Content Externalism *without* Thought Experiments?

[Forthcoming in *Analysis*. Please cite the published version.]

Abstract: A recent argument against content internalism bucks tradition: it abandons Twin-Earth-style thought experiments and instead claims that internalism is inconsistent with plausible principles relating belief contents and truth values. Call this (for reasons that will become obvious) the *transparency argument*. Here, it is shown that there is a structurally parallel argument against content internalism's foil: content externalism. Preserving the transparency argument while fending off the parallel argument against externalism requires that (1) content-determination and truth-value-determination are implausibly linked together and that (2) eternalism about belief contents is true. Given these requirements, there may be reason to prefer simple, thought-experiment-based arguments against internalism—the sort of arguments that the transparency argument is meant to supersede.

1. Introduction

Content internalism—roughly, the thesis that mental content is fully determined by intrinsic features of a thinking subject—is often believed to fall at the hands of Twin-Earth-style thought experiments. A recent argument against internalism (due to Yli-Vakkuri 2018) abandons these thought experiments. The argument, at a certain level of abstraction, is this: given plausible assumptions, internalism implies that a belief and its content can differ in truth value. But this is absurd: a belief and

its content cannot differ in truth value.[1] Thus, internalism is false. Call this the *transparency argument*. If successful, it would mark welcome progress in a decades-old debate. So, is it successful?

Some think not, arguing that (among other things) the internalist may deny one or more of the transparency argument's premisses (Sawyer 2018; Rieppel 2019; Woodling 2019). Perhaps this is so—I am sympathetic to extant criticisms. However, I want to raise a separate issue. On the one hand, the transparency argument is meant to appeal to the *externalist*—someone who thinks that mental content is determined in part by extrinsic features of a thinking subject. On the other hand, the transparency argument suggests an exactly parallel argument against externalism itself. That is, the externalist is *also* apparently committed to the absurdity that a belief and its content can differ in truth value—and for roughly the same reasons that the internalist is.

If the externalist wishes to retain the transparency argument against internalism and reject the parallel argument against her own view, her options are limited. And, as far as I can tell, she will need to hold that content-determination and truth-value-determination coincide in unexpected ways. Moreover, she will likely have to claim that *eternalism* is true of belief contents—belief contents are never (in a sense to be made precise) temporally neutral. For some, the cost of these commitments may be worth the benefit of the transparency argument against

[1] The transparency argument is also supposed to target the contents of utterances in natural language. I do not address this here since it might be complicated by potential dissimilarities between belief content and linguistic meaning (Pietroski 2020).

internalism. For others, they will be reason to avoid the argument and retreat to the thought experiments that gave rise to externalism in the first place.

2. The Argument Against Internalism

Consider my token belief that tomatoes are red. Most think that the content of this belief is (something like) the proposition that tomatoes are red. What makes it the case that my belief has this content as opposed to some other content or no content at all? If answerable, there are exactly two possibilities. Either my belief's content is fully determined by the way that I am intrinsically or it is not. And if it is not, then it must instead be determined at least in part by my non-intrinsic features (plausibly, in conjunction with my intrinsic features).

The first sort of answer belongs to the *internalist.* She holds that my belief has the content it does purely because of how I am intrinsically. To illustrate, imagine an intrinsic duplicate of me called 'Dup'. If Dup is indeed my intrinsic duplicate, then each part of Dup must correspond to a part of me. If I have a head, Dup has a corresponding head. If I have a hand, Dup has a corresponding hand. Similarly, since beliefs are parts of my mental economy, if I have a belief, Dup has a corresponding belief. The internalist holds that if the content of my belief is that tomatoes are red, then the content of Dup's corresponding belief must also be that tomatoes are red (and vice versa). In fact, if we are intrinsic duplicates, and if belief contents are fully determined by intrinsic features of believers, all of our corresponding beliefs must share their contents. In other words, for each token belief, internalism identifies some feature of the belief that would survive intrinsic

duplication of this sort and then claims that it is this feature that determines the belief's content. Call any such feature an *internal content-determining feature*—or just *I-feature* for short.

Tradition has it that internalism falls at the hands of certain thought experiments (I will review one of them shortly). But Yli-Vakkuri (2018) claims to have found a thought-experiment-free argument against internalism. He begins by identifying a consequence of, or else a claim closely associated with, internalism:

$$\textsc{Narrow}_{C}\text{: } \Box \forall x \forall y(Ixy \rightarrow c(x) = c(y))^{2}$$

The domain of quantification here is restricted to beliefs; '*Ixy*' means that *x* corresponds to *y* in that they possess the same I-feature; and the function '*c*(*x*)' picks out the truth-evaluable content of *x*. Accordingly, we read $\textsc{Narrow}_{C}$ as saying that, necessarily, beliefs with the same I-feature have the same content.

