis in this way, for we cannot attribute Allison's claim to Kant.

[16]Henry E. Allison, *Kant's Transcendental Idealism: An Interpretation and Defense* (New Haven: Yale University Press, 1983), 82-87.

Taking the First Argument as about the connection between objective individuation and space, Allison finds in it the claim that space is necessary for us to tell apart qualitatively identical but numerically distinct objects: space is necessary for us to represent things "not merely as different" (i.e., qualitatively different) "but as in different places" (i.e., numerically different).[17]

With Allison's claim as a premise, we may reconstruct the First Argument as follows:

> (3.1) The representation of space is necessary for numerical difference;
>
> (3.2) Numerical difference is necessary for objectivity (suppressed premise);[18]
>
> (3.3) Therefore, the representation of space is necessary for objectivity.

We self-consciously supply (3.2), knowing that, when it's combined with (3.1), we get the conclusion, a variation of the thesis, which we can attribute to Kant. That is, we may supply (3.2), a claim *not found* in the First Argument but one that holds true for Kant, *only if* we can attribute (3.1) to Kant.[19] But we cannot attribute (3.1) to Kant.

This is because Allison has misinterpreted the phrase "not merely as different but as in different places" to be about the individuation of objects. As we have seen, Warren argues that "outside" should be given a spatial but not an "ontological sense".[20] That is, *pace* Allison, he argues that "outside" just means "spatially other than" but not "non-identical to". If so, Kant does not claim (3.1). In support of this, Warren gives several convincing reasons; I will mention two.

First, the language in the First Argument is distinctively spatial.[21] "Outside", "another place in space", "next to", and "different places" are all spatial, not ontological. Against Warren, one might object: these descriptions *do entail* ontological differences. In response, we merely need to note that, although "spatial difference" and "numerical difference" may be co-extensional, what Kant *means* in the First Argument (i.e., his intension revealed by the use of language) has strictly to do with spatial difference.

Second, Warren acutely observes that the very purpose of Kant putting the explanatory bracket after "outside me" is to make explicit the distinction

---

[17] Allison, *Kant's Transcendental Idealism: An Interpretation and Defense,* 82*ff.*

[18] For Allison, (3.2) would be understood epistemologically (10*f.*), e.g., having an *idea* of objectivity.

[19] Warren (1998) would probably question the truth of B2: for a minimal kind of objectivity, it's enough that I distinguish between qualitatively different things (*cf.* 188). However, Kant likely holds B2 true, for objects of intuition only possess extrinsic or relational features (A285/B341). The only way to tell these objects apart is by their non-qualitative features.

[20] Warren, "Kant and the Apriority of Space," 184.

[21] Ibid., 185.

between "outside" and "outer" which appears later in the First Argument.[22] "Outer" may have an ontological sense, but not "outside". Warren's observation resonates with mine in §2 that Kant intends to distinguish between two cognates of "*aus*" at A22/B37.

So, we cannot attribute to Kant (3.1) nor, therefore, (3.3). That is, Allison's interpretation of the First Argument is not modifiable into an argument for the Spatiality Thesis. There are other interpretations of the First Argument to consider.[23] But due to the space limitations, I will now attempt to locate Kant's argument elsewhere.

§4 A Corollary of Transcendental Idealism

We have seen that an argument for the thesis which Cassam attributes to Kant cannot be found in the First Argument, and that, on Allison's interpretation, Kant cannot be read as arguing for the thesis with the First Argument either. The other likely location for his argument is in the passages directly before A29, where Kant states the thesis: "Through space alone is it possible for things to be outer objects for us" (A29).

Cassam suggests (but dismisses, as we will see) a connection between Kant's Spatiality Thesis and the main conclusion of the *Transcendental Aesthetic*,[24] which is also a doctrine of "Transcendental Idealism": "Space is nothing other than merely the form of all appearances of outer sense, i.e., the subjective condition of sensibility, under which alone outer intuition is possible for us" (A26/B42).

Taking seriously Cassam's suggestion, and using the doctrine above as the premise, we may incorporate Kant's clarifications of the doctrine in A27/B43 and interpret Kant as arguing for the thesis in either of the following two ways:

(4.1) Space is an *a priori* form of outer intuition (A26/B42);

(4.2) A "form" is a "necessary condition" for representing "appearances" (A27/B43);

Either	*Or:*
(4.3a) Appearances are "objects of sensibility" [*Gegenstände der Sinnlichkeit*] (A27/B43);	(4.3b) "Outer intuition" comprises objects independent of mental representation (see §1);
(4.4a) Therefore, space is a necessary condition for representing objects of sensibility (A29).	(4.4b) Therefore, space is the necessary condition for representing objects of intuitions (A29).

[22] Ibid., n.10.

[23] E.g., Falkenstein (1995, 161*ff.*) and Warren himself. They argue that, since we represent things as located before we can represent spatial relations, the origin of space is non-empirical (the first conclusion). Their interpretations appear to have no direct bearing on the Thesis.

[24] Cassam, "Space and Objective Experience," 236.

Kant, then, seems to give us two routes to reach the thesis. Via either route, a doctrine of Transcendental Idealism (4.1) *analytically* implies the Spatiality Thesis as the conclusion. This is because the other premises (except (4.3b)) are Kant's definitional elaborations of the doctrine.

I will assess my proposal by highlighting its merits and responding to two objections. The main merit is that it receives the best textual support thus far. The entirety of the argument, on my reconstruction, is traced back to claims Kant explicitly makes. The second merit is the close textual proximity between the conclusion and the premises. The whole argument is contained in section (b) of "Conclusions from the above concepts" (A26/B42 to A29/B45). The interpretation thus meets the expectation that an argument directly precedes the conclusion.

A tempting objection is that (4.4a) is not an instance of the thesis. This is because "objects of sensibility", being something close to sensations, are dependent on mental representation. But the thesis, at the very least, is a claim about mind-independent objects (see §1).

In response, one may compare (4.4a) with (4.4b), which is more evidently a version of the thesis. Specifically, we may ask whether "objects of sensibility" and the representation of "objects of intuition" are something similar. They are: "representing objects of intuition" describes how the *representation-independent* objects are put under a subjective form, which is roughly the meaning of "objects of sensibility".

Considering this, I concede that the phrase "objects of sensibility" indeed does not denote representation-independent objects. But I maintain that "objects" in "objects of sensibility" denote objects, just as "objects of intuition" do. That is, I am distinguishing between "objects" and "appearances"; the latter are representation-dependent. So, (4.4a) in effect says that space is necessary for representation-independent objects to *become sensible*.[25] If so, (4.4a) is still an instance of the Spatiality Thesis, for it makes a claim about representation-independent objects.

Another objection, one that Cassam is likely to raise, is that my interpretation misrepresents the relation between the Spatiality Thesis and (4.1), that space is the form of sensibility. Cassam maintains that, since Kant's aim is to "explain in more basic terms" the correctness of the Spatiality Thesis, it cannot be a paraphrase (a "notational equivalent") of (4.1).[26] Rather, to ground the thesis, (4.1) must be more basic than and "grasped independently from" the thesis.[27] However, unlike the reconstructions in §2 and §3, my account effectively represents (4.1) and the thesis as equivalent.

Cassam seems right about the basicness of (4.1).[28] However, my account is

[25]Cf. the non-trivial conclusion I attributed to the First Argument (§2).

[26]Cassam, "Space and Objective Experience," 264.

[27]Ibid., 264.

[28]E.g., see Strawson 1997, 237: space being a form of sensibility is 'an ultimate *fact* [...] not

compatible with this. Without (4.2), (4.3a) and (4.3b), it is not clear at all that the thesis follows from (4.1). That is, the connection between *Transcendental Idealism* and the thesis is not explicit and certainly not "notationally equivalent". To see this, note that (4.1) can be explained independently of the thesis, but the reverse might not hold true. Moreover, this is not controverted by my suggestion that (4.1) analytically implies the thesis: X may analytically entail Y, where X is grasped independently of and is more basic than Y.

Moreover, my account is compatible with the thought motivating this objection. The thought is perhaps that, since objects considered prior to experience are non-spatial, the connection between the spatial and the objective must be explained by the non-objective, i.e., subjective sensibility.[29] So, the form of sensibility, being the condition making the connection between space and objectivity possible, must conceptually precede a thesis about this connection. The version of the thesis, on my interpretation of Kant, is that when the representation-independent objects become sensible, they are *thus* represented as spatial. This claim conceptually separates sensibility from objectivity, and objectivity from space, and does not put the connection between objectivity and space prior to sensibility.

Conclusion

I have offered a reading on how Kant argues for the connection between space and objective. Attempting to locate Kant's argument for the Spatiality Thesis in the First Argument, I first questioned whether it implicitly contains an argument Cassam attributes to Kant, then entertained and modified Allison's interpretation of it. I have, finally, argued that the argument for the thesis, construed as an analytic corollary of Transcendental Idealism, is best located in A27/B43. This interpretation receives the best textual support and can stand up to objections.

If my account is accurate, we can confidently attribute the thesis that "space is a necessary condition for objective experience" to Kant. Moreover, my contentions on how to understand sensibility, objectivity, and the "outer" in context of Kant's argument for the thesis suggest an unconventional way of reading the opening arguments of *Critique of Pure Reason*.

capable of further explanation'.

[29] *cf.* Cassam, "Space and Objective Experience," 263.

Bibliography

Allison, Henry E. *Kant's Transcendental Idealism: An Interpretation and Defense*. New Haven: Yale University Press, 1983.

Cassam, Quassim. "Space and Objective Experience." *Thought, Reference, and Experience*, edited by Jose Luis Bermudez, 258-89. Oxford: Oxford University Press, 2005.

Evans, Gareth. "Things without the Mind—A Commentary upon Chapter Two of Strawson's Individuals." In *Collected Papers: Gareth Evans*, 249–90. New York: Oxford University Press, 1985.

