

Against Metasemantics-First Moral Epistemology

Jesse Hambly & Shang Long Yeo

(Forthcoming in *The Journal of Ethics*, please cite published version)

Abstract: Moral metasemantic theories explain how our moral thought and talk are about certain properties. Given the connection between what our moral terms are about and which moral claims are true, it might be thought that metasemantic theorising can justify first-order ethical conclusions, thus providing a novel way of doing moral epistemology. In this paper, we spell out one kind of argument from metasemantic theories to normative ethical conclusions, and argue that it fails to transmit justification from premises to conclusion. We give three reasons for this transmission failure, which together pose a serious challenge to such metasemantic arguments.

Moral metasemantic theories tell us in virtue of what our moral thought and talk pick out a property.¹ Because of this foundational role in fixing what our moral terms are about, metasemantic theories seem to yield a distinctive way of doing normative ethics: we simply argue from the metasemantic theory, plus the premise that a certain descriptive property fulfils the theory's reference-determining conditions, to the conclusion that this descriptive property co-occurs with a moral property. In this paper, we spell out and critically examine this way of theorising – which we call *metasemantics-first moral epistemology*. We argue that such theorising faces serious challenges.

¹ Metasemantic theories tell us about (1) what makes it the case that *terms* or *sentences* are about a certain property/object, or (2) what makes it the case that *concepts* or *beliefs* (psychological representations) are about a property/object. Our discussion is largely framed around (1) but everything we say could be put in terms of (2). We will remain neutral in this paper between referentialist (Braun 2016; Edwards 2014) and non-referentialist (Chalmers 2011; Segal 2000) views of content, which differ on whether the content of a term or concept is exhausted by its referent.

Our investigation of metasemantics-first moral epistemology is important for several reasons. First, it illuminates moral metasemantics by making explicit a line of argument that yields the surprising implication that metasemantic theorising can determine which properties are morally relevant. Second, it furthers our understanding of moral epistemology by giving reasons why this surprising implication fails to hold. Third, it sheds light on the relationship between metasemantics and moral epistemology by exploring how the two types of theorising might interact. The structure of this paper is as follows: In Section (1) we illustrate how moral metasemantic theories can imply normative ethical conclusions, using Richard Boyd's causal theory of reference for moral terms as an example. Section (2) contends that the metasemantic argument from Boyd's theory does not provide sufficient evidence for its normative ethical conclusion, because the argument fails to transmit justification from premises to conclusion. We give three different reasons for this transmission failure, which together pose a serious challenge to the metasemantic argument. Section (3) explores whether the reasons for transmission failure apply to arguments from other metasemantic theories, including paradigmatically internalist ones.² In Section (4) we deal with objections. Section (5) concludes.

1 | The Boydian Metasemantic Argument

The truth conditions of a moral sentence are determined by what the sentence is about. This opens the possibility of using moral metasemantics to derive normative ethical conclusions. We first illustrate this possibility using Richard Boyd's (1988) metasemantic theory, which claims that the referent of a moral term is whatever causally regulates our use of that term in the appropriate way. Specifically, Boyd suggests that "a term *t* refers to a kind (property,

² Internalist metasemantic theories hold that semantic content supervenes on the internal features of agents. Externalist metasemantic theories deny this supervenience thesis (Farkas 2008: 362–68; cf. Wikforss 2008: 161). Some theorists hold that there is both internalist and externalist content. See Chalmers (2003) and Jackson (2003). Views which recognize both kinds of content are usually classed as internalist (Brown 2016: sec. 3.2; Segal 2009: 369).

relation, etc.) k just in case there exist causal mechanisms whose tendency is to bring it about, over time, that what is predicated of the term t will be approximately true of k " (1988: 195). What this means is that, if term t refers to property k , it must be the case that our t beliefs (our beliefs predicating t) have better tracked which things have k over time, as a result of our interactions with k .³

Suppose that what causally regulates our use of 'is right' in the appropriate way is some descriptive property d (such as the property of maximising net desire-satisfaction, or of practical universalisability).⁴ We can then make the following *Boydian Metasemantic Argument*:

(*Boyd's Metasemantic Theory*) The sentence 'x is right' attributes the property which causally regulates our use of 'is right' in the appropriate way.

(*Boydian Metasemantic Grounds*) Property d causally regulates our use of 'is right' in the appropriate way.

Therefore,

(C1) The sentence 'x is right' attributes property d . (From Boyd's Metasemantic Theory and Metasemantic Grounds.)

(*Aboutness and Truth*) If the sentence 'x is right' attributes property d , then 'x is right' is true iff x is d .

³ To make this a little more concrete, consider Boyd's own suggestion about the referent of 'good'. Boyd suggests that the term 'good' might be causally regulated in the appropriate way by a 'homeostatic property cluster' consisting of things which satisfy various human needs like love, friendship, health, control over one's own life, physical recreation, intellectual and artistic appreciation, etc (1988: 203–5). Boyd tells us that whether this proposal succeeds depends on whether it is "plausible that the homeostatic cluster of fundamental human goods has, to a significant extent, regulated the use of the term 'good' so that there is a general tendency... for what we say about the good to be true of that cluster". He contends that "we can observe [this tendency] in the way in which our concept of the good has changed in the light of new evidence concerning human needs and potential" (Boyd 1988: 210).

⁴ Note that d may be a conjunctive, disjunctive, or higher-order property.

Therefore,

(C2) The sentence ‘ x is right’ is true iff x is d . (From C1 and Aboutness and Truth.)

(*Equivalence Schema*) The sentence ‘ p ’ is true iff p .

Therefore,

(C3) x is right iff x is d . (From C2 and the Equivalence Schema.)

Boyd’s Metasemantic Theory, combined with a premise about which property satisfies its reference-determining conditions, gives an intermediate conclusion about which property is attributed by moral sentences. The argument then deploys two more uncontroversial assumptions: Aboutness and Truth is an instantiation of the principle that if a sentence attributes a property to some object, then that sentence is true iff the object in fact has that property.⁵ (A non-moral example: if ‘Tibbles is asleep’ attributes the property of being asleep, then it is true iff its object, Tibbles, is in fact asleep.) The Equivalence Schema is similarly uncontroversial – deflationists about truth and their opponents agree that this Schema tells us something about the nature of truth, even though they disagree about whether it exhausts the nature of truth (Armour-Garb, Stoljar, and Woodbridge 2022: sec. 1.1). The argument gives the conclusion (C3), which is the sort of biconditional that normative ethical theories are in the business of delivering – it tells us which descriptive properties must obtain in order for something to be right.⁶

⁵ Dickie (2016: 102–4) proposes a similar principle ranging over objects and beliefs.

⁶ Nothing in this metasemantic argument is special to ethics – we could, in principle, attempt a metasemantics-to-substantive argument of this kind in any domain. However, some of our later replies turn on considerations that might be idiosyncratic to ethics – for this reason, we focus on ethics, but acknowledge that some of what we say might apply to other domains too. Thanks here to an anonymous referee for suggesting the potential generality of our arguments.