According to Yli-Vakkuri, the falsity of $\textsc{Narrow}_{c}$ follows from just two principles: (1) a belief's truth value is the truth value of its content and (2) corresponding beliefs of intrinsic duplicates may differ in truth value. Letting the function '*v*(*x*)' pick out the truth value of *x,* we formalize each as follows:

$$\textsc{Transparency}\text{: } \Box \forall x\ v(x) = v(c(x))$$

$$\text{I-}\textsc{Difference}\text{: } \neg\Box \forall x \forall y(Ixy \rightarrow v(x) = v(y))^{3}$$

---

[2] We may, as Yli-Vakkuri (2018) does, remain neutral on the species of objective necessity invoked here—*i.e.* whether it is nomological, metaphysical, or something else.

[3] This is the principle others have called '$\textsc{Broad}_{T}$'. Rieppel (2019: 471-3) offers an interesting discussion of Yli-Vakkuri's (2018) defence of this principle.

TRANSPARENCY says that, necessarily, a belief's truth value is identical with the truth value of its content. This seems sufficiently obvious. But I-DIFFERENCE says that it is not necessary that beliefs with the same I-feature have the same truth value, and this is not immediately obvious. Yli-Vakkuri defends this by claiming that "truth is a paradigmatic broad semantic property" (2018: 83-84). A belief's truth depends (or can depend) on the way the external world is. From this, he infers that beliefs with the same I-feature may differ in truth value.

The problem is that NARROW$_C$, TRANSPARENCY, and I-DIFFERENCE are inconsistent in modal logics as weak as K. Yli-Vakkuri (2018: 86, fn. 10) offers a formal proof, but the intuitive idea is simple enough. Suppose a belief's I-feature is sufficient to determine its content (by NARROW$_c$) but that its I-feature is *not* sufficient to determine its truth value (by I-DIFFERENCE). It follows from these two claims that it is possible that beliefs with the same content differ in truth value. But this is impossible—a belief's truth value *is* the truth value of its content (by TRANSPARENCY), and so beliefs with the same content must have the same truth value. Hence, if TRANSPARENCY and I-DIFFERENCE are true, then NARROW$_C$ is false. And since NARROW$_C$ is a consequence of internalism, internalism is false. This is the transparency argument against internalism.

3. A Transparency Argument Against Externalism

The source of the problem for internalism seems clear: if a belief's I-feature is sufficient to determine its content but not its truth value, then it is possible that

beliefs with the same content differ in truth value. Interestingly, a structurally analogous argument also applies to *externalism*.

Why has the analogous argument gone unnoticed? I suspect it is because, for the purposes of the present debate, many have understood externalism as the mere denial of NARROW$_C$. But fleshed out varieties of externalism must do more than issue a negative claim to the effect that NARROW$_C$ is false. Instead, they must offer a positive claim about how the contents of belief are determined. In broad outline, externalists are united by the idea that belief contents are determined by subjects' intrinsic features in conjunction with their extrinsic features. To illustrate, consider the Twin Earth thought experiment (Putnam 1975). It is intuitive that my beliefs about what I call 'water' are beliefs about the chemical substance H_2O. But my duplicate's beliefs on a distant, H_2O-less planet—beliefs about what *he* calls 'water'—are about the chemical substance XYZ. Although we are intrinsically the same, our beliefs differ in content. Intuitively, the difference in content is due to differences in our respective environments—or, more precisely, our being related to distinct environments makes for differences in belief content. This is the core insight of externalism: extrinsic features of subjects determine that their beliefs have the contents that they do (again, in conjunction with certain intrinsic features of subjects). Call any feature that is at least partially extrinsic to a subject and that fully determines the content of her belief an *external content-determining feature*—or just *E-feature* for short. It is this sort of feature, whatever it may be, that determines the content of a belief—at least if externalism is true.

As with internalism, there is a principle either entailed by or closely associated with externalism:

BROAD_C: $\Box \forall x \forall y(Exy \rightarrow c(x) = c(y))$

As before, the domain of quantification is restricted to beliefs, and the two-place predicate '*Exy*' means that *x* and *y* have the same E-feature. The principle thus reads: necessarily, if two beliefs have the same E-feature, then they have the same content. For example, even though my twin on Twin Earth has different belief contents than I do, had our environments *both* contained H_2O and not XYZ, our "water-related" belief contents would have been the same (ignoring, for the moment, indexical contents). Our intrinsically identical internal constitutions conjoined with the fact that we are related to type-identical environments would ensure this.