Falkenstein, Lorne. *Kant's Intuitionism: a Commentary on the Transcendental Aesthetic*. Toronto: University of Toronto Press, 1995.

Kant, Immanuel. *Critique of Pure Reason (The Cambridge Edition of the Works of Immanuel Kant)*. Edited by Paul Guyer and Allen Wood. Cambridge: Cambridge University Press, 1998.

Mandik, Pete. "Objectivity Without Space." *Electronic Journal of Analytic Philosophy,* vol. 6 (1990).

McWherter, Dustin. *The Problem of Critical Ontology: Bhaskar Contra Kant*. Basingstoke: Palgrave Macmillan, 2013.

Strawson, Peter Frederick. *Entity and Identity and Other Essays*. Oxford: Oxford University Press, 1997.

Warren, Daniel. "Kant and the Apriority of Space." *The Philosophical Review* 107, no. 2 (1998): 179.

Assert What You Know: The Problem with Bald-Faced Lies

Joanne Park

Abstract. In "Bald-Faced Lies! Lying Without the Intent to Deceive," Roy Sorensen argues that the existence of bald-faced lies, or lies that do not involve deceit, challenge conventional definitions of lying. As part of his argument, he claims that bald-faced lies are not in themselves bad, as most ethical condemnations of lying target deceit and not the lie itself. In this paper, I challenge this aspect of Sorensen's claim by using Timothy Williamson's distinction from "Knowing and Asserting" between conventional and constitutive rules. I argue that bald-faced lies *are* in themselves bad because they place too much emphasis on conventional rules of assertion at the expense of the constitutive rule of assertion: assert only what one knows. Ultimately, Sorensen is wrong to say that bald-faced lies are morally neutral, as bald-faced lies devalue the constitutive rule of assertion and therefore the practice of assertion itself, which plays a critical role in the sharing of knowledge.

Introduction

In his "Bald-Faced Lies! Lying Without the Intent to Deceive," Sorensen claims bald-faced lies challenge standard accounts of lying, which assume lying involves an intent to deceive. As part of his analysis, Sorensen assesses the moral badness of bald-faced lies as compared to lies involving deception. In his view, bald-faced lies are not in themselves bad because most ethical condemnations of lying either target deceit or badness that is "correlated with the lie." In this paper, I ask whether this assessment is correct. Contrary to Sorensen, I will argue that bald-faced lies are in themselves bad by invoking Timothy Williamson's distinction between conventional and constitutive rules. Ultimately, I identify the wrongness of bald-faced lies in how they value conventional rules of assertion over the constitutive rule of assertion. For this paper, I will stipulate that an assertion is a kind of utterance that directly expresses one's belief.[30]

I will begin by outlining Sorensen's argument for why bald-faced lies are merely correlated with wrongdoing, as well as some of my initial reservations with his approach. I will then introduce Williamson's distinction between conventional and constitutive rules as they pertain to assertion. Based on this

[30]Bernard Williams, "Truth, Assertion, and Belief," in *Truth and Truthfulness* (United Kingdom: Princeton University Press, 2002), 78.

distinction, I will propose my alternative account of the badness of bald-faced lying, applying this account to three examples of bald-faced lies that appear in Sorensen's paper. Finally, I will end by presenting several advantages of my account.

Bald-faced lies

Sorensen defines bald-faced lies as undisguised lies or lies that "do not fool anyone."[31] Bald-faced lies occur when the intended audience of the lie knows that the utterance is a lie, and the liar knows that the audience knows that the utterance is a lie. As such, the liar has no intention to deceive the audience. One commonly used example of bald-faced lying involves a student who has been accused of plagiarism by their dean. The dean knows that the student has plagiarized, and the student knows that the dean knows that she has plagiarized. But, because the dean has a reputation for not punishing students who deny plagiarizing, the student lies and says that she did not plagiarize. As a result, she goes unpunished.[32] In this example, the student has no intention to deceive the dean that she has not plagiarized; instead, she lies to avoid being punished.

Deceitful lies, on the other hand, are "disguised" as genuine assertions. For example, if I lie to my parents that I have gone to the library when I have gone to my friend's house, I intend for my parents to believe that I have, in fact, gone to the library. If I know that my parents can see my location, I no longer have any reason for telling this lie. Standard definitions of lying, based on such examples, include intentional deception as a necessary condition of lying. Sorensen's paper is primarily a revision of such accounts. Appealing to the intuition that bald-faced lies must be lies, Sorensen re-defines lies as false assertions that could be believed by someone who only had access to the assertion and not any other evidence: e.g., in the student's case, a random bystander who only heard the assertion "I did not plagiarize" could plausibly believe her.[33]

Sorensen's defense of bald-faced lies

The secondary question in Sorensen's paper is about the morality of bald-faced lying; specifically, he refutes the conventional intuition that bald-faced lies are more wrong than lies with an intention to deceive. He first points out how most moral condemnations of lying are derived from the deception involved. For instance, deontological explanations of why lying is bad is based

[31]Roy Sorensen, "Bald-Faced Lies! Lying Without the Intent to Deceive," *Pacific Philosophical Quarterly* 88 (2007): 252.

[32]Don Fallis, "What is Lying?," *The Journal of Philosophy* 106, no.1 (Jan. 2009): 43.

[33]Sorensen, "Bald-Faced Lies!," 255; Because this paper does not focus on Sorensen's argument, I only offer a rough sketch of Sorensen's proposed definition of lying.

on how lies betray trust, while utilitarian explanations rely on how deception has bad consequences—neither of which apply to bald-faced lies.[34] Therefore, Sorensen argues, intuitions about the wrongness of bald-faced lies are not based on any common moral theories.

Second, Sorensen addresses the apparent moral wrongness of situations involving bald-faced lies. He proposes that, in such cases, the bald-faced lies are symptomatic of or correlated with some other wrongdoing.[35] Under this proposal, the wrongness of the plagiarism case does not come from the student's bald-faced lie, but from the fact that she was attempting to avoid just punishment.[36] Sorensen justifies this by comparing the student's case to an alternative one, in which the dean's unofficial rule is that he will only punish students who confess. If a student stays silent in the second case, the silence would be as wrong as the bald-faced lie, as both equally correlate with the wrongness of avoiding just punishment.[37] Ultimately, Sorensen labels bald-faced lies as "morally neutral": while they might be correlated with morally wrong action, there are no features of bald-faced lies that render them intrinsically immoral.[38]

I find this approach unsatisfying for two reasons. First, there seems to be an intuitive difference between the student who brazenly lies about whether she has plagiarized and the student who chooses to stay silent, even if both are avoiding punishment. By committing to the obviously false assertion, the student seems to be misusing the practice of assertion and is doubling down on their (false) innocence. Furthermore, even if both students are in the wrong, there is certainly something worse about the dean's rule in the first case, which incentivizes students to avoid just punishment by making an obviously false assertion. But, by not distinguishing between the two cases, Sorensen's account does not allow us to condemn such norms.

Second, Sorensen's proposal does not allow us to condemn bald-faced lies that are not straightforwardly correlated with wrongdoing. To see why this is problematic, we can take up a third case: the school has a rule that students are not allowed to raise their voice, but a student will not be punished for doing so if she makes an assertion denying that she has raised her voice. A student who raises her voice to defend a bullied classmate might, to avoid punishment, utter a bald-faced lie about doing so. In this case, there is no clear correlation to wrongdoing; some may even argue that the student is avoiding *unjust* punishment. If Sorensen means to assign rightness or wrongness to bald-faced lies based on how they correlate with some other goal or consequence,

[34]Ibid., 252.
[35]Ibid., 260-262.
[36]Ibid., 261.
[37]Ibid., 261.
[38]Ibid., 263.

the student's bald-faced lie might even be morally *right*.[39] Yet there is still something undesirable about the student's approach of asserting that which is blatantly false—as opposed to acknowledging that she raised her voice to defend her classmate or contesting the rule.

Williamson's account of conventional and constitutive rules

Before I outline my account of the wrongness of bald-faced lies, I will offer some background on Williamson's distinction between conventional and constitutive rules. According to Williamson, conventional rules are contingent and arbitrary, meaning they can be replaced by other conventions or change over time.[40] For example, a conventional rule of soccer might be to shake the referee's hand when you lose. The rule is followed by most players, but not shaking the referee's hand does not mean you are deviating from the game of soccer or that you are no longer playing soccer at all. Over time, the rule might even change, and one might be expected to not shake the referee's hand.

In contrast, constitutive rules are essential to the act to which they pertain.[41] A constitutive rule of soccer is that the ball should only be moved around the field by using one's foot. A player who runs around the field with the ball in their hands goes against the very nature of soccer as a game. Furthermore, if the rules of soccer changed to allow players to play with their hands, the game would be constitutively different from the game of soccer that exists now. However, a constitutive rule being broken does not mean that the rule-breaker is no longer participating in the act; a soccer player who touches the ball is still playing soccer, just in a way that is contrary to the game. Thus, while a constitutive rule does not need to be followed in every instance of an act, if it is eliminated or widely ignored, the act is no longer coherent. Williamson uses this distinction to argue that the constitutive norm of assertion is the knowledge norm of assertion: you must only assert what you know.[42] The knowledge norm best captures how ordinary speakers make assertions, which is to convey what they believe to be true—even if they might not fully justify their assertion, or their belief is wrong. For the sake of this paper, I accept that the knowledge norm is the constitutive norm because Sorensen himself takes it to be true.[43]

---

[39]Ibid., 263; I find this interpretation to be plausible because Sorensen deems bald-faced lies as themselves morally neutral, while considering some of the examples (e.g., the plagiarism case) to be morally wrong based on the wrong action it correlates with.

[40]Timothy Williamson, "Knowing and Asserting," *The Philosophical Review* 105, no. 4 (October 1996): 488.

[41]Williamson, "Knowing and Asserting," 488.