2 | Why the Boydian Metasemantic Argument is Weak Evidence

We believe that the Boydian Metasemantic Argument is weak evidence for its normative ethical conclusion in the sense that the argument does not provide *sufficient* evidence for its conclusion – metasemantic considerations alone don't justify ethical claims about what is right. In this section, we give several reasons in support of this claim. We focus on the transmission of justification from premises to conclusion – which happens when one's (propositional) justification for the conclusion of an argument derives from one's justification for its premises. When an argument transmits justification in this way, we can use that argument to learn something new – by acquiring or strengthening justification for its conclusion (Moretti and Piazza 2023). By contrast, when an argument cannot transmit justification, it cannot give us new (or more) justification for its conclusion, and hence is methodologically inert. In what follows, we argue that the Boydian Metasemantic Argument cannot transmit justification for various reasons – due to its simultaneous generation of a defeater, due to its information dependence, and due to its indirectness.⁷

2.1 | Simultaneous Generation of a Defeater

One way an argument can fail to transmit justification is because of defeaters. To illustrate, suppose you use deductive reasoning to derive some mathematical theorem as a conclusion. However, suppose you also acquire justification that someone spiked your coffee with a drug that's known to cause errors in reasoning. Having this defeater would prevent your deductive reasoning from transmitting justification to its conclusion, or at least reduce the amount of

⁷ It might be argued instead that the Boydian Metasemantic Argument fails because we just aren't justified in believing Boyd's Metasemantic Theory – perhaps because the epistemic standing of metasemantics is generally quite poor (Cappelen 2018: 72–74). This undermines the metasemantic argument as it currently stands, but it leaves open the possibility of the argument working in principle, once we have acquired better justification for our metasemantic theories. In contrast, we prefer our challenges of transmission failure because they target the very structure of the metasemantic argument, and would work even when we have good justification for Boyd's theory. Thanks here to an anonymous referee.

justification transmitted (Tucker 2010: sec. 3a). We'll suggest that any metasemantic theorising that justifies Boyd's Metasemantic Theory (a premise of the argument) will simultaneously also generate a defeater for the very pattern of reasoning involved in the Boydian Metasemantic Argument.

Moral metasemantic theorising doesn't just take a subject's first-order, object-level judgments as inputs. It takes into account all of a subject's commitments about the moral domain – this includes higher-order dispositions to see their own first-order judgments as fallible (Schroeter and Schroeter 2014: 14), and to revise their first-order judgments in various circumstances, such as when they receive certain kinds of empirical information, when they hear testimony or debate their views with others, or when they conduct theoretical moral reflection (Schroeter, Schroeter, and Toh 2022: 188). Such dispositions reflect a subject's epistemic commitments regarding the domain, and where possible, should be respected when generating a metasemantic theory.⁸

We contend that one such higher-order disposition with respect to the moral domain is a disposition not to accept that some descriptive property makes a moral difference without having *checked* the moral relevance of that property. (When actions are (say) right, it is because they have certain descriptive properties, e.g., being an instance of helping someone in need.) To check the moral relevance of a descriptive property *d* is to try to ascertain whether *d* is, for instance, a right-making property. This check is commonly performed in several ways: 1) by forming first-order moral judgments about which actions are right, and then looking to see if the right actions have *d*, 2) by forming a direct judgement about whether *d* is morally relevant (this happens when we judge, for instance, that an action's effect on overall happiness is a morally relevant property – though perhaps not the only one),⁹ or 3) by forming moral

⁸ Schroeter and Schroeter (2014: 14) goes as far as to claim that “In effect, we can build our semantic determination theory from the first person reflective epistemology of the topic in question.”

⁹ Nye (2015) argues that ethical theorising should directly evaluate the ethical relevance of descriptive properties, and only use cases to clarify and illustrate claims about such properties. Many contemporary debunking

judgments at various levels of generality and pursuing reflective equilibrium with them (Rawls 1974, 1999), which delivers an output about whether *d* is morally relevant. We make no claim that the methods of checking we've identified in fact confer justification on our moral beliefs. Rather, our suggestion is just that these methods of checking are central to our ordinary moral practices.

This commitment to checking the relevance of descriptive properties emerges from the fact that when we say that an action is right, we represent it as having a normatively significant evaluative status – of being choiceworthy and being worthy of approval and admiration by others (FitzPatrick 2018: 552–53). Attributing this status to an action is distinct from merely attributing a descriptive property that makes the action right. For instance, judging that some action maximises happiness is distinct from evaluating that action to be right – even if all and only those things which are right maximise happiness.

Notice that the Boydian Metasemantic Argument doesn't incorporate any of the possible ways of checking for moral relevance. Consequently, the argument clashes with the higher-order epistemic commitment to checking, which generates a defeater for the argument.¹⁰ We believe this defeater generates a strong presumption, stemming from factors internal to metasemantic theorising, against recognising the Boydian Metasemantic Argument as a legitimate way of arriving at normative ethical conclusions. To strengthen our case, we highlight two features of this defeater. First, the defeater isn't grounded in abstract theoretical

arguments appear to rely on direct judgements about the ethical (ir)relevance of certain descriptive properties. For discussion see Konigs (2020) and Huemer (2008: 372–74).

¹⁰ Boyd's theory does impose an epistemic constraint on the assignment of a referent to moral terms, since it dictates that the referent of a term be assigned such that subjects' beliefs about the term become truer over time (van Roojen 2006: 170). However, this constraint does not involve checking as we've explained the idea here. We expect that some philosophers will dispute our interpretation of Boyd's view. For instance, Väyrynen (2019: 208) appears to suggest that Boyd's metasemantic theory might include something like a check for moral relevance in its reference-determining conditions. We disagree with this interpretation of Boyd, but in any case there exist moral metasemantic theories that clearly don't do this – for instance, the theories which result when we apply Kripke's (1980) metasemantics for names or Fodor's (1994b) asymmetric dependence theory to ethical terms. Thanks to an anonymous referee for pushing us to clarify this issue.

considerations, or in an external theorist's intuitions about the subject matter – instead, it draws directly on the subject's own epistemic commitments. Secondly, the disposition that grounds this defeater seems to be rather forceful and entrenched.¹¹¹²

2.2 | Information Dependence

Another kind of transmission failure occurs due to dependence on background information: when some evidence justifies a proposition *P* only when combined with independent justification for some background information, but this background information is itself also entailed by *P* (perhaps along with other premises) (Wright 2003: 58–59). In this case, the justification from the evidence for *P* will fail to transmit to the background information. To illustrate, consider the following example. Suppose we're visiting a zoo – we look into a pen and acquire evidence that the animals within look like zebras:

(Zebra Appearance) The animals in the pen look like zebras.

This evidence seems to justify:

(Actual Zebras) The animals in the pen are actually zebras.

¹¹ Our discussion has assumed that this defeater is a high-order defeater for the Boydian Metasemantic Argument. This is our official position, but we could alternatively hold that the defeater is an undercutting defeater for the Boydian Metasemantic Theory. That is to say, we might hold that the fact that the Boydian Metasemantic Theory licences the Boydian Metasemantic Argument counts against believing the theory. The defeater would still threaten the Boydian Metasemantic Argument because arguments with premises which are unjustified obviously fail to transmit justification to their conclusions.