In short, we have a pair of parallel modal principles: NARROW_C and BROAD_C. The former is closely associated with internalism and the latter with externalism. They differ only in that one principle appeals to I-features and the other appeals to E-features.

BROAD_C is threatened by an argument that parallels the argument against NARROW_C. The argument begins with a relatively simple thought: content-determination and truth-determination are two distinct, and presumably independent, things. Whatever extrinsic features are sufficient for determining the content of a belief, those same features are not always, or even typically, sufficient for determining the belief's truth value. This intuitive idea is reflected in the history of psychosemantics: despite a wide variety of theories, no one (to my knowledge) has advocated a theory where content- and truth-determination necessarily coincide.

Consider, for the sake of illustration, a simple tracking account on which a belief's content is determined by the proposition whose truth it tracks under optimal conditions (Stalnaker 1984: 17-19). And suppose for concreteness that I believe that there is a rabbit in the woods. On a tracking account, my belief tracks the truth of the proposition that *there is a rabbit in the woods*. That is, when I have the belief under optimal conditions, it is true that there is a rabbit in the woods, and so my belief is true. But another situation is also possible: I fail to be in optimal conditions, and my belief is false. Unbeknownst to me, all trees have been burned to the ground and rabbits have gone extinct. So, my beliefs in each scenario, though they track the same proposition, do not have the same truth value.

The idea generalizes. Content-determination and truth-determination are independent affairs. So, even if a partly extrinsic property determines belief content, it is nonetheless possible for a belief to have that extrinsic property in situations where it is true and situations where it is false. Or, slightly more accurately, for any property that is a plausible candidate for an E-feature, it is possible that beliefs x and y possess that E-feature and yet differ in truth value.

E-DIFFERENCE: $\neg\Box\forall x\forall y(Exy \rightarrow v(x) = v(y))$

We are now in a familiar situation. For the same reason that NARROW_C, I-DIFFERENCE, and TRANSPARENCY are inconsistent, BROAD_C, E-DIFFERENCE, and TRANSPARENCY are also inconsistent. And if we grant E-DIFFERENCE and TRANSPARENCY, then we must reject BROAD_C, and with it externalism.

Now, Yli-Vakkuri points out that the transparency argument against internalism is "not, of course, psychologically impossible to resist—no philosophical argument

is. A sufficiently dedicated internalist will find a way to resist it" (2018: 86-87). Likewise, the parallel argument against externalism can be resisted by a sufficiently dedicated externalist. What strategies might she employ? That depends. If she wishes to retain the transparency argument against internalism, her options are few.

To begin, two options will not do. First, the externalist cannot deny TRANSPARENCY, for then she loses the transparency argument against internalism—which I am assuming she wishes to retain. Second, the externalist should not deny BROAD_C. For if she denies that her view has the modal consequences codified by BROAD_C, then the internalist may reasonably deny that her view has the modal consequences codified by NARROW_C. For example, the externalist could hold that belief content is grounded in, but not necessitated by E-features (see Schaffer 2010 for a discussion of grounding and necessitation). But if she does this, the internalist could make a parallel move and say that belief content is grounded in, but not necessitated by I-features.

The only plausible option is for the externalist to find some way of resisting E-DIFFERENCE. Now, E-DIFFERENCE is *prima facie* plausible. As I have suggested, content-determination and truth-determination are independent affairs. That is, the properties of a belief that determine its content are distinct from the properties that determine its truth value. A belief is true in virtue of the fact that its content is true. But it does not have its content in virtue of the fact that its content is true. Denying E-DIFFERENCE does not sit comfortably with this. Its denial is equivalent to the claim that, necessarily, beliefs with the same E-feature have the same truth value (at a world). There is thus a necessary connection between content-determination and

truth-value. But given that a belief's content and its truth value have different determinants, this necessary connection is *prima facie* puzzling, and perhaps even undesirable.