[42]Ibid., 495.

[43]Sorensen, "Bald-Faced Lies!," 255.

The knowledge norm proposal

Given this background, I will now lay out my proposal. I first begin with the premise that, by definition, constitutive rules take priority over conventional rules. The knowledge norm is essential to the practice of assertion because the practice of assertion would break down without it, much like how the game of soccer would break down if there was no constitutive rule about playing with your feet. This is because a fundamental aspect of assertions is that they are used to transmit information, and that there is some intellectual authority attached to the person who makes an assertion.[44] For example, if I assert to my classmate that PHIL 3912 meets on Thursdays, my classmate will take me to know that this is the case, and she might even act based on that assertion. Without this the aspect, there would be no purpose to making assertions, as there would be no utility or consequence to making such claims. Conventional rules of assertion, on the other hand, are inessential. For example, take a convention that you should avoid impolite assertions about other people in a social setting. One can make an assertion that breaks this rule without violating the purpose of assertions; for example, you can correctly assert that someone's child is bad at soccer, thus conveying that information to the child's parents, even if the assertion comes off as offensive.

Second, bald-faced lies violate the knowledge norm. Sorensen states that lying involves asserting what one does not believe, and one cannot know something which one does not believe. For the sake of this paper, I stipulate that beliefs generally aim for truth (and that one cannot believe something one does not consider true), because this account of belief is most consistent with Sorensen's paper. Thus, when one utters a bald-faced lie, one asserts what one knows not to be true, which violates the knowledge norm of assertion.

Third, bald-faced lies are often used to follow conventional rules of assertion, as demonstrated by commonly used examples of bald-faced lies. To evidence this claim, I will go over three examples that Sorensen brings up in his paper: the plagiarism case, which I have already introduced, the chicken stealing case, and the accountant case.

In the plagiarism case, student utters the bald-faced lie because of the dean's policy, which incentivizes the student to follow it at the expense of the knowledge norm. As such, the bald-faced lie occurs when the conventional rule is prioritized over the constitutive rule.

The chicken-stealing case is as follows: a chicken thief is caught by his father-in-law with two dead chickens in his hand at 5AM. The father-in-law knows that the son-in-law is the thief, and the son-in-law knows that his father-in-law knows that he is a thief. Yet, the son-in-law asserts that he was checking in on the coop and scared off the real thief. To preserve family unity, the

[44]Williams, "Truth, Assertion, and Belief," 78.

father-in-law does not further interrogate his son-in-law. [45] In this case, it is less clear whether the son-in-law was explicitly aware of the fact that, if he lied, the father-in-law would leave him unpunished; however, he was likely betting on the fact that the importance of preserving family unity would incentivize the father-in-law to acquiesce. In this case, the conventional rule is asserting that which is best for family unity; once again, the conventional rule is prioritized over the knowledge norm when the two are in conflict.

The final case is the accountant case, where a corporate accountant tells the public that he had no knowledge of the company's fiscal improprieties. Even though there is public knowledge that the accountant had knowledge of the improprieties, which the accountant is aware of, he makes the assertion because the legal standard of evidence is higher than the standard of evidence for knowledge; until he is legally proven to have known of the improprieties, he cannot be punished, even if most people know that he did.[46] The accountant tells the bald-faced lie to be consistent with the legal norm for what one can assert—i.e., a conventional rule—at the expense of the knowledge norm.

Thus, the problem with bald-faced lying is that when conventional norms of assertion come into conflict with the constitutive norm of assertion, i.e., the knowledge norm, they involve prioritizing the conventional norms. The practice of assertion does not depend on institutional, social, or legal norms, because such norms are arbitrary and subject to change; it is possible for a family to value honest character, or for a school to demand politeness. But, as stated above, the knowledge norm is intrinsic to the practice of assertion.

The value of truth

To elaborate on why violating the knowledge norm is problematic, we can apply a primitive version of Kant's universalizability test to the constitutive norm of assertion. Kant's universalizability test envisions a world in which there is a "law" that permits everyone to lie whenever it is in their self-interest. My primitive version does not involve the creation of a law but envisions the widespread violation of the knowledge norm. This would result in a world in which individuals feel no obligation to make assertions based on knowledge, and feel no obligation to make it appear as though their assertions are based on knowledge. In such a world, the practice of assertion would break down, because statements would no longer be a reliable way to convey information. Deeming bald-faced lies to be morally neutral undermines the practice of assertions and, by extension, their critical role in transmitting knowledge.

This approach addresses the two concerns I identified with Sorensen's account. First, it more precisely explains why bald-faced lies misuse and ultimately undermine the practice of assertions. Additionally, by clarifying the

[45] Sorensen, "Bald-Faced Lies!," 254

[46] Ibid., 261.

superior status of the knowledge norm of assertion, it offers a way to condemn constitutive norms that incentivize lying — the most flagrant example being the dean's plagiarism rule. Conventional norms of assertion should not undermine constitutive norms. Second, by identifying what is intrinsically wrong with bald-faced lies, this approach does not rely on the goodness or badness of the actions, circumstances, or consequences that are correlated with bald-faced lies. Regardless of whether the bald-faced lie is told to achieve a good or bad objective, it is wrong because it undermines the knowledge norm of assertion; widespread use of the bald-faced lie to achieve good objectives would still lead to the breakdown of information transmission.

Sorensen objects that this approach fails to consider lying to oneself. He argues that "[t]here is no hope of a successful (synchronic) deception when you are lying to yourself," even if you might believe the lie later. Because the perpetuator of the lie is identical to the recipient, there is no asymmetry that makes the lie problematic.[47] Such an objection, however, still operates on the assumption that lies are only wrong if they involve a power imbalance or violation of trust between the perpetuator and the recipient. This assumption is problematic for two reasons. First, there are many instances in which lies are condemned, even if the recipient is in a position of power over the perpetuator: a child who lies to their parent, and even the plagiarism kid of the student lying to the dean. Second, does not contest that there might be something intrinsically wrong with making a false assertion for "external reasons," which is what my account is concerned with. Finally, Sorensen's argument only proves that lying to oneself might be less bad than lies to others; it does not contest the common intuition that self-lying can be harmful to oneself.

Another objection might be that bald-faced lies cannot truly harm the practice of assertion because they occur in isolated instances in response to specific conventional rules. For instance, the dean knows not to believe the student's statement—even if he uses it to determine whether she will be punished—because he is aware of his own conventional rule. However, the examples given for bald-faced lying indicate that there are a wide-ranging set of contexts in which conventional rules disincentivize telling the truth. Indeed, the scope of social norms that concern assertion, whether it be politeness, respect, or family unity, suggest that bald-faced lies do not require specifically delineated rules, like in the plagiarism case.

Furthermore, this objection does not consider an additional dimension of bald-faced lying: what Sorensen calls "escalating absurdity." Because the bald-faced liar wants to be on the record for their lie, they are willing to continually make statement that "double down" on their original lie and are increasingly more absurd.[48] If, after the student stated that she did not plagiarize, the dean asked her if her paper was identical to the paper she plagiarized (assuming the

[47]Ibid., 263.
[48]Ibid., 253.

student did copy it word-for-word), the student might further state that they are different. This escalates the absurdity of the original lie, because the fact that the papers are identical is even more obvious to the recipient than the fact that she did not plagiarize. Not only does this phenomenon evidence the self-reinforcing nature of bald-faced lies, but it further emphasizes the importance of condemning bald-faced lies in themselves, as "escalating absurdity" is inherent to the bald-faced lie and how it clearly contradicts what all relevant parties know to be the case.

Conclusion

Ultimately, though Sorensen may be right to criticize those condemnations of bald-faced lies that merely rely on the wrongness of deception, he is wrong to conclude that they are morally neutral. By focusing on the relationship between bald-faced lies and the knowledge norm of assertion, it is possible to condemn bald-faced lies in themselves—as well as the conventions that incentivize them to occur.

Bibliography

Fallis, Don. "What is Lying?" *The Journal of Philosophy* 106, no.1 (Jan. 2009): 43.

Sorensen, Roy. "Bald-Faced Lies! Lying Without the Intent to Deceive." *Pacific Philosophical Quarterly* 88 (2007): 252.

Williams, Bernard. "Truth, Assertion, and Belief." In *Truth and Truthfulness*, 78. United Kingdom: Princeton University Press, 2002.

Williamson, Timothy. "Knowing and Asserting." *The Philosophical Review* 105, no. 4 (October 1996): 488.

Slabs, SMS, and Semantic Dependence

Liana Raguso

Abstract. What does it mean for one linguistic expression to be 'short for' another? This paper claims that 'short for' constitutes a psychological rather than formal relationship. To develop this definition, I will turn to later Wittgenstein, seeking to adjudicate a debate between Wittgenstein and his interlocutor in *Philosophical Investigations*: while the interlocutor claims that a builder's cry of "Slab!" must be 'short for' "Bring me a slab," Wittgenstein calls this view into question. Their debate puts forth two plausible definitions of the 'short for' relationship, one based on formal necessity and the other on mental behavior. We will evaluate these definitions by examining a case study which was not available to Wittgenstein: the variable uses of texting abbreviations. Texters' use of such abbreviations independently from their parent phrases discredits the logical definition of 'short for'. Through an examination of how non–texters use the same abbreviations, I propose an alternate definition of 'short for' centered around the act of mentally translating between one phrase and another.

In *Philosophical Investigations*, Wittgenstein argues with his interlocutor that the sentence "Slab!" is not 'short for' the longer sentence "Bring me a slab." Wittgenstein's claims rest on his implicit definition of 'short for,' which differs from that of the interlocutor. This paper will seek to evaluate Wittgenstein's stance through the process of developing a definition of 'short for' that does not imply logical dependence between terms. A case study of texting abbreviations as used by texters versus non-texters reveals that Wittgenstein is largely correct in his definition of 'short for', though he leaves this definition incomplete. When an abbreviation is short for another expression, a language user must mentally translate from the shorter to the longer to understand the abbreviation, so Wittgenstein is correct in saying that "Slab!" is not short for "Bring me a slab" in this case.