¹² It's instructive to contrast our argument here with a recent discussion of Boyd's Metasemantic Theory in Zhao (2021). Zhao suggests that the primary problem with Boyd's view is that it delivers referents for our moral terms which are inconsistent with a 'reference defeater' – where, roughly speaking, a reference defeater is a non-negotiable commitment about the features of the referent of a term *t*. (An example Zhao gives in the case of the term 'witch' is the belief that witches have supernatural powers.) The reference defeater that Zhao identifies in the case of moral terms is the belief that moral properties are not identical to descriptive properties (Zhao 2021: 1174–75). Our focus is on the commitments of ordinary speakers with respect to moral epistemology rather than their commitments with respect to moral metaphysics. Also, we don't take the epistemic commitment that we've identified to conclusively defeat Boyd's Metasemantic Theory.

But from this, we can also infer that:

(No Illusion) The animals in the pen are not mules cleverly painted to look like zebras.

Something seems to have gone wrong. We shouldn't be able to get from Zebra Appearance to No Illusion. One prominent diagnosis is that the justification from Zebra Appearance to Actual Zebras fails to transmit to No Illusion, due to information dependence of the kind outlined earlier: we can acquire justification for Actual Zebras on the basis of Zebra Appearance only if we have independent justification for No Illusion (the background information) (Moretti and Piazza 2023: sec. 3.2; Wright 2003). When we reason from Zebra Appearance to No Illusion, our reasoning *presupposes* independent justification for No Illusion, but it does not *establish* such justification. Hence, the argument fails to transmit justification to No Illusion.

We believe that a similar structure – and a similar failure of transmission – obtains in the Boydian Metasemantic Argument. To show this, first recall that one of the argument's premises is Boyd's Metasemantic Theory. One important way of justifying such a moral metasemantic theory is by appealing to its extensional adequacy – that is, by showing that the theory delivers extensional implications that are deemed credible by our first-order moral judgments.

This desideratum of extensional adequacy can be justified on several grounds. First, it follows from the function of metasemantic theorising – which typically takes certain semantic facts for granted and seeks to explain why these facts obtain (Burgess and Sherman 2014: 3; Lewerentz and Marschall 2018: 1670–72). Such semantic facts are given by our first-order moral judgments – for instance, I might judge that donating money to charity is morally right; the corresponding semantic fact, then, is that my judgement attributes the property of moral rightness to the act of donating.¹³ Seeking to explain these semantic facts just is to endorse

¹³ We are primarily concerned with first-order moral judgments like these, which use terms like 'is right' and are hence more akin to ordinary speakers' use of expressions. We leave open the possibility that extensional fit with other kinds of evidence is not required – for instance, our extensional adequacy desideratum is compatible with Dowell's (2016) conclusion that higher-order semantic intuitions (for instance about whether we genuinely

extensional adequacy as a desideratum. Secondly and relatedly, the practice of metasemantic theorising is replete with concrete examples of treating certain semantic facts as given and then assessing whether metasemantic theories deliver the correct reference assignment. For instance, in the debate over the metasemantics of proper names, the referent of a name like ‘Madagascar’ is taken for granted and used a datum which needs to be accommodated (Dickie 2011; Evans 1973: 195–96). Similarly, when theorists discuss the disjunction problem for causal theories of mental content, they assume that the mental representation *dog* refers to dogs not to dogs or cats-on-a-dark-night (Fodor 1994a: 59). Examples like this can be multiplied. Thirdly, extensional adequacy is desirable on grounds of interpretive charity. As far as possible, we want to charitably interpret object-level judgments as true, because the more a metasemantic theory makes these judgments come out as true, the more the theory can fulfill its theoretical roles of predicting and explaining behaviour, and of providing an appropriate normative standard for thought and talk (L. Schroeter and Schroeter 2019: 198–99).¹⁴

If extensional adequacy is a desideratum, then a moral metasemantic theory is justified to the extent that it delivers extensional implications that are deemed credible by certain first-order moral judgments, which were taken as given – that is, assumed to be true – in the context of metasemantic theorising.¹⁵ We thus infer from facts about our judgments:

(Our Judgments) We make first-order moral judgments about what is right, M1, ..., Mn.

disagree with a hypothetical community that speaks a different language) have no probative value in metasemantic theorising. Thanks here to two anonymous referees.

¹⁴ Charity can be interpreted in a variety of strengths: for instance as saying that an interpreter should maximise the number of truths a subject believes (Quine 1960), or as saying that the interpreter should make some (possibly very minimal) part of a subject’s total understanding come out true (L. Schroeter and Schroeter 2019: 198–99). Proponents of charity agree, however, that it is better for a metasemantic theory to assign referents that make our object-level judgments come out true, other things being equal. Charity features prominently in Turner’s (2013) metasemantic argument for free will, a non-moral example of metasemantics-first epistemology.

¹⁵ To motivate the extensional adequacy desideratum, we haven’t assumed that we have first-order beliefs which are justified independently of metasemantic theorising. However, if this assumption is made, it would supply a separate and straightforward justification for extensional adequacy.

plus an assumption about the truth of some of them:

(Truth) Our first-order moral judgments M1, M2, M3 (for instance) are true.

to justify a metasemantic theory:

(Boyd's Metasemantic Theory) The sentence '*x* is right' attributes the property which causally regulates our use of 'is right' in the appropriate way.

Notice that Truth is indispensable for theorising of this kind, which supports a metasemantic theory on grounds of its extensional fit. Our Judgments alone cannot justify Boyd's Metasemantic Theory, because a metasemantic theory concerns when certain sentences are *about certain properties*. Our Judgments only tells us what our moral sentences are like, but doesn't contain any information about the actual distribution of properties that these sentences are putatively about. For information of this latter kind, we need an assumption like Truth, which is presupposed – but not established – when we endorse extensional adequacy as a desideratum.

We now have the ingredients to diagnose transmission failure. In metasemantic theorising, we presuppose Truth (the claim that certain first-order moral judgments of ours are true) so that Boyd's Metasemantic Theory can be justified on the basis of Our Judgments (the claim that we make certain first-order moral judgments). But in the Boydian Metasemantic Argument, we then use Boyd's Metasemantic Theory to infer the normative ethical claim (C3). (C3) includes, among its entailments, moral claims whose contents overlap with those of Truth, which was presupposed in the process of justifying Boyd's theory. Just like in the zebras example, then, the entire chain of reasoning merely *presupposes* independent justification for the moral claims in Truth and (C3), but it does not *establish* such justification. Thus when

Boyd's Metasemantic Theory is justified on grounds of extensional adequacy, it fails to transmit justification through the metasemantic argument to (C3).¹⁶

How might we supply independent justification to Boyd's Metasemantic Theory? We might use the checking process mentioned in the previous section. When we check the moral relevance of a descriptive property (like maximal net desire-satisfaction), we evaluate its normative significance, thus potentially supplying justification to claims about actions with this property (for instance, the claim that a donation which maximises desire satisfaction is morally right). These claims can then be used to independently justify Boyd's Metasemantic Theory. The process of checking doesn't presuppose certain moral claims – as is typically done in metasemantic theorising – but rather aims to establish justification for them.¹⁷

We see the challenge from information dependence as working in tandem with the defeater-based challenge from the previous section. To the extent that we respect subjects' higher-order epistemic commitments when doing metasemantic theorising, the Boydian Metasemantic Argument simultaneously generates a defeater that prevents it from transmitting justification. On the other hand, to the extent that we respect subjects' first-order commitments, the Boydian Metasemantic Argument runs into the problem of information dependence, which also leads to a failure to transmit justification. Therefore, regardless of whether we respect first-order or higher-order commitments in metasemantic theorising, the Boydian Metasemantic Argument fails to transmit justification.

