Noncognitivism and agent-centered norms

Alisabeth Ayars[1] · **Gideon Rosen**[2]

Abstract This paper takes up a neglected problem for metaethical noncognitivism: the characterization of the acceptance states for agent-centered normative theories like Rational Egoism. If Egoism is a coherent view, the non-cognitivist needs a coherent acceptance state for it. This can be provided, as Dreier (Aust J Philos 74: 409–422, 1996) and Gibbard (Thinking how to live, Harvard University Press, 2003) have shown. But those accounts fail when generalized, assigning the same acceptance state to normative theories that are clearly distinct, or assigning no acceptance state to theories that look to be intelligible. The paper makes the case for this and then asks: What should we conclude if the problem cannot be solved? We might conclude that since Egoism is clearly a coherent (if mistaken) view, the argument amounts to a refutation of noncognitivism. But we suggest another possibility. There is, on reflection, something incoherent, or at least odd, in standard formulations of Egoism; noncognitivism predicts this and so provides an intriguing explanation for this fact.

Keywords Noncognitivism · Expressivism · Agent-centered norms · De se attitudes · James Dreier · Alan Gibbard · Egoism

1 Introduction

Understood as a thesis in the philosophy of mind, noncognitivism is the view that normative judgments consist (at least in part) in noncognitive pro- or con-attitudes: desires, intentions, states of approval and disapproval, or something of the sort.

✉ Gideon Rosen
grosen@princeton.edu

[1] Department of Philosophy, University of British Columbia, Vancouver, Canada

[2] Princeton University, Princeton, USA

Understood as a thesis in the philosophy of language, it is the thesis that normative statements like "Mary should mind her own business" express noncognitive attitudes in the sense in which a descriptive statement like "snow is white" expresses the ordinary factual belief that snow is white.[1] These more or less equivalent formulations set the agenda for noncognitivist accounts of normative thought and language. In constructing a detailed theory, the goal is to specify the noncognitive state associated, in this sense, with each normative statement S within the scope of the theory—i.e., the noncognitive state in which the judgment that S consists.[2] Following Dreier (1996), we can call this the acceptance state for S.

In this paper we take up a neglected issue for noncognitivism so-understood: the characterization of the acceptance state for agent-centered norms. By a norm we mean a general claim about what people should do in a given situation: a principle of the form *Everyone should φ in C*. Agent-neutral norms specify a single state of affairs that everyone should aim to realize or promote, and so "give everyone the same goal", in Parfit's phrase (Parfit, 1984, p. 27). The utilitarian claim that happiness is to be maximized instructs everyone to act so as to maximize happiness, and so implies that if you should do A (because your doing A would maximize happiness), then others should promote or at least not hinder in you in doing A, because according to the norm, we share a common goal that your doing A would realize. Agent-centered norms, by contrast, though general in form, give different agents different goals in a corresponding sense. A norm that requires everyone to keep her promises tells you to keep yours and us to keep ours, but says nothing about whether we should take an interest in your promise-keeping. (Contrast a norm that requires everyone to ensure that promises are kept.) The purest example of an agent-centered norm, and our main example here, is Rational Egoism, the view that everyone should do what's best for herself without non-instrumental regard for others.

---

[1] It is always a bit unclear what it means to say that a statement type S *expresses* a given attitude. The idea is roughly that S expresses A when a competent sincere assertoric utterance of S is caused in the right way by a token of A on the part of the speaker. But a sincere assertoric utterance of S will normally be caused by many attitudes, not all of which count as attitudes expressed by the statement. To our knowledge there is no standard characterization of the expression relation with which this sort of attitude-theoretic semantics is concerned; and yet there is no way to formulate noncognitivism as a thesis in the philosophy of language without invoking some such notion.