While Wittgenstein's interlocutor characterizes 'short for' as a formal relationship, Wittgenstein calls this characterization into question. In paragraph 19, Wittgenstein discusses with an interlocutor the expression "Slab!", used when a builder wants his assistant to bring him a slab:

> Wittgenstein: is the call "Slab!" in example (2) a sentence or a word? [...]
>
> Interlocutor: "Slab!" is surely only a shortened form of the sentence "Bring me a slab."

> Wittgenstein: But why should I not on the contrary have called the sentence "Bring me a slab" a *lengthening* of the sentence "Slab!"?
>
> Interlocutor: Because if you should say "Slab!" you really mean: "Bring me a slab".
>
> Wittgenstein: But how do you do this: how do you *mean that* while you *say* "Slab!"? Do you say the unshortened sentence to yourself? And why should I translate the call "Slab!" into a different expression in order to say what someone means by it? And if they mean the same thing—why should I not say: "When he says 'Slab!' he means 'Slab!' "? Again, if you can mean "Bring me the slab", why should you not be able to mean "Slab!"?
>
> Interlocutor: But when I call "Slab!", then what I want is, *that he should bring me a slab*.
>
> Wittgenstein: Certainly, but does 'wanting this' consist in thinking in some form or other a different sentence from the one you utter?[49]

The interlocutor seems to think that the sentence "Slab!" would be incoherent on its own were it not 'short for' "Bring me a slab." This position is embedded within a broader view of language according to which words basically represent objects; treated as stable representational units, meaningful words combine to form meaningful sentences in which each word conveys an antecedently specified meaning and serves a limited structural role. As a result, the interlocutor thinks that a speaker must incorporate words representing each constituent part of the concept of wanting to be brought a slab — the command ("bring"), the recipient ("me"), and the object to be brought ("a slab") — in order to express this desire. According to the interlocutor's understanding, the word "slab" on its own serves a merely ostensive function; that the exclamation "Slab!" can express completely the speaker's desire of being brought a slab only makes sense if "Slab!" somehow *really means* or is 'short for' "Bring me a slab." This seems to be what the interlocutor wants to express by saying, "Because if you should say 'Slab!' you really mean: 'Bring me a slab'." For the interlocutor, being 'short for' is thus a formal relationship of part to whole invoked in contexts where the part makes no sense on its own. Whether this relationship involves certain mental processes is largely irrelevant: being 'short for' is a logical necessity demanded by the structure of language.

For Wittgenstein, being 'short for' is primarily a mental relationship, involving meaning or form only secondarily. Since Wittgenstein sees word-meaning as determined by usage rather than strict representative relations (to present a simplified account of his position), he finds no problem with "Slab!" expressing in its entirety the concept which might otherwise be expressed as "Bring me a

[49]Ludwig Wittgenstein, *Philosophical Investigations*, trans. G.E.M. Anscombe, P.M.S. Hacker, and Joachim Schulte. 4th ed. (N.p.: Wiley-Blackwell, 2009), 11-12.

slab." It is possible but not logically necessary that "Slab!" could be 'short for' "Bring me a slab" under this view. Wittgenstein questions what it would mean experientially for "Slab!" to be 'short for' something else—by asking, "Do you say the unshortened sentence to yourself?", he implies that saying "Slab!" would have to involve thinking, "Bring me a slab" if the 'short for' relationship were in place. Since we do not think anything other than "Slab!" when we say "Slab!", Wittgenstein seems to conclude that "Slab!" is not 'short for' anything and just as meaningful as "Bring me a slab." At best, we can say that "Slab!" and "Bring me a slab" have identical meanings, but if we let this constitute "Slab!" being 'short for' "Bring me a slab," equally so does it constitute "Bring me a slab" being 'long for' "Slab!" Importantly, both phrases still stand independently without need of each other. The validity of Wittgenstein's position depends in one sense on a larger argument over whether linguistic meaning is determined by use, but it depends more specifically on whether 'short for' should be understood as a formal relation or a mental behavior.

In order to evaluate the interlocutor and Wittgenstein's contrasting notions of 'short for', we will see how well they apply to the case of texting abbreviations as used by texters and non-texters. Let us lay out the parameters of this case study: though texters and non-texters are by no means definitive or exact demographic groups, within this essay, texters (as opposed to non-texters) will be those who possess greater comfort with texting abbreviations and use them regularly and fluidly within daily life. Non-texters are not necessarily those who never text but rather those who text less frequently and less naturally. Texters may be young people who grew up with an awareness of texting abbreviations, while non-texters may be older. This essay will focus on texting abbreviations which are acronyms constructed from the first letter of each word in a phrase (e.g., "lol" = "laugh out loud"), as opposed to those which are shortened forms of individual words (e.g., "nvm" = "never mind"). The phrase or sentence from which a texting abbreviation derives will be referred to as its "parent phrase." The relationship of texting abbreviations to their parent phrases is thus similar to the relationship between "Slab!" and "Bring me a slab"; in both cases, one linguistic object is composed from components of the other, though the components are letters in the case of texting abbreviations and a word in the case of "Slab!" Through an analysis of how texting abbreviations relate to their parent phrases in various contexts, I will determine which definition of 'short for' better fits the structure of language.

The interlocutor's understanding of language and the state of being 'short for' implies that texting abbreviations must be 'short for' their parent phrases. Composed from fragments of their parent phrases, texting abbreviations stand to their parent phrases in the relationship of part (or collected parts) to whole. Because these parts are fragments of the whole expression, they cannot stand independently. Thus, we are worse off than in the case of "Slab!" where the word "slab" at least had a referent, even if it did not constitute a full

sentence. Individual letters lack representative power out of context, and their arrangement in an acronym is meaningless on its own: if "lol" did not 'stand for' "laugh out loud," it would have no meaning. Yet because texting abbreviations do in fact have meaning, this meaning must come from their ability to represent the parent phrases they are short for.

However, texting abbreviations demonstrate themselves not to be 'short for' their parent phrases when used by texters; the interlocutor's theory thus fails. Rather than relying on their parent phrases for meaning, texting abbreviations take on full semantic independence in the vernacular of texters. The most extreme form of this independence manifests in texting abbreviations whose significations deviate from those of their parent phrases. To provide one example, "lol" is typically used equivalently with "I find this slightly funny" rather than "laugh out loud," or else used as a filler term, somewhat comparable to "um" in verbal expression. Thus, "lol" cannot be 'short for' "laugh out loud" when "lol" and "laugh out loud" do not mean the same thing. Since abbreviations can have different meanings than their parent phrases, they do not need their parent phrases to provide them with meaning. Despite their etymological relationship, we must conclude that texting abbreviations and parent phrases stand as separate expressions with independent meanings which may or may not converge. Texting abbreviations' status as independent units can also be seen in their verbal/sonic expression: for example, texters typically pronounce "lol" (whether verbally or through an interior voice) as a singular term, "lawl," instead of spelling out the letters, "L-O-L" or expanding to the parent phrase "laugh out loud." This sonic independence remains even when the meanings of a text abbreviation and its parent phrase converge — "ngl" and "not gonna lie" have equivalent meanings, and "ngl" is typically read out loud as "N-G-L," but texters may still read "ngl" in a manner phonetically similar to "niggle" in their own heads. These actions disregard the part-to-whole relationship and instead treat abbreviations as independently coherent units. In a similar vein, texters also assert the independence of texting abbreviations through their writing practices, writing texting abbreviations in all lowercase like any other word ("ngl" instead of "NGL" or "N.G.L."). Linguistic expressions formed from fragments of longer expressions do not necessarily attain meaning through reference to their parent phrases, from which their meaning may diverge as the parent phrase/abbreviation relationship is deemphasized and becomes vestigial. The interlocutor's conception of 'short for' has collapsed.

In response, the interlocutor might say, "I cannot deny that texters use texting abbreviations as independent linguistic units, but they are wrong to do so and confused about what they are really saying." I would point out that if texters view texting abbreviations incorrectly, there is in fact a group of people who view them correctly (in the interlocutor's terms). It is non-texters who tend to see a relationship of dependence between texting abbreviations and their parent phrases — as we will examine later in more detail. The interlocutor thus

finds themself claiming that a group of people who are less familiar with texting abbreviations, who use them with less ease, and who are generally aware of only a limited number of texting abbreviations are more correct in their use and understanding of these abbreviations than a group of people who are familiar with these abbreviations, who use them naturally and accurately, and who, in fact, invent them. The interlocutor's stance has lost its viability.

Having dispensed with the interlocutor's definition of 'short for', we will now develop a viable definition by examining what occurs between two expressions when the 'short for' relationship is actually in place. Though texting abbreviations as used by texters are not 'short for' their parent expressions, texting abbreviations as used by non-texters seem to be genuine examples of the 'short for' relationship. When using texting abbreviations which have come to hold modified meanings when used by texters, non-texters typically remain unaware of these modified meanings and continue to use such abbreviations interchangeably with their parent phrases. When reading text messages out loud, non-texters often speak the separate letters or the parent phrase instead of treating the abbreviation as one word; when texting, they often capitalize each letter in the abbreviation, calling attention to the relationship between the letters of the abbreviation and the words of the parent phrase rather than treating the abbreviation as an independent word. Through these speech and writing practices, non-texters emphasize the relationship between texting abbreviations and the parent terms from which they derive rather than treating abbreviations as singular, independent linguistic units. I will thus take texting abbreviations in their use by non-texters as an entry point into considering what 'short for' means.