¹⁶ Could the metasemantic argument still transmit justification to those entailments of (C3) that do not overlap with the claims in Truth? We think the prospects of this objection are weak, so long as metasemantic theorising presupposes the truth of moral claims, rather than supplying any independent justification for them. Moreover, even if this objection works, our argument will still have reduced the amount of justification transmitted – justification still does not transmit to the entailments of C3 that *do* overlap with Truth.

¹⁷ Note that the criticism from information dependence applies only to a specific class of metasemantic arguments – namely those where the metasemantic theory is justified by appeal to extensional adequacy. This criticism might not apply when the theory is justified in a different way. This specificity is recognised in the transmission failure literature – Pryor (2001: n. 5) points out that in the zebra case, if we justified Actual Zebras on the basis of how the animal *sounds* (rather than what it looks like), then this justification could transmit to No Illusion.

2.3 | Indirectness

But suppose the problems with defeat and information dependence could somehow be avoided. In this section, we pose a further obstacle to the Boydian Metasemantic Argument's ability to transmit justification, stemming from (what's known as) indirectness. The idea is here that sometimes evidence justifies a proposition Q directly, independent of any further argument. If this were so, an argument that proceeds from the evidence through some intermediate proposition P, and then to Q (supposing we know P implies Q) would fail to transmit. The evidence directly justifies Q without having to go through the inferential relation from P to Q, so no justification for Q is based on this inferential relation (Moretti and Piazza 2023: sec. 3.2; Wright 2002). One example of such an indirect argument is the following:

(White Posts) Jones just kicked the ball between the white posts.

(Goal) Jones just scored a goal.

(Soccer) A game of soccer is taking place.

White Posts justifies Soccer directly, independent of the inferential relation from Goal to Soccer. To see this, imagine a scenario where White Posts didn't justify Goal – for instance, because a referee noticed that Jones committed an infringement that nullifies the goal – we would still have the same justification for Soccer from White Posts, without having any justification for Goal (Wright 2002: 334–35). The argument thus takes an unnecessary detour through Goal, and no justification for Soccer is derived from Goal – all its justification comes directly from White Posts (Davies 2009: 372–74; Moretti and Piazza 2023: sec. 3.2).¹⁸

We believe the Boydian Metasemantic Argument is similarly indirect – at least if one supports Boyd's Metasemantic Theory on the grounds of its extensional fit with certain first-

¹⁸ Wright (2002: 334–35) proposes, quite plausibly, that White Posts counts as evidence simultaneously for both Goal and Soccer, but not because it is a justification for Goal which then transmits through to Soccer.

order moral judgments. Suppose one supports Boyd's theory in this way and then uses the theory in a metasemantic argument for the normative ethical claim (C3). Notice, however, that the entire argument takes an unnecessary detour through Boyd's Metasemantic Theory: supposing that our first-order moral judgments didn't support Boyd's Metasemantic Theory – for instance because we adopted the radical view that a subject's object-level judgments don't count as input to metasemantic theorising at all – these first-order moral judgments will still justify (C3) to the same extent. This suggests that no part of the justification for (C3) is in fact derived from Boyd's Metasemantic Theory and the Boydian Metasemantic Argument – all its justification comes directly from the first-order moral judgments that were used to justify Boyd's theory in the first place. The Boydian Metasemantic Argument is indirect in the same way as the soccer argument, and thus fails to transmit justification.

Cumulatively, these three considerations – from a simultaneously generated defeater, from information dependence, and from indirectness – pose a strong challenge to the Boydian Metasemantic Argument's ability to transmit justification. Importantly, they stem from argument templates which are widely taken to be sufficient for transmission failure (Moretti and Piazza 2023: sec. 3.2; Tucker 2010: sec. 3a), and which were developed independently of our application to the Boydian Metasemantic Argument – thus strengthening the overall plausibility of our case.

3 | Extensions

In this section, we consider whether and how our three considerations extend to other instances of metasemantics-first moral epistemology – specifically to metasemantic arguments that rely on different metasemantic theories. To start with, let's generalise the Boydian Metasemantic Argument. Take any moral metasemantic theory which says that the sentence 'x is right' attributes a property if that property satisfies certain reference-determining conditions. And

suppose we know that descriptive property d satisfies these conditions. We can then replace the first two premises of the Boydian Metasemantic Argument with:

(*Metasemantic Theory*) The sentence ‘ x is right’ attributes the property which satisfies reference-determining conditions C_1 - C_n .

(*Metasemantic Grounds*) Property d satisfies reference-determining conditions C_1 - C_n .

Keeping the other premises fixed, this gives us:

(C3) x is right iff x is d .

Call the resulting schema the *metasemantic argument schema*, and its instances, *metasemantic arguments*. We now consider various representative metasemantic theories that could be substituted in for Metasemantic Theory, and whether the problems with simultaneous defeat, information dependence, and indirectness apply.

3.1 | Moral Functionalism

Consider first Jackson and Pettit’s (1995) Moral Functionalism, which specifies the reference-determining conditions in terms of the rightness role in folk morality:

(*Moral Functionalism*) The sentence ‘ x is right’ attributes the property which plays the rightness role in folk moral theory.¹⁹

This rightness role is determined by the folk’s inferential and judgmental dispositions with respect to ‘right’. For instance, the folk readily apply ‘right’ to certain paradigmatically right

¹⁹ It may be inaccurate to describe Moral Functionalism as a metasemantic view (although cf. Schroeter and Schroeter (2018: 524) and Werner (2020: 147)). Jackson and Pettit themselves characterise the position as a view about the meaning of moral terms (Jackson 1998: 131; Jackson and Pettit 1995: 22). What matters for our purposes is that Jackson and Pettit’s theory allows one to construct the sort of argument from the theory plus the reference facts to a first-order ethical conclusion sketched in Section (2). Instead of calling our paper ‘Against Metasemantics-First Moral Epistemology’ we could have equally titled it ‘Against Reference Determination Theory-First Moral Epistemology’.

actions, such as the act of saving a life costlessly. The folk also distinguish things with respect to rightness only if these things display some descriptive difference, and if the folk take an action to be right, they will prefer that action to its alternatives, other things being equal. These “commonplaces” about rightness specify the rightness role in folk moral theory. The referent of ‘is right’ is whatever property that makes these commonplaces come out true (Jackson and Pettit 1995: 22–29).

Suppose one learns that descriptive property *d* plays the rightness role: the actions which the folk take to be paradigm instances of rightness are also *d*, and *d* is consistent with folk dispositions about moral supervenience and moral motivation, and so on. One can then construct a metasemantic argument from Moral Functionalism to the conclusion that *x* is right iff *x* is *d*.

Is the metasemantic argument from Moral Functionalism vulnerable to the three challenges to the Boydian Metasemantic Argument we identified in Section (2)? Take, first, the simultaneous generation of a defeater. It doesn't appear that Moral Functionalism requires checking the moral relevance of property *d* – the theory allows you to learn about the referent of ‘is right’ simply by learning facts about folk usage.²⁰ Therefore, the metasemantic argument from Moral Functionalism similarly is vulnerable to the checking defeater we raised for the Boydian Metasemantic Argument.