Moreover, even if there is a way to render this commitment less puzzling, there is a further commitment that one must take on in denying E-DIFFERENCE. Specifically, denying E-DIFFERENCE requires *eternalism* about belief contents—*i.e.* the position that a belief's content cannot change truth value over time at a world. The reason is that if eternalism is false, a case that supports E-DIFFERENCE is relatively easy to construct. To illustrate, we can consider a single, token belief evaluated at two different times relative to a world *w*. For concreteness, assume the belief has the temporally neutral content *there is sriracha in the fridge* and that this content is determined by some E-feature of the belief. In *w*, sometimes there is sriracha in the fridge and sometimes there is not. The content of the belief, and thus the belief itself, varies in truth value at *w* depending on the time of evaluation. So its truth value changes but its E-feature does not. Hence, beliefs with the same E-feature need not have the same truth-value—that is, E-DIFFERENCE is true. To deny E-DIFFERENCE, there can be no case like this whatsoever—cases of this sort must be impossible. And this seems to require that no contents are temporally neutral and, accordingly, eternalism.

Of course, eternalism is not problematic in itself. There is precedent for various forms of the view (*e.g.* Moore 1962, Richard 1981, and Salmon 1986). Equally, however, there is precedent for the denial of eternalism or *temporalism* (*e.g.* Prior

1959, Kaplan 1989, and Brogaard 2013).[4] Some externalists will not mind committing to eternalism. But those who accept temporalism will need a different strategy. And those who, like myself, find the eternalist v. temporalist debate frustratingly subtle should be wary. For it is plausible that "coverage of the data will be exactly the same for each [view]" (Dever 2015: 2), making the choice between eternalism and temporalism especially difficult.[5]

The overall point, then, is that if the externalist tries to retain the transparency argument against internalism while rejecting a parallel argument against externalism, then she must (1) reject E-DIFFERENCE and plausibly (2) accept eternalism about belief contents. This seems risky. Denying E-DIFFERENCE implausibly links together content-determination and truth-value. And eternalism, though not itself implausible, is a substantial commitment to make in arguing for content externalism. Thankfully, the transparency argument is not mandatory. There are other, more stable arguments for externalism—albeit ones that still rely on thought experiments.[6]

Jonathan Brink Morgan

Montclair State University, USA

morganj@montclair.edu

[4] To be clear, this debate is typically framed as a debate about contents *in general* and not just about the contents of belief.

[5] The issue concerning eternalism also raises questions about whether truth is fundamentally monadic or relational. See Cappellen and Hawthorne 2009 and MacFarlane 2014.

[6] Many thanks to two anonymous referees for generous feedback that improved this paper significantly. Thanks also to Chelsey Deisher for reading multiple drafts.

References

Brogaard, B. 2012. *Transient Truths: an essay in the metaphysics of propositions*. Oxford: Oxford University Press.

Cappellen, H. and J. Hawthorne. 2009. *Relativism and Monadic Truth.* Oxford: Oxford University Press.

Dever, J. 2015. Eternalism, temporalism, neutralism. *Inquiry* 58: 608-618.

Kaplan, D. 1989. Demonstratives. In Themes from Kaplan, eds. Joseph Almog, John Perry and Howard Wettstein, 481-563. New York: Oxford University Press.

MacFarlane, J. 2014. *Assessment Sensitivity: relative truth and its applications.* Oxford: Oxford University Press.

Moore, G. E. 1962. Facts and propositions. In his *Philosophical Papers,* 60–88. New York: Collier Books.

Richard, M. 1981. Temporalism and eternalism. *Philosophical Studies 39*: 1–13.

Salmon, N. U. 1986. *Frege's Puzzle*. London, England: MIT Press

Schaffer, J. 2010. The least discerning and most promiscuous truthmaker. *Philosophical Quarterly* 60: 309–24.

Pietroski, P.M. 2020. A narrow path from meanings to contents. *Philosophical Studie*s.

Prior, A. N. 1959. Thank goodness that's over. *Philosophy 34*: 12–17.

Putnam, H. 1975. The meaning of 'meaning'. *Minnesota Studies in the Philosophy of Science* 7: 131–93.

Rieppel, M. 2019. Broad properties of beliefs. *Analysis* 79: 470-476.

Sawyer, S. 2018. Is there a deductive argument for semantic externalism? Reply to Yli-Vakkuri. *Analysis*, 78(4), 675-681.

Stalnaker, R. 1984. *Inquiry.* Cambridge, MA: MIT Press..

Woodling, C. 2019. Content externalism, truth conditions, and truth values. *Philosophia* 48: 821-830

Yli-Vakkuri, J. 2018. Semantic externalism without thought experiments. *Analysis* 78: 81-90.

Yli-Vakkuri J. and J. Hawthorne. 2018. *Narrow Content.* Oxford: Oxford University Press.