By analyzing non-texters' reading practices regarding texting abbreviations, I can confirm and expand on Wittgenstein's implied definition of 'short for' as a mental process of translation, exploring the role of semantic dependence in 'short for' relations. Suppose a non-texter reads a text containing "lol" out loud. Plausibly, they may read the separate letters of the abbreviation while exhibiting signs of confusion, then verbally 'translate' the abbreviation into its parent phrase, which they will then respond to with signs of realization and understanding. Even when this 'translation' process does not occur verbally, there may be a temporal lag after non-texters read an abbreviation during which they appear confused; they may then realize the meaning, presumably after translating the abbreviation in their head, and carry on reading the text. This temporal lag while the abbreviation is 'decoded', or 'translated', reveals that the abbreviation is not understood on its own grounds but rather understood in reference to the parent phrase, which must be mentally recalled and substituted into the message. In accordance with Wittgenstein's definition of 'short for', texting abbreviations are mentally translated when they are heard, and presumably a non-texter who types a texting abbreviation is *really saying* the parent phrase in their head. To complete Wittgenstein's definition, we must also note that these mental

processes are necessary for the comprehension of text abbreviations rather than occurring coincidentally. I add to Wittgenstein's definition the interlocutor's notion of semantic dependence: when one expression is 'short for' another, it has no comprehensible meaning independently of the expression for which it is short. However, this dependence is not a necessary structural dependence founded on the relationship of part to whole; it is a circumstantial dependence predicated on the speaker or listener's linguistic reference frame. Our notion of semantic dependence can be thought of as comprehensional dependence, with comprehension approached as it is experienced by language-users. Thus, semantic dependence is recognizable by language-users in their own experience and exhibited in their actions; the independence of texting abbreviations as used by texters does not reduce to an automated or subconscious dependence, since an imperceptible comprehensional dependence does not count as dependence in our terms.

I now finally return to the case of "Slab!" According to the definition I have developed, "Slab!" is 'short for' "Bring me a slab!" if and only if understanding "Slab!" necessarily entails mentally translating "Slab!" to "Bring me a slab!" for a given understander. Thus, in most contexts, "Slab!" will not be 'short for' "Bring me a slab!" — the builder will yell "Slab!" without thinking "Bring me a slab" to him or her self, and the assistant will respond to this call without having to translate in their mind. In other words, the psychological processes entailed by 'short for' will not occur, and "Slab!" will serve as an independent unit of meaning. However, "Slab!" might indeed be 'short for' "Bring me a slab" under certain contexts. For example, on the assistant's first day, they might be confused upon hearing "Slab!" — in order to respond successfully, they might have to think, "Oh, what the builder means is, 'Bring me a slab,' that's what I should do." As time passes, the assistant will likely cease this process of mental translation, and "Slab!" will cease to be 'short for' "Bring me a slab" in their mind. In Wittgenstein's example, "Slab!" is understood without a psychological process of translation, so we can conclude that "Slab!" is not 'short for' "Bring me a slab."

Through exploring the use of texting abbreviations by both texters and non-texters, I have validated the Wittgensteinian claim that "Slab!" is not necessarily short for anything else and arrived at a new definition of 'short for' as a relation of mental semantic dependence. The interlocutor's conception of 'short for' as a formal relationship has been discredited through its incompatibility with texters' treatment of texting abbreviations as independently meaningful. A better definition of 'short for' has emerged from non-texters' treatment of these same abbreviations: I posit 'short for' as a psychological (and not formal) relationship of semantic dependence, as this dependence is actually experienced by language-users (and not merely assumed to be necessary). Further inquiry may find that our argument carries broader implications for the theory of language which produced the interlocutor's definition of 'short for' and lead to

further insights in our understanding of language's formal structure.

Bibliography

Wittgenstein, Ludwig. *Philosophical Investigations*. Translated by G.E.M. Anscombe, P.M.S. Hacker, and Joachim Schulte. 4th ed. N.p.: Wiley-Blackwell, 2009.

A Defence of the Sensitivity Analysis of Knowledge

Eric Wallace

Abstract. It is widely considered that Nozick's sensitivity analysis of knowledge fails. This is largely due to arguments proposed by Sosa. In this paper, I defend Nozick's sensitivity condition on knowledge. To do this, I define and motivate sensitivity, then explain Sosa's definition of safety and its supposed advantages. For all three of Sosa's purported counterexamples, we will find that the sensitivity theorist can offer responses that are more satisfactory than those available to the proponent of Sosa's safety theory. Having motivated sensitivity and shown that it deals with these purported counterexamples better than safety, I will conclude that Sosa fails to motivate a move from sensitivity to safety.

One key question in epistemology is "what is knowledge?". Traditionally, a response to this question has been given in the form of necessary and sufficient conditions. One early analysis of knowledge was Nozick's sensitivity analysis. While this theory enjoyed a short-lived popularity, it has fallen into disrepute in large part due to the objections of Sosa. This paper aims to show that these objections miss their mark. The sensitivity analysis of knowledge comes out unscathed.

The Intuition Behind Sensitivity

The basic statement of sensitivity is as follows:

> **S**'s belief that **P** is sensitive iff, if **P** were false, **S** would not believe that **P**.[50]

The conditional on the right hand side is a subjunctive conditional. There are several proposed methods for modelling these, but Nozick opts to model them as follows: The conditional is satisfied iff, in the closest possible world where **P** is false, **S** does not believe that **P**.[51] The insight here is that beliefs that do not track truth cannot count as knowledge, and we can model truth tracking using a subjunctive conditional. A key reason for thinking that knowledge must track truth is that lucky knowledge, that is a belief that constitutes knowledge by chance, is implausible. A truth tracking condition rules out such cases of

[50]Robert Nozick, "Knowledge and Skepticism," In *Epistemology: an anthology* ed. Ernest Sosa and Kim Jaegwon, (Oxford: Blackwell, 2000), 79.

[51]Nozick, "Knowledge and Skepticism", 80.

lucky knowledge because beliefs that are true as a result of luck will not track truth.

The basic statement of sensitivity above needs to be amended to factor in methods of knowing. The amended definition is as follows:

> **S**'s belief that **P**, arrived at via method **M**, is sensitive iff, if **P** were false and **S** used **M** to come to a belief about **P**, **S** would not believe that **P**.[52]

Consider a person whose friend phones to say that they will be late. This friend is normally late, so, had their friend not phoned, they still would have assumed their friend was going to be late. While the belief formed based on the friend's testimony seems like knowledge, it fails the unamended definition of sensitivity. The amended definition classifies the belief that the friend will be late, formed based on testimony, as sensitive because if their friend had not been late, they would not have believed that they were going to be late via their testimony. I will use this amended definition of sensitivity throughout the paper. Sosa agrees with Nozick that knowledge must track truth. He also agrees that sensitivity is at least initially plausible as an attempt to capture truth tracking.[53]

Safety and Its Advantages

Sosa characterises safety as follows:

> **S**'s belief that **P** is safe iff, if **S** were to believe **P**, it would be so that **P**.[54]

Sosa's safety condition aims to capture some of the appeal of Nozick's sensitivity condition while circumventing some perceived problems. A safe belief tracks truth in the sense that it could not easily have been wrong. When working out whether a belief is safe, one considers all nearby[55] possible worlds

[52]Ibid., 81.

[53]Ernest Sosa, "How to Defeat Opposition to Moore," In *Philosophical Perspectives* no. 13 (1999) 141, 147; Sosa opts to use Nozick's unamended condition. Presumably this is because he did not think considerations of methods would be relevant to his criticisms. As we will see later, they are.

[54]Sosa, "How to Defeat Opposition to Moore," 141-153. Some use a definition that is amended in a similar way to Nozick's methods amendment. I opt to use this simpler definition for clarity, but everything I say also applies to the amended definition.

[55]One might worry that there is no principled way of working out how nearby a world must be to be relevant. Even though Sosa insists that the relevant alternatives theorist responds to an analogous point, I set this worry aside. This is because this seems to be an instance of a more general problem in the analysis of knowledge. Examples: How relevant does a relevant alternative have to be? How reliable does a cognitive mechanism have to be? How good does justification have to be? How closely tied to an epistemic virtue does a belief have to be? Given the generality of the problem, we had all better hope it can be fixed.

where **S** believes that **P**. If it is the case that **P** in these worlds, then **S**'s belief that **P** is safe.

Sosa gives several arguments that safety has advantages over sensitivity. I will focus on Sosa's counterexamples as his other arguments have been discussed elsewhere. To give reason for maintaining sensitivity in the face of safety, I will show that sensitivity gives better responses than safety.

False Belief Counterexample

I start with Sosa's false belief counterexample because he believes it to be the most compelling, and I agree with him.[56] Sosa asks us to consider a subject **S** who knows some proposition **P** and believes that they do not falsely believe that **P**.[57] I shall call **S**'s belief that they do not falsely believe that **P** their metabelief.[58] Sosa makes the following argument:

> (P1) All metabeliefs about propositions that constitute knowledge are knowledge.
>
> (P2) All metabeliefs about propositions that constitute knowledge are insensitive.
>
> ---
>
> (C3) Some knowledge is insensitive.

Should Sosa successfully show that some knowledge is insensitive, this is sufficient to show that the sensitivity theory is false. Sosa thinks that (P1) must be true because a minimally attentive and reflective agent who knows **P** surely must know that they do not falsely believe that **P**. He defends (P2) on the grounds that in the most likely scenarios where I falsely believe that **P**, I will still believe that **P**. Based on that, I will form a belief that I do not falsely believe that **P**, making my metabelief insensitive.

(P1), however, is not true. To see this, consider someone who has beliefs about the world formed in a manner such that some of them constitute knowledge. They then form metabeliefs not based on reflection upon how they formed their beliefs about the world but based on an arrogant conviction that all their beliefs are correct. Intuitively, it does not seem that a belief formed purely on this arrogant conviction can be knowledge. Such a metabelief not only does not intuitively feel like knowledge, but also is not safe, so the safety theorist cannot maintain that it is knowledge. Considering this, we must revise (P1) to factor in that some metabeliefs about knowledge are not knowledge as follows:

[56] Ibid., endnote 11.