However, our case must be tempered by the recognition that there are different versions of Moral Functionalism. The version just considered says that the referent of ‘is right’ is whatever plays the rightness-role in *current* folk morality. It takes the commonplaces to be fixed by the folk’s actual moral judgments – consequently, it supports a metasemantic argument that clearly doesn’t involve a check for moral relevance, and is problematic in the way detailed

²⁰ It might be objected that you are among the folk, so you bring to bear your own judgments about what falls into the extension of ‘right’ when inquiring into the reference of ‘right’ – and that might count as a check. However, the sense in which you are bringing to bear such judgements is that you are taking as an input the fact that some agent (i.e., you) thinks something is right, rather than directly considering what is right.

above. On the other hand, consider a version of Moral Functionalism that says that the referent is what plays the rightness-role in *mature* folk morality, which is arrived at after subjecting current folk morality to debate and critical reflection (Jackson 1998: 133–34).²¹ Such debate and reflection might involve, for instance, thinking about which actions are in fact paradigmatic instances of right action, rather than actions merely taken to be so by the folk.²² A metasemantic argument from this version of Moral Functionalism may avoid the checking defeater.

Similar issues play out with information dependence. If we're thinking of a Moral Functionalism that uses current folk morality, then the relevant metasemantic argument merely presupposes that actual folk moral judgments are right, but does not establish this – thus it fails to transmit justification to its conclusion. By contrast, if Moral Functionalism assigns referents based on functional roles obtained from subjecting current folk morality to critical debate and reflection, then the corresponding metasemantic argument may have established justification for moral claims, thus avoiding the problem with information dependence.

The problem with indirectness, however, still plagues metasemantic arguments from Moral Functionalism. For suppose that our first-order moral judgments didn't support Moral Functionalism for whatever reason²³ – these judgments would still support a normative ethical claim like (C3), and plausibly to the same extent as before. This reveals that such metasemantic arguments take an unnecessary detour through Moral Functionalism in any of its forms – all

²¹ This other version of Moral Functionalism draws on Jackson and Pettit's (1995: 26–27) discussion of moral disagreement, where they distinguish between moral disagreement as disagreement about which descriptive property fulfils some fixed rightness role, versus disagreement about which commonplaces specify the rightness role.

²² Another way that Moral Functionalism could require a check is through the internal role clauses that form part of the rightness role. These concern the relationships between rightness and other moral properties (Jackson 1998: 130–31). Suppose 'If x is right, then it would be wrong not to choose x ' is one such clause. Assessing whether descriptive property d satisfies this clause might require judging whether it is wrong to choose options which don't have d , which in turn might require something like a check for moral relevance.

²³ For instance, this might happen if the input clauses of folk morality were argued to be irrelevant to assigning a referent.

the justification for (C3) comes directly from the moral judgments which were taken as given in metasemantic theorising, rather than from the intermediate step through Moral Functionalism.

3.2 | The Connectedness Theory

Another theory which licences a problematic metasemantic argument is Laura and Francois Schroeter's (2014) connectedness theory. The theory builds its reference-determining conditions around the idea of a representational tradition – a set of token thought elements connected by relations of *de jure* sameness. Two thought elements stand in this relation of *de jure* sameness if thinkers take them to have the same subject matter.²⁴ For instance, when other English speakers describe things as 'right', we assume their thoughts are about the same topic as ours when we call things 'right'. Similarly, individuals take their current beliefs about what's right to be about the same topic as their past beliefs about what's right. Schroeter and Schroeter (2014: 14) propose:

(The Connectedness Theory) The sentence 'x is right' attributes the property picked out by a holistic rationalising interpretation of the representational tradition associated with 'is right'.

A holistic rationalising interpretation assigns a referent by taking into account the attitudes, dispositions, practices, and environmental feedback loops associated with a historically and socially extended representational tradition, aiming to vindicate the most important aspects of

²⁴ A representational tradition doesn't require that every thought element bears *de jure* sameness to every other, they just need to be connected by overlapping chains of *de jure* sameness (Schroeter and Schroeter 2014, 11-12). Also, Schroeter and Schroeter build into their theory of concept-identity that two token thought elements express the same concept only if the understanding and historical context associated with them does not diverge so radically as to undermine a univocal interpretation. This condition rules out, for instance, that the concept an ordinary British speaker expresses with 'corn' and the concept an ordinary American speaker expresses with 'corn' count as the same concept - given that 'corn' is generally applied by British speakers to any type of grain, while American speakers use the term exclusively for maize (Schroeter and Schroeter 2014, 16-17).

the tradition. To arrive at this holistic rationalising interpretation, we seek a wide reflective equilibrium that resolves various tensions within the communal representational tradition (L. Schroeter and Schroeter 2014: 20–22).

Suppose one discovers that property d best rationalises the representational tradition associated with ‘is right’. Using the Connectedness Theory in a metasemantic argument, one can conclude that x is right iff x is d .

Despite the fact that Schroeter and Schroeter are keen to stress that their metasemantic theory is built around our epistemic commitments concerning the moral domain (L. Schroeter and Schroeter 2014: 14), we think that a metasemantic argument from their theory generates a defeater. This is because the theory’s story about reference determination doesn’t involve checking the moral relevance of property d . The fact that the theory doesn’t involve such a check is vividly brought out by Sinhababu (2019: 6) who asks us to imagine that anti-egalitarian influences on folk moral views become more and more entrenched, shaping the representational tradition associated with ‘is right’. When this happens, Schroeter and Schroeter’s theory could well lead to the conclusion that ‘is right’ refers to the property of entrenching the advantage of the rich, or furthering the subjection of minorities. Here, the holistic rationalising interpretation doesn’t just say that ‘is right’ refers to a property that’s morally irrelevant – rather it picks out a referent that seems morally repugnant.

The metasemantic argument from the Connectedness Theory also suffers from information dependence, since it effectively assumes that the extended representational tradition contains more truths than falsehoods, or that the process of finding a holistic rationalising interpretation will assign more weight to true moral judgments than to false ones. In using the Connectedness Theory in a metasemantic argument, then, we merely presuppose certain moral claims rather than establishing them, thus preventing justification from transmitting to the normative ethical conclusion (C3).

Finally, the argument is also problematically indirect. Our moral judgments support (C3) directly, regardless of whether they also support the Connectedness Theory. If this is right, then none of the justification for (C3) is in fact transmitted through the metasemantic argument from the Connectedness Theory.

3.3 | The Epistemic Theory of Content

The third metasemantic theory we'll discuss is developed by Preston Werner (2020) – this theory is built from a claim about the relationship between metasemantics and epistemology, which was first recognised by Imogen Dickie (2015). Werner (2020: 157) proposes

(The Epistemic Theory of Content) The sentence 'x is right' attributes the property which the proprietary means of justification (for rightness beliefs) non-luckily converge on, in hospitable conditions.

Here, the proprietary means of justification for rightness beliefs are the non-derivative means of justification that generally trump other pieces of evidence for those beliefs.²⁵ Werner (2020: 157) cites intuitions, rational insight, and reflective equilibrium as examples of proprietary means of justification for our moral beliefs. Werner is clear that the proprietary means of justification explain what our moral beliefs refer to. Dickie's theory (2015: 108–13), by contrast, holds that there is no explanatory priority between what our moral beliefs refer to and what the proprietary means of justification converge on.