[2] As is familiar, the noncognitivist should not deny that normative statements express normative beliefs. Someone who *judges* that Fred should keep his promise ipso facto *believes* that Fred should keep his promise. So it is always true, if unilluminating, to say that a sincere competent utterance of 'X should φ' expresses the belief that X should φ. The nontrivial task is to provide a characterization of the normative belief that reveals the nature of the noncognitive state that underlies it (Rosen, 1998).

As is less familiar, the noncognitivist can understand this task in either of two ways. She can say that the judgment/belief that Fred should keep his promise is *identical* to the noncognitive state she associates with the judgment. (This has the awkward consequence that the state is technically cognitive after all, since beliefs are paradigmatically cognitive: truth-evaluable, aimed at truth, etc.) Alternatively, as we prefer, she can say that when someone judges/believes that Fred should keep his promise, she is in that cognitive state *in virtue of being* in a more fundamental, noncognitive state. This noncognitive state is distinct from the belief—as it must be if it is to ground it. However the relation is not causal or contingent, but rather constitutive: the belief is grounded in the non-cognitive state that underlies it in roughly the same sense in which the flower's being red is grounded in its being crimson.

In this paper we argue that existing noncognitivist accounts of the acceptance state for agent-centered norms fail for systematic reasons. We focus mainly on Dreier (1996), the original statement of the problem in its relevant form. Dreier's paper is a response to an old paper by Brian Medlin (Medlin, 1957), which takes a version of noncognitivism as a premise and argues for the incoherence of Egoism. Dreier sees Medlin's argument as a problem, not for egoism, but for noncognitivism. Metaethical theories should be neutral—or substantially neutral—about first-order normative issues. An account of the nature of normative judgment should not rule out views that look coherent, even if those views are substantively mistaken. So if noncognitivism rules out Egoism, as Medlin argues, noncognitivism is less neutral than it should be. Dreier offers a response on behalf of the noncognitivist, the main innovation of which is to treat the noncognitive states judgments of the form *X should φ* as *de se* attitudes in the sense of Lewis (1979), rather than *de dicto* attitudes as Medlin supposes. Our main claim is that despite the many virtues of this *de se* gambit, it fails. Any noncognitivist account must satisfy two desiderata: it should assign coherent acceptance states to coherent normative views, as Dreier emphasizes; but it should also assign distinct acceptance states to normative judgments that are clearly distinct, in the sense that each can be held without the other. We argue that Dreier's approach conflates distinct judgments, as does Alan Gibbard's rather different proposal, and so flouts the second desideratum.

2 Medlin's problem

As noted, Medlin's aim was to show that Rational Egoism—the perennial enemy of morality, with its other regarding demands—cannot coherently be accepted. The argument is straightforward. Rational Egoism says that everyone should do what's best for herself, where the relevant "should" is the all things considered practical "should" (as distinct from the moral "should" or the "should" of prudence narrowly conceived). So suppose Tom and Harry are running a race and that it would be best for each to win. The egoist must then accept (1) and (2):

(1) Tom should win (because that would be best for him)
(2) Harry should win (because that would be best for him)

The noncognitivist thinks these "should"-judgments express desire-like attitudes. In Dreier's cleaned-up version of Medlin, the relevant attitudes are preferences defined over states of affairs. On this view, to judge that X should $φ$ is to prefer X's $φ$ing to the alternatives. The egoist who accepts (1) and (2) must therefore have the following preferences:

P1$_M$ a preference for Tom's winning over Harry's winning
P2$_M$ a preference for Harry's winning over Tom's winning

But these preferences are incoherent given the facts. It's incoherent to prefer A to B and B to A when each excludes the other, and plainly irrational when the

incompatibility is obvious. Egoism is thus untenable, since it commits anyone who holds it to incoherent preferences in simple cases.

For Medlin, that's the conclusion. For Dreier, the challenge is to identify an acceptance state for "Everyone should benefit herself" that does not saddle the egoist with incoherent attitudes in simple cases. Medlin's argument shows that the acceptance state for statements of the form "X should φ" can't be a preference ordering that ranks <X φs> ahead