[57] Ibid., 145.

[58] I introduce this nomenclature for brevity and legibility, though I appreciate that the term 'metabelief' is usually used more broadly.

> (P1*) Some metabeliefs about propositions that constitute knowledge are knowledge.

The new argument with (P1*) replacing (P1) is still valid. Consider someone who knows that they have over-brewed their tea via the method of tasting it, and suppose that they form a belief that they know that they do not falsely believe that they have over-brewed their tea via the method of reflecting on their tasting the tea. To evaluate whether this metabelief is insensitive, we must consider whether the agent would still hold their metabelief in the closest possible world where they do falsely believe that the tea is over-brewed *via the method of reflecting on their tasting it*. One plausible candidate for such a world is a world where the agent misremembers how long the tea has been brewing, forms the belief that the tea is over-brewed when it is not via deduction from their knowledge of how long it takes tea to over-brew, and then forms their metabelief by reflecting on their deduction. Importantly, in such a world the agent does not form their metabelief via the same method. In other words, the example meets Nozick's methods definition of sensitivity. The example above is an example of a sensitive metabelief. Hence, (P2) is false. If we amend (P2) as follows:

> (P2*) Some metabeliefs about propositions that constitute knowledge are insensitive.

Sosa is faced with the problem that the argument is now invalid.

Now I shall compare how safety responds to these cases. Suppose that knowledge is safe true belief. Consider a case where I believe that my friend will do as I asked them based on their testimony. Suppose that this belief is knowledge. In all likely scenarios where I believe that they will do as I ask, they will do as I ask. Suppose also that I believe that I do not falsely believe that they will do as I asked, not because I reflect on their testimony and their honesty, but because I believe myself to be a master manipulator, which I am not. Intuitively, it seems that my metabelief is not knowledge because it is formed based on a belief that is false. In all close worlds my metabelief arises only when I believe that they will do as I ask, and, by stipulation, there is no close possible world where I believe that they will do as I ask and they do not do so. Hence there is no close possible world where I hold my metabelief, and it is false. The metabelief is a safe true belief, so, for the safety theorist, it is knowledge. Yet it does not seem to be knowledge as it is based on false considerations. This does not show that safety is not a necessary condition for knowledge, but it does show that a fourth condition on top of safety, truth, and belief would be required to eliminate such a case. Hence, this case illustrates a problem facing the safety theorist that does not face the sensitivity theorist.

The Garbage Chute Counterexample

I will now consider Sosa's garbage chute case. In this counterexample, a person releases a bag of garbage down a garbage chute, and the garbage makes it to the bottom of the chute. The person believes that the garbage makes it to the bottom of the garbage chute. Sosa claims that this is a case of knowledge that is not sensitive as, if the garbage had somehow snagged, the person would have believed that it made it to the bottom of the chute.[59] There are three issues with this counterexample. First, it is not safe, so the safety theorist cannot claim that it is a case of knowledge. Second, we can construct structurally similar cases that are not knowledge. Third, cases like the garbage case are cases where belief does not track truth.

To be safe, the belief in the garbage chute case would need to be such that in all close possible worlds where the agent believes that the garbage is at the bottom of the chute, the garbage is at the bottom of the chute. The cases where the garbage does not get to the bottom of the chute are very unlikely so, Sosa claims, the worlds where these cases hold are not close worlds. Hence, even though the agent would believe that the garbage is at the bottom of the chute in cases where it snags, the belief is safe. This is all well and good so long as the worlds where the garbage snags are distant worlds. These scenarios being unlikely does not seem like good reason for thinking that these are distant worlds. To see this, suppose for a reductio that worlds at which unlikely events occur are distant. Imagine that I run a raffle with 500,000 of my friends. It is a fair raffle and exactly one of us will win. The chance that any of us win is 0.0002%, which is unlikely. In all worlds where our raffle takes place, one of us will win. That event had a 0.0002% chance of occurring, so all worlds are worlds where an unlikely event occurs. By our supposition, this means that all worlds are distant worlds. Given that all worlds are distant, no worlds are close. Hence, for any belief I hold, there is no close world where I am mistaken. Hence, all beliefs would be safe, simply because my friends and I ran a raffle. This is absurd, so our assumption that worlds at which unlikely events happen are distant was false. The safety theorist must now concede that the unlikeliness of the garbage snagging does not show that worlds where it snags are distant.

Standardly, worlds that we choose to model as close to the actual world are worlds that are objectively similar to the actual world.[60] If we adopt this view, then worlds where the garbage snags and very little else is changed are close as they are only slightly different to the actual world. All that varies is the location of the garbage and the exact shape of the chute. Given that worlds where the garbage snags are close and they are worlds where the subject in Sosa's example still believes that the garbage is at the bottom of the chute, the subject's belief that the garbage is at the bottom of the chute is not safe.

[59] Ibid., 145.

[60] This view is not without its issues, but it fairs much better than Sosa's unlikelihood view.

Hence, the safety theorist must concede that the garbage chute case is not a case of knowledge. Hence, the safety theorist cannot maintain that it is a counterexample to sensitivity.

Aside from the belief in question not being safe, we had better hope there is independent reason to think that the garbage chute case is not a case of knowledge. Without independent reason, it is hard to see why the garbage chute case should not be considered a counterexample to both sensitivity and safety.

First, it should be acknowledged that not everyone shares the intuition that this is a case of knowledge. I, for example, was very confused by Sosa's assertion that this belief is knowledge.

The case I propose is inspired by Lewis's talk of stakes when ascribing knowledge. We are less willing to tolerate margin for error in ascriptions of knowledge when the stakes are higher. Some choose to explain such cases in terms of context. As Sosa himself opposes contextualist theories of knowledge, this option is not available to Sosa. The following is structurally similar to a garbage chute case but with much higher stakes. Imagine someone working in a nuclear power plant. When they press a button, safety rods slide down a chute and stop the nuclear reactor overheating and killing thousands of people. Only very rarely do the rods snag and not get to the bottom of the chute. Suppose that the worker presses the button and forms the predictive belief that the rods will be at the bottom of the chute, and the disaster has been averted. Also, suppose that the rods do indeed make it to the bottom of the chute. Intuitively, it seems wrong to ascribe knowledge to the worker. Someone who claims knowledge because there is only a 0.001% chance that thousands will die appears to be following the wrong epistemic norms.

Following Sosa in supposing that contextualism is implausible, we have three options. (I) We can accept that garbage chute cases and high stakes cases are knowledge. The implausibility of this claim in the latter case and its incompatibility with otherwise plausible constraints on knowledge (for example sensitivity, safety, relevant alternatives, reliabilism) give good reason to reject this option. (II) We can accept that neither are knowledge. (III) We can deny that the cases are structurally similar. While it may be a failure of imagination on my part, I cannot see how this could be done. For those who do not have the intuition that the garbage chute case is knowledge, there is no issue here; (II) is the obvious choice. For those who are unconvinced, there are several considerations in favour of this option. First, (I) — admitting that the reactor case is knowledge — looks dire. Second, when one is pressed on whether one *really* knows that the garbage is at the bottom of the chute, many are inclined to admit that they do not; they were speaking loosely. The same consideration sits nicely with the reactor case. Third, if we find the initial claim that knowledge tracks truth plausible, then we are committed to rejecting the claim that cases like the garbage chute are knowledge. Consider both the garbage case and the nuclear reactor case. Once the garbage is put in the chute or the button is

pressed, the belief formation methods used neither interact with the progress of the garbage or rods, nor the things that cause garbage or rods to progress down the chute as they do (e.g. the roughness of the chute). There is no sense in which the belief formation of the agent tracks the facts of the matter except in a crude statistical way. Hence, we have reason to reject garbage cases as knowledge: beliefs in garbage chute cases (and structurally similar ones) are not knowledge because they do not track truth.

To reiterate, Sosa's counterexample is not safe, so it cannot give reason to accept safety over sensitivity. Also, it is structurally like cases that do not seem like knowledge, and the claim that it constitutes knowledge is incompatible with the claim that knowledge tracks truth, which is a key motivation for safety.

The Necessary Propositions Counterexample

Sosa believes that necessary propositions cause problems for sensitivity.[61] Consider some proposition that is necessarily true, e.g., Fermat's Last Theorem (FLT), and a subject who believes that proposition, such as Fermat. Fermat believes FLT; FLT is true; but is Fermat's belief sensitive? Sosa believes that there is no way of working out whether Fermat's belief is sensitive because we cannot evaluate the conditional 'if FLT were false, Fermat would not believe FLT via the same method'. This is because there is no closest world where FLT is false, as FLT is true at all possible worlds. This is not problematic for two reasons. First, there are many proposed ways that we can model counter-possible conditionals like the one above. Second, Sosa conflates the technical apparatus with which we are modelling language with the language. Subjunctive conditionals are claims in language that we model using a possible world semantics. The failure of the modelling tool to evaluate the conditional in some cases is not reason to believe that the conditional cannot be evaluated or is not well defined. As an analogy, consider a mathematician who wishes to model a physical system. She sets up an equation to do this and the equation accurately models the system. She then finds that for certain values the equation requires her to divide by zero. In this case she ought to consider this reason to think that her modelling method is flawed, not reason to think that the world exhibits strange behaviour at those values. In practice, this is what mathematicians do. Similarly, the failure of our methods to evaluate the conditional 'if FLT were false, Fermat would not believe FLT via the same method' tells us that our modelling methods are flawed, not that the conditional cannot be evaluated. Hence, necessarily true propositions do not cause problems for the sensitivity theorist. They do, however, cause problems for the safety theorist.