A metasemantic argument from the Epistemic Theory of Content is not defeated by subjects' higher-order epistemic commitments concerning checking. When determining whether our intuitions or moral insight converge on some property, we must either compare

²⁵ Werner (2020: 156–57) is unsure whether to define the proprietary means of justification as generally trumping other pieces of evidence, but we read him as specifying a necessary but insufficient condition for a proprietary means. The specifics of how these proprietary means are defined does not matter for our later arguments – all we need is for these means to involve a check for moral relevance.

different intuitions about (or insight into) specific cases, or else directly intuit (or see) the relevance of some descriptive property – both of which count as checks for relevance. Similarly, the method of reflective equilibrium involves finding a set of principles that, when applied competently, would yield our considered moral judgments (Rawls 1951: 184–87) – this requires identifying the morally relevant properties that would rationalise our considered judgments, which constitutes a check.

For similar reasons, the metasemantic argument also evades the problem of information-dependence. When we determine that the proprietary means converges on a descriptive property, we will have effectively established – rather than merely presupposed – justification for moral claims involving this property.

However this metasemantic argument still seems problematically indirect, on two counts. Firstly, whatever moral judgments were taken to justify the Epistemic Theory of Content would also justify the normative ethical conclusion (C3) directly, thus this argument takes an unnecessary detour through the Epistemic Theory. Secondly, establishing the premise that the proprietary means of justification converges on some descriptive property would suffice for justifying (C3), regardless of whether the Epistemic Theory is also true. For suppose that the Epistemic Theory were false – perhaps because its claim about explanatory priority doesn't hold (that is, our moral beliefs don't refer to some property *because* their proprietary means of justification converges on that property). The mere fact that the proprietary means of justification does converge on some property will still justify (C3). This suggests that the Epistemic Theory itself does no work in justifying (C3), despite appearances to the contrary.

To conclude, we've argued in this section that metasemantic arguments from a variety of different metasemantic theories fail to transmit justification from premises to conclusion. We take the lesson of our discussion to be that metasemantic arguments are unlikely to succeed. Of course, we haven't surveyed metasemantic arguments from all the metasemantic theories

on offer. Still, the theories we have considered are a good sample, suggesting that no such argument will transmit justification.

In the next section we consider objections to our case against metasemantic-first epistemology. However, before we move on, we want to contrast our discussion with similar work in the literature. Other philosophers have raised problems for moving from metasemantic theories to moral conclusions. Neil Sinhababu (2019: 3), in arguing against causal externalist metasemantic theories, contends that “The facts about causal regulation don’t settle the moral issue. Learning that I’m from the linguistic community where aggregate happiness causally regulates moral concepts intuitively doesn’t settle moral questions in favour of consequentialism.”²⁶ He concludes that Boyd’s theory (along with other externalist views) gives the natural, social, or historical environment too large a role in determining the referent of moral concepts, and is hence false. Similarly, Barry Maguire maintains that “even if some purely sociological account of the extension of some ethical word were offered, this would not *itself* settle any ethical question without admitting an *ethical* defence” (Maguire 2018: 441).

We agree with the general spirit of these comments, but our critique differs in important ways. First, we base our critique on an explicit reconstruction of the structure of metasemantic arguments, and we give three novel reasons for why metasemantic arguments might fail. Second, unlike Sinhababu, we believe the problem extends beyond externalist metasemantic views, to internalist views like Moral Functionalism. As long as a metasemantic argument satisfies the conditions for generating a defeater, for information dependence, and for indirectness, it will fail to transmit justification to its normative ethical conclusion. Third, Sinhababu (2019: 5–6) is concerned with both the epistemological and metaphysical implications of moral metasemantic theories, whereas our focus is exclusively epistemological; we focus on whether we can use metasemantic theories to extend our stock of justified moral beliefs, rather than on the truth-conditions of moral claims.

²⁶ See also Brink (2001), Sayre-McCord (1997), and Gampel (1997).

4 | Objections

We now consider some objections to our arguments against metasemantics-first moral epistemology. These objections are posed to the defeater-based challenge, but they suggest analogous worries for the challenge from information dependence too.²⁷

Objection #1: Suppose an oracle tells you that some moral metasemantic theory is true. You now appear to be independently justified in using (e.g.,) the Boydian Metasemantic Argument, without having to check the moral relevance of property *d* in the argument.

Reply: We concede that, in this case, you can be justified using the Boydian Metasemantic Argument without *yourself* checking the moral relevance of the property picked out as the referent of ‘is right’. Still, this doesn’t threaten our overall argument, for we maintain that a metasemantic argument from this theory will be threatened by the checking defeater unless the check for moral relevance is performed by *someone*. The oracle must themselves have performed a check for moral relevance.²⁸ When we understand the requirement to check in this more expansive sense, we can easily accommodate cases of testimony about moral metasemantics.²⁹

A different kind of testimony is testimony about the truth of first-order moral claims. Perhaps such testimony could independently justify moral claims – consequently, one might suggest that our list of methods of checking in Section (2) is incomplete. However, there is a straightforward response here: we are concerned with *non-derivative* methods of checking for moral relevance. Moral testimony is a derivative method for checking, in the sense that

²⁷ These objections all concern the necessary and sufficient conditions for checking moral relevance, and can be replaced with analogous objections about the necessary and sufficient conditions for establishing (rather than merely presupposing) justification for a moral claim.

²⁸ We assume here a weak version of the Transmission View in the epistemology of testimony (Leonard 2021: sec. 2.1); a hearer has justification on the basis of a speaker's testimony only if the speaker has justification too.

²⁹ These same points apply against the objection that once we have the oracle’s testimony, we don’t need to appeal to extensional adequacy to support a moral metasemantic theory (thus undermining the information dependence and indirectness challenges).

someone must have acquired non-testimonial justification for a moral claim before being able to transmit it via testimony.

Objection #2: Perhaps subjects have a commitment to checking the moral relevance of descriptive properties, but one can check the moral relevance of a descriptive property by investigating whether agents are inclined to classify things which they believe have *d* as right and *vice versa*. If this is correct, then metasemantic arguments from Moral Functionalism and the Connectedness Theory will count as incorporating a check in their reference-determining conditions – given that they take agents’ judgements about what is right as inputs.

Reply: Even if we concede this is a form of checking, this proposal runs into a defeater generated by another one of subjects’ epistemic commitments concerning the moral domain, which we will call *Fallibility*: moral agents can be mistaken about morality. We (individually and collectively) can make moral mistakes (Bengson, Cuneo, and Shafer-Landau 2022: 69; Cuneo 2020: 223). *Fallibility* is among the core data that constitute the theory-neutral starting points for metaethical theorising. Such data are defeasible but it is a cost to a theory if it fails to vindicate them.³⁰ This proposal is in tension with *Fallibility* because it suggests that we can determine which facts are the right-making facts simply by learning about what some group of agents believe about what is right.

Objection #3: The check for moral relevance must itself be supported by a theory of metasemantics, so we inescapably have to start with a metasemantic theory that isn’t supported by such a check. In particular, when checking for moral relevance, we need to know which judgments to perform the check with. But if we don’t know what our moral judgments are about, how can we tell which judgments are justified, reliable, or otherwise appropriate for use in checking moral relevance?