Consider Fermat, who believes FLT, which is necessarily true, but he does so because he believes all conjectures that are named after him. Presumably, a

[61] Ibid., 146.

belief formed based on vanity alone is not knowledge as it does not track truth. Despite this, the safety theory declares this belief to be knowledge. Fermat believes that FLT is true; it is true; and there is no close possible world at which Fermat believes FLT, but it is false. Note: this is a slightly different problem to the one Sosa claimed the sensitivity theorist faced. Sosa worried that the sensitivity conditional could not be evaluated in cases of necessary truths. The safety conditional, on the other hand, can be evaluated; it just gives the wrong results. The safety theorist must say that Fermat's belief is knowledge when it is not. To avoid such problems, the safety theorist must introduce some fourth condition to deal with beliefs about necessary truths. In the absence of such a fourth condition, this gives reason to support sensitivity over safety.

Conclusion

Given that I have motivated sensitivity as a necessary condition for knowledge and shown that it deals with Sosa's counterexamples better than safety, I conclude that Sosa does not give reason to adopt safety over sensitivity. In other words, the sensitivity theorist can give adequate responses to all of Sosa's counterexamples. While there are other attacks on sensitivity theory that need to be dealt with, I have shown that many counterexamples thought to rule out sensitivity do not rule it out. There are other purported counterexamples to sensitivity, which I aim to deal with in future work, but I take the considerations above to conclusively show that Sosa's purported counterexamples pose no serious problems for the sensitivity theorist.

Bibliography

Nozick, Robert (1981). "Knowledge and Skepticism". In Sosa, Ernest; Kim Jaegwon (ed.) *Epistemology: an anthology*. 2000. Oxford: Blackwell, 79-101.

Sosa, Ernest. "How to Defeat Opposition to Moore". *Philosophical Perspectives*, no. 13 (1999): 141-153.

Meaning just ain't in any individual head, an inter–subjective approach to meaning.

Guillem Adrover Clar

Abstract. Putnam's Twin Earth thought experiment creates a division between externalist and internalist approaches to meaning regarding whether we believe that internal duplicates can differ in meaning when they utter the same word or not. I argue that Putnam's externalist approach is wrong because he treats individuals as the subjects of the thought experiment when the appropriate subject for such thought experiments should be communities of individuals, thus acknowledging the social dimension of meaning. I reconstruct Putnam's argument and show that it depends crucially on whether Oscar1 and Oscar2 mean the same thing when they say water. By means of analogy to another thought experiment, Archimedes' Gold, I show that indeed they mean the same thing, relative to the context provided in the thought experiment. In the process of doing so I highlight what we lose if we take an invariantist approach to meaning.

I also provide an example against internalism and in favour of inter–subjectivism about meaning, arguing that the external elements to meaning are not in the world, but are properties of communities of speakers and their contexts. This is done through showing a proposition whose truth value does not seem to depend neither solely on internal characteristics nor the state of the external world, but rather on the characteristics of the relevant community of speakers. In this way I highlight what is wrong with the usual interpretations of what Putnam's Twin Earth experiment shows.

Introduction

In 1975, Hilary Putnam published a paper called "The Meaning of 'Meaning' ". Among the claims made in the paper we find him expressing that "Meaning just ain't in the head".[62] This phrase is meant to summarize the semantic view he proposes in the paper, namely Semantic Externalism. In this essay I will explain, analyse, and evaluate that view as well as propose an alternative view focused on the inter-subjectivity of meaning as a middle ground between Semantic Externalism and Semantic Internalism.

When Putnam claims that meaning is not in the head, he is saying that there are elements of meaning that are external to our mental processes. From this a distinction is drawn between views that support or reject the statement. Semantic Externalism would be the view that supports the statement, i.e. there is

[62]Hilary Putnam, "The meaning of 'meaning'," In *Mind, Language and Reality* ed. Putnam (Cambridge: Cambridge University Press,1975) 227.

some element to meaning that is external to the mind, and Semantic Internalism would deny it, i.e. everything in meaning is mentally internal, i.e. relative and reducible to psychological states and ideas.

Putnam supports his view mainly by means of a thought experiment now referred to in the literature as Twin Earth. Summarizing, Putnam asks us to imagine that there is a distant planet called Twin Earth where everything is exactly the same as here on Earth except some key factors. One such difference between Earth and Twin Earth is that water is not H2O on Twin Earth but a liquid with a very intricate formula abbreviated to XYZ. Furthermore, the water in Twin Earth has the same behaviour as H2O in the normal temperatures and pressures as to be indistinguishable in normal contexts to our water here on Earth. Then Putnam asks us to think of the context of Earth in 1750 when the chemical structure of water was unknown and assume something similar of Twin Earth, then he presents Oscar1 and Oscar2, which are internal duplicates (they share all internal features, in particular they possess identical psychological states) from Earth and Twin Earth respectively. Putnam's point is that while Oscar1 and Oscar2 have the same psychological state towards water and waterte respectively they do not mean the same when they say the word "water", since the first is referring to H2O and the latter is referring to XYZ. The point Putnam was trying to show in his paper is that psychological state does not determine extension. Putnam's argument, if sound, would show that there are external elements to meaning, proving the externalist thesis (Putnam 1975).[63]

The argument is simple, since most of the heavy lifting is done by the context provided by the thought experiment:

> (P1) Oscar1 and Oscar2 mean different things when they say the word "water"
>
> (P2) If Oscar1 and Oscar2 mean different things when they say the word "water", then meaning is not purely internal.
>
> (C) Meaning is not purely internal.

The argument is a straightforward application of Modus Ponens, and therefore is undoubtedly valid. If we want to deny the conclusion of the argument we need to either deny (P1) or (P2). If Oscar1 and Oscar2 are internal duplicates (which they are by assumption) it would be very difficult to deny (P2), since that would require to affirm that meaning is purely internal but that internal duplicates can differ in meaning when they speak. This seems to lead to contradiction with the concept of "purely internal". Therefore, if someone were to maintain that meaning is purely internal, he would need to deny (P1), i.e. he would need to present an argument explaining why Oscar1 and Oscar2 mean the same thing when they say the word "water".

[63] Putnam, "The meaning of 'meaning'."

In this essay I will argue that Oscar1 and Oscar2 mean the same when they say the word "water" but that nonetheless meaning is not purely internal.

To establish that I will also discuss a different thought experiment presented in the Putnam paper. I will refer to said thought experiment as Archimedes' Gold. Summarizing, it consists of imagining Archimedes, or any man of science for that matter, who had a number of methods for determining if a given piece of metal was gold or not. Said methods were more rudimentary than those available to us nowadays, therefore there were some metal pieces, which were for Archimedes, as far as his methods for identifying gold were concerned, undistinguishable from gold and that consequently Archimedes did refer to with the ancient Greek word for gold. What Putnam argues from this is that we want to able to say that Archimedes was asserting falsehoods when he claimed that a non-gold metal piece was gold, even if everyone would have agreed with him that it was gold and it met the standards of its time for being considered gold. This example is relevant to the discussion at hand since it provides good intuition on why it is desirable to affirm that Oscar1's and Oscar2's utterances have different meaning when referring to different things even if water from Earth and Twin Earth are, as far as their methods for distinguishing them are concerned, indistinguishable. That is the crux of the meaning of Putnam's "meaning just ain't in the head": he means that we need certain facts about the state of affairs of the world to be able to tell the meaning of certain utterances. This is so because when Archimedes says that something is gold, he is saying that that thing has a certain chemical structure consisting of gold atoms in a certain proportion and structure and so on, even if he or no one is aware of that fact. There is also a very clear parallelism between the Twin Earth example and the Archimedes Gold example, since both rely on comparing the meaning that we perceive in our current situation (either temporal or spatial) with an alien situation in which some terms seem to differ in meaning. Putnam's argument is especially concerned with natural kinds, like water or gold, which he argues have definitions for their extension that do not depend on our knowledge of them.[64] In a sense when we start referring to water as "water" we are designating the set of chemical and physical properties that constitute the extension of the natural kind of water. That is why, for Putnam, "water" refers to water H2O here, in Twin Earth and thousands of years ago, and it did so since we picked a word for that distinctive kind of thing that nowadays would translate to "water". Concisely, for Putnam, it is contingent that "water" refers to water but it is necessary that water is H2O[65].

[64]Ibid., 222.

[65]It may help to elucidate the point to think of "water" as a rigid designator (Kripke 1979), where our word "water" refers to water (i.e. water in the actual world, H_20) across all possible worlds. Putnam also follows Kripke insofar as identities between rigid designators are necessary (although it is contingent that "water" happens to pick out water).

Archimedes and the time traveller

What I want to argue is that Archimedes can speak truthfully when he says that a non-gold piece of metal is gold if certain conditions are fulfilled when he says it. As I have explained, Archimedes' situation is analogous to that of Oscar1 and Oscar2, so establishing that Archimedes can speak truthfully in this way implies that "gold" means somethings that is not gold, then "water" can mean XYZ under certain circumstances. Putnam is mistaken in the way he frames these two thought experiments since in both of them he treats the subjects (Oscar1, Oscar2 and Archimedes) as isolated subjects failing to recognize the social web that underpins meaning. I want to argue that meaning has an external component but that said external component does not lie in the world as such (as an object) but rather in the relations of concepts we establish as a society. If tomorrow we all decided to start using the word "water" differently (say to refer to H_20_2 or to XYZ) it would be very hard to argue that the meaning of "water" has not changed, at least in some sense. Hence, I would like to argue that the non-internal component of meaning should be inter-subjective rather than objective.

I want to propose a follow-up thought experiment to Archimedes' Gold. It involves two subjects, Archimedes and a time traveller, who we will name Bob. Let us suppose that Bob, a regular nowadays human with extensive chemical and physical knowledge, meets Archimedes and informs him that a given piece of metal that Archimedes had identified as gold is not in fact gold. Furthermore, he provides Archimedes with the reasoning for claiming so and with the adequate scientific methods shows Archimedes the difference in behaviour between the non-gold piece and the true gold piece. In this moment, Archimedes can recognize the error he made and can claim that he was mistaken when he identified that piece as gold. The thing is, the meaning of gold has changed during the conversation that Archimedes and Bob have had. At time t1, the beginning of the conversation, "gold" for Archimedes had a meaning that was in line with what his peers considered to be gold, and at a subsequent t2, at the end of the conversation, Archimedes had bettered his understanding of gold and had aligned it to the nowadays understanding of the term. Still, the meaning of the term "gold" generally in 3rd century B.C.E. Greece had not yet changed, it would be customary for that, that Archimedes went out and shared his discovery with the rest of the scientific community so that the meaning of "gold" could be updated[66].

One may wonder why go through so much trouble when we could just affirm,

[66]It may be argued that this would lead to arbitrariness in regards to the question of which percentage of the scientific community/ general population needs to agree to update the meaning of a term. I think I am content to keep these boundaries fuzzy or context-relative (within reason). Some communities of speakers may be more or less resistant to updating their terms on the basis of scientific agreement. In any case, the solution will involve considering groups of speakers as the subject in regards to meaning.

like Putnam does, that meanings are invariant in time. I will call this being an invariantist (in regard to meaning). Consider the case in which Archimedes, by himself, finds out, through some properties or experiment, that a non-gold piece of metal behaves different than gold. If we take the invariantist approach we want to say that when he asserts "I was wrong, this metal is not gold" he is right and speaking truthfully. Then imagine that he possesses a piece of gold, that while being gold is, say, a different isotope of gold which leads the piece to behave slightly different than gold in certain circumstances, and in a similar vein Archimedes may notice this difference and assert "I was wrong, this metal is not gold". Putnam would have to say that in this case Archimedes is wrong, but on what ground is he wrong? Archimedes does not have any means of differentiating between the two cases, and the only way that we can differentiate is if we know *a priori* by assumption which pieces of metal are gold and which are not in regard to our current definition of the term. But said definitions, the cutting points between what is and what is not an X, are somewhat arbitrary. In other words, Putnam takes the set of natural kinds as a given, but even what counts as a natural kind in a theory is arbitrary[67]. It is conceivable that in a possible world, maybe where different isotopes of gold are more prevalent, these are given concrete terms to pick them apart. In that world Archimedes is right in both cases when he says "I was wrong, this metal is not gold", but the natural kind of gold is unchanged. The problem here with semantic externalism, like with externalism about knowledge and justification, is that it requires to know truth a priori to be able to assess truth, but we would like to be able to assess truth without knowing truth to begin with, since that is when we need the most to be able to assess truth. Without knowing a priori which pieces are gold and which aren't, semantic externalism cannot tell us a thing about the proposition "I was wrong, this metal is not gold".

Example against the Internalist Approach

Therefore, as I argued, we ought to reject the externalist approach to meaning which relies in objective facts about the world. Some may enquire then, why not embrace a fully internal conception of meaning? I will present an example against this, which in turn supports the inter-subjectivity of meaning. I assume that this example represents a common intuition that could be generalized further into a proper argument. The example relies in the proposition "The arbitrary dog is a golden retriever". I assume that for most people this

[67]Here I am taking a more radical position than Putnam and Kripke. It is not only that "water" contingently refers to water, but the fact that we choose to differentiate certain natural kinds from others, by giving them different names (the cutting points for something to count as an X or not X) are themselves contingent, and to some extent, arbitrary. The tendency of treating the world (objective reality) as a given (extremely common in contemporary analytical philosophy) is criticized by Nagel (1984, 27).

proposition seems false, since it is rather arbitrary, even for those for which the term "arbitrary dog" conjures in their mind a golden retriever. Hence, there has to be more to meaning than the ideas and the psychological states of a given individual. One reason for which someone, for whom his own *personal* "arbitrary dog" is indeed a golden retriever, would be hesitant to affirm that the proposition is true is in part because he knows that that is not necessarily the case for everyone. Here we can appreciate that we believe that meaning is a shared construct that can sometimes differ to the ideas and states that conform what we could call the "private meaning" of words, which I would argue is out of the scope of language. We could think of internal copies making this claim and the truth-value varying and the sufficient condition for this change would not be the state of affairs of the world but rather the aggregate of knowledge of the subject in said world. Let me explain: It is not necessary that all dogs are golden retrievers for the proposition to be true. It is sufficient that the community of people for which that utterance is meant to make sense has only come into contact with golden retrievers, so that they all associate "golden retriever" with "arbitrary dog". The subject in the case of meaning is not an individual but a community of speakers.

Conclusion

When I underline the social aspect of meaning I do not want to imply that all there is to meaning are social language games, as some readings of Wittgenstein may imply (1953: §43)[68]. I believe that the rules (presupposition, accommodation, etc.) that our language games possess, as showcased by Lewis,[69] are an integral part of language and conversation but are certainly not the whole picture. In my example, I have hinted to the idea of our own "private meanings" and the aggregate collective of shared ideas and concepts interact to form what we call "meaning". In the time-travelling example, Archimedes knowledge of the natural kind gold has advanced, but until he communicates that discovery, the meaning of the term "gold" at his time has not changed yet. In Archimedes Gold, the pieces that are thought to be gold act as if they were gold for all their social and relevant physical uses and Archimedes would be certainly deceiving someone at his time were he to tell him that that wasn't gold, and in a similar vein, as established, Oscar1 and Oscar2 mean the same when they say the word "water", since their respective communities of speakers are identical.

[68]The referenced passage is "For a large class of cases—though not for all—in which we employ the word "meaning" it can be defined thus: the meaning of a word is its use in the language". For more detail on Wittgesteinian "Use" theories and their relationship with language games, see Lycan (2018, 77-79). More examples of the kind of thinking I am eluding to can be found in the "slab" and "block" languages in §§2-10 and in the Private Language Argument (§§243-315).

[69]David Lewis, "Scorekeeping in a Language Game," In *Journal of Philosophical Logic* (1979).

Therefore, if the arguments, ideas, and examples I have presented are correct I would have succeeded in showing that Oscar1 and Oscar2 mean the same when they say the word "water" and if so, (P1) of Putnam's argument would turn out to be false. Nonetheless, in the process of showing this we have established the need for external elements of meaning, namely those of inter-subjective nature, drawing attention to the social aspects of meaning and to the advantages of thinking of communities of speakers as the true subjects of the thought experiments regarding meaning. As explained, if this is the case, meaning would not be purely internal. In conclusion, meaning just ain't in the head, at least not in any individual head, but not in the way Putnam thought.

Bibliography

Frege, Gottlob. "Sense and Reference." *The Philosophical Review.* Durham, NC: Duke University Press (1948).

Kripke, Saul. *Naming and Necessity*. Oxford: Blackwell, 1979.

Nagel, Thomas. *The View from Nowhere.* Oxford: Oxford University Press, 1986.

Lewis, David. "Scorekeeping in a Language Game." *Journal of Philosophical Logic*. New York: Springer Publishing (1979).

Lycan, William. "Use theories." *Philosophy of Language: A Contemporary Introduction*. New York: Routledge (2018).

McKinsey, Michael. "Skepticism and Content Externalism." *The Stanford Encyclopedia of Philosophy* ed. Edward N. Zalta. (Summer 2018 Edition). https://plato.stanford.edu/archives/sum2018/entries/skepticism-content-externalism/.

Pessin, Andrew, and Sanford Goldberg. *The Twin Earth Chronicles: Twenty years of reflection on Hilary Putnam's "The Meaning of 'Meaning'."* New York: Routledge, 2015.

Putnam, Hilary. "The meaning of 'meaning'." *Mind, Language and Reality* ed. Hilary Putnam. Cambridge: Cambridge University Press (1975).

Rowlands, Mark, Joe Lau, and Max Deutsch. 2020. "Externalism About the Mind." *The Stanford Encyclopedia of Philosophy* (Winter 2020 Edition), Edited by Edward N. Zalta, https://plato.stanford.edu/archives/win2020/entries/content-externalism/.

Wittgenstein, Ludwig. *Philosophical Investigations.* Oxford: Basil Blackwell Publisher Ltd., 1958.

Contributors

Guillem Adrover Clar

Guillem Adrover Clar is a third year Joint Honours Mathematics and Philosophy student at St. Andrews. He is from Palma de Mallorca, Spain. He decided to pursue both maths and philosophy due to his inability to decide between the two subjects that he loves most. His main philosophical interests are philosophy of mind, logic and metaphysics. He is also interested in historical philosophy, especially that of Plato and Kant. Regarding maths, he is heavily biased towards the pure branch, especially algebra and analysis. In his spare time he loves playing board games, especially chess. He finds it hard to write a biography about himself, be it from his lack of real accomplishments up to now, his shyness or his adherence to Kant's words in the Introduction to the Critique of Pure Reason (widely echoed by most analytical philosophers) "about ourselves we shut up".

Joanne Park

Joanne Park is a junior at Columbia University studying history and philosophy, whose primary interests include meta-ethics, philosophy of law, and the history of philosophy. Outside of philosophy, Joanne is also interested in police reform and children's rights, and is Co-Editor-in-Chief of the Columbia Undergraduate Law Review.

Liana Raguso

Liana Raguso is a third-year undergraduate studying Fundamentals and English at the University of Chicago. She is a frequent texter and a less frequent student of philosophy of language.

Eric Wallace

Eric is a joint honours mathematics and philosophy student at the University of St Andrews. He is primarily interested in fundamental questions, such as the analysis of knowledge, what good philosophical methodology is, the grounding of mathematics, and which inference rules are the correct ones to use. Looking forward, he hopes to carry on the project of developing and vindicating the sensitivity analysis of knowledge.

David Zheng

David Zheng has recently graduated from University of Cambridge, and he is thankful for Dr Matt Dougherty for supervising this paper. His philosophical interests lie in the intersections of metaphysics with music and with law. Otherwise, he enjoys usquebaugh, chess, and motorcycling.