³⁰ See Bengson, Cuneo, and Shafer-Landau (2022: chaps. 2–3) for discussion of philosophical data.

The force of this objection can be illustrated with an analogy from Schroeter and Schroeter (2013: 6). Suppose two experimental philosophers, Faith and Charity, are studying a dart thrower. Faith wants to identify which throwing methods and conditions will help the thrower hit the target reliably – for instance, overhand versus underhand methods, windy versus windless conditions. Charity wants to identify which target they are aiming at – for instance, the hexagon or the ovoid. Schroeter and Schroeter contend that, in order for Faith to determine that a method or condition is conducive to reliable hits, they need the information that Charity is seeking – information about which exact shape the thrower is aiming at. Analogously, in order for us to tell whether a method (or condition) of forming moral judgments is reliable, it seems that we require a moral metasemantic theory which tells us what these judgments are about (which is akin to information about which target the thrower is aiming for).

Reply: It's plausible that performing the check for moral relevance requires knowing that certain methods and conditions of moral-judgement-formation are reliable. Thus if Schroeter and Schroeter are right, performing this check requires prior commitment to a moral metasemantic theory too. Fortunately, we believe that they've drawn the wrong lessons from their analogy, and that developing it reveals ways to identify reliable methods and conditions without assuming a full metasemantic theory. For example, suppose there are only two targets – the hexagon and the ovoid – and suppose that when the thrower uses the overhand throwing method, most of their darts land in the overlap between these two targets. In this case, Faith can deduce that overhand throwing is highly reliable, even without information about which target the thrower is aiming at: regardless of which target was aimed at, overhand throwing would likely have hit it. Analogously, we can identify reliable methods and conditions of moral-judgement-formation without assuming a specific metasemantic theory. If the different metasemantic theories overlap³¹ (in the sense that there are significantly many acts that these

³¹ The claims in this overlap are first-order moral claims rather than higher-order claims about the existence of genuine disagreement, so our argument doesn't clash with Dowell's (2016) conclusion that semantic intuitions

theories agree have the property of rightness) and if there exist methods and conditions that deliver verdicts in this overlap (that is, they yield judgments that these acts are right), then we can say that these methods and conditions are reliable, without assuming a specific metasemantic theory.³² These assumptions seem quite plausible: the moral metasemantic theories will likely agree on certain paradigmatically right actions (like benefitting someone at no cost to yourself), and there likely exist methods and conditions (like the use of intuitions under conditions where one has full information about the non-moral facts) that would yield the judgement that these paradigmatic actions are right. We conclude, then, that performing a check does not require one to be independently justified in believing a specific metasemantic theory.

5 | Conclusion

Moral metasemantic theories tell us what our moral terms are about – this might seem to offer hope that these theories could thereby also help us acquire and reinforce justification for our moral beliefs. We've argued, however, that this project faces serious difficulties. We first elucidated the structure of metasemantic arguments which purport to justify normative ethical conclusions. We then raised three challenges which suggest that metasemantic arguments fail to transmit justification from their premises to their conclusions. Such arguments are vulnerable to at least one of the following problems: of simultaneously generating a defeater, of information dependence, or of indirectness. For better or for worse, we suspect that we're stuck

about disagreement with a hypothetical community have no probative value. Moreover, we only require that this overlap be large enough to support assessment of reliable methods and conditions for checking – this is compatible with significantly revisionary first-order ethical implications, and thus does not necessarily conflict with Silverstein's (2019) arguments. Thanks here to an anonymous referee.

³² Conversely, we can also deduce that certain methods and conditions are unreliable if they consistently yield verdicts that are agreed to be wrong by all metasemantic theories. This requires assuming the falsity of a simple subjectivism, which says that the moral truth is what our actual moral judgments say they are (or, put in terms of the darts analogy, that the target's contours are not drawn exactly around where the darts actually land).

with standard methods of moral theorising – metasemantic theorising won't help us extend our stock of justified moral beliefs.

Acknowledgements We thank two anonymous referees for the Journal of Ethics for their perceptive and generous comments, which have improved this paper greatly. We would also like to thank Garrett Cullity, Frank Jackson, Neil Sinhababu, Zachary Swindlehurst, Daniel Stoljar, Hezki Symonds, Alexander Sandgren, James Willoughby, and an audience at the ANU Philosophy of Mind Work-In-Progress Group, for their comments and questions. The two authors contributed equally to this paper. This research was supported by an Australian Government Research Training Program (RTP) Scholarship.

References

- Armour-Garb, Bradley, Daniel Stoljar, and James Woodbridge. 2022. "Deflationism About Truth." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Summer 2022. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/sum2022/entries/ruth-deflationary/>.
- Bengson, John, Terence Cuneo, and Russ Shafer-Landau. 2022. *Philosophical Methodology: From Data to Theory*. Oxford, New York: Oxford University Press.
- Boyd, Richard N. 1988. "How to Be a Moral Realist." In *Essays on Moral Realism*, edited by Geoffrey Sayre-McCord, 181–229. Ithaca: Cornell University Press.
- Braun, David. 2016. "The Objects of Belief and Credence." *Mind* 125 (498): 469–97. doi:10.1093/mind/fzv137.
- Brink, David O. 2001. "Realism, Naturalism and Moral Semantics." *Social Philosophy and Policy* 18 (2): 154–76.
- Brown, Curtis. 2016. "Narrow Mental Content." In *The Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/archives/sum2016/entries/content-narrow/>.
- Burgess, Alexis, and Brett Sherman, eds. 2014. *Metasemantics: New Essays on the Foundations of Meaning*. Oxford, New York: Oxford University Press.
- Cappelen, Herman. 2018. *Fixing Language: An Essay on Conceptual Engineering*. Oxford, New York: Oxford University Press.
- Chalmers, David J. 2003. "The Nature of Narrow Content." *Philosophical Issues* 13 (1): 46–66.
- . 2011. "Propositions and Attitude Ascriptions: A Fregean Account." *Noûs* 45 (4). Wiley: 595–639.
- Cuneo, Terence. 2020. "Can Expressivism Have It All?" *Philosophical Studies* 177 (1): 219–41.
- Davies, Martin. 2009. "Two Purposes of Arguing and Two Epistemic Projects." In *Minds, Ethics, and Conditionals: Themes from the Philosophy of Frank Jackson*, edited by Ian Ravenscroft, 0. Oxford University Press. doi:10.1093/acprof:oso/9780199267989.003.0015.
- Dickie, Imogen. 2011. "How Proper Names Refer." *Proceedings of the Aristotelian Society* 111: 43–78.
- . 2015. *Fixing Reference*. Oxford; New York: Oxford University Press.
- . 2016. "The Essential Connection Between Epistemology and the Theory of Reference." *Philosophical Issues* 26 (1): 99–129.
- Dowell, Janice L. 2016. "The Metaethical Insignificance of Moral Twin Earth." *Oxford Studies in Metaethics* 11. Oxford University Press. doi:10.1093/acprof:oso/9780198784647.003.0001.
- Edwards, Kevan. 2014. "Keeping (Direct) Reference in Mind." *Noûs* 48 (2): 342–67. doi:10.1111/nous.12009.
- Evans, Gareth. 1973. "The Causal Theory of Names." In *Proceedings of the Aristotelian Society, Supplementary, Volumes 47:187–225*.
- Farkas, Katlin. 2008. "Semantic Internalism and Externalism." In *The Oxford Handbook of the Philosophy of Language*, edited by Ernest Lepore and Barry C. Smith, 323–40. Oxford; New York: Oxford University Press.
- FitzPatrick, William J. 2018. "Representing Ethical Reality: A Guide for Worldly Non-Naturalists." *Canadian Journal of Philosophy* 48 (3–4): 548–68.
- Fodor, Jerry. 1994a. "A Theory of Content, I: The Problem." In *A Theory of Content and Other Essays*. Cambridge, Massachusetts: MIT Press.
- . 1994b. "A Theory of Content, II: The Theory." In *A Theory of Content and Other Essays*. Cambridge, Massachusetts: MIT Press.
- Gampel, Eric. 1997. "Ethics, Reference, and Natural Kinds." *Philosophical Papers* 26 (2): 147–63.
- Huemer, Michael. 2008. "Revisionary Intuitionism." *Social Philosophy and Policy* 25 (01): 368–92. doi:10.1017/S026505250808014X.
- Jackson, Frank. 1998. *From Metaphysics to Ethics*. Oxford: Oxford University Press.
- . 2003. "Narrow Content and Representation, or Twin Earth Revisited." *Proceedings and Addresses of the American Philosophical Association* 77 (2): 55–70.

- Jackson, Frank, and Philip Pettit. 1995. "Moral Functionalism and Moral Motivation." *The Philosophical Quarterly* 45 (178): 20–40.
- Königs, Peter. 2020. "Experimental Ethics, Intuitions, and Morally Irrelevant Factors." *Philosophical Studies* 177 (9): 2605–23.
- Kripke, Saul. 1980. *Naming and Necessity*. Cambridge, Massachusetts: Harvard University Press.
- Leonard, Nick. 2021. "Epistemological Problems of Testimony." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Summer 2021. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/sum2021/entries/testimony-episprob/>.
- Lewerentz, Lukas, and Benjamin Marschall. 2018. "Metasemantics, Intentions and Circularity." *Synthese* 195 (4). Springer Nature: 1667–79. doi:10.1007/s11229-016-1290-y.
- Maguire, Barry. 2018. "The Autonomy of Ethics." In *The Routledge Handbook of Metaethics*, edited by Tristram McPherson and David Plunkett, 431–42. New York: Routledge.
- Moretti, Luca, and Tommaso Piazza. 2023. "Transmission of Justification and Warrant." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta and Uri Nodelman, Summer 2023. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/sum2023/entries/transmission-justification-warrant/>.
- Nye, Howard. 2015. "Directly Plausible Principles." In *The Palgrave Handbook of Philosophical Methods*, edited by Chris Daly, 610–36. Houndmills, Basingstoke: Palgrave Macmillan.
- Pryor, James. 2001. "Is Moore's Argument an Example of Transmission-Failure?" <http://www.jimpryor.net/research/papers/Moore2001.pdf>.
- Quine, Willard Van Orman. 1960. "Word and Object." *MIT Press*. <https://mitpress.mit.edu/9780262670012/word-and-object/>.
- Rawls, John. 1951. "Outline of a Decision Procedure for Ethics." *The Philosophical Review* 60 (2): 177–97. doi:10.2307/2181696.
- . 1974. "The Independence of Moral Theory." *Proceedings and Addresses of the American Philosophical Association* 48: 5–22. doi:10.2307/3129858.
- . 1999. *A Theory of Justice*. Revised edition. Cambridge, Mass: Belknap Press.
- Roojen, Mark van. 2006. "Knowing Enough to Disagree: A New Response to the Moral Twin Earth Argument." In *Oxford Studies in Metaethics, Volume 1*, edited by Russ Shafer-Landau, 161–94. Clarendon Press.
- Sayre-McCord, Geoffrey. 1997. "'Good' On Twin Earth." *Philosophical Issues* 8: 267–92.
- Schroeter, François, Laura Schroeter, and Kevin Toh. 2022. "The Limits of Metalinguistic Negotiation: The Role of Shared Meanings in Normative Debate." *Canadian Journal of Philosophy* 52 (2): 180–96. doi:10.1017/can.2022.27.
- Schroeter, Laura, and François Schroeter. 2013. "Normative Realism: Co-Reference without Convergence?" *Philosophers' Imprint* 13: 1–24.
- . 2014. "Normative Concepts: A Connectedness Model." *Philosophers' Imprint* 14 (25): 1–26.
- . 2018. "Metasemantics and Metaethics." In *The Routledge Handbook of Metaethics*, edited by Tristram McPherson and David Plunkett, 519–33. New York: Routledge.
- . 2019. "The Generalized Integration Challenge in Metaethics." *Noûs* 53 (1): 192–223. doi:10.1111/nous.12216.
- Segal, Gabriel. 2000. *A Slim Book About Narrow Content*. Cambridge, Massachusetts: The MIT Press.
- . 2009. "Narrow Mental Content." In *The Oxford Handbook of the Philosophy of Mind*, edited by Ansgar Beckermann, Brian P. McLaughlin, and Sven Walter. Oxford; New York: Oxford University Press.
- Silverstein, Matthew. 2019. "Revisionist Metaethics." In *Methodology and Moral Philosophy*, edited by Jussi Suikkanen and Antti Kauppinen, 214–33. Routledge.
- Sinhababu, Neil. 2019. "One-Person Moral Twin Earth Cases." *Thought: A Journal of Philosophy* 8 (1): 16–22.
- Tucker, Chris. 2010. "Transmission and Transmission Failure in Epistemology." In *Internet Encyclopedia of Philosophy*. <https://iep.utm.edu/transmis/>.

- Turner, Jason. 2013. "(Metasemantically) Securing Free Will." *Australasian Journal of Philosophy* 91 (2). Routledge: 295–310. doi:10.1080/00048402.2012.695381.
- Vayrynen, Pekka. 2019. "Normative Commitments in Metanormative Theory." In *Methodology and Moral Philosophy*, edited by Jussi Suikkanen and Antti Kauppinen, 193–213. New York: Routledge.
- Werner, Preston. 2020. "Getting a Moral Thing into a Thought: Metasemantics for Non-Naturalists." In *Oxford Studies in Metaethics*, edited by Russ Shafer-Landau, 15:140–69. Oxford: Oxford University Press.
- Wikforss, Asa. 2008. "Semantic Externalism and Psychological Externalism." *Philosophy Compass* 3 (ue 1): 158–81.
- Wright, Crispin. 2002. "(Anti-)Sceptics Simple and Subtle: G. E. Moore and John McDowell." *Philosophy and Phenomenological Research* 65 (2). [International Phenomenological Society, Philosophy and Phenomenological Research, Wiley]: 330–48.
- . 2003. "Some Reflections on the Acquisition of Warrant by Inference." In *New Essays on Semantic Externalism and Self-Knowledge*, edited by Susana Nuccetelli, 57–78. MIT Press.
- Zhao, Xinkan. 2021. "Metasemantics and Boydian Synthetic Moral Naturalism." *Synthese* 199 (3): 11161–78. doi:10.1007/s11229-021-03283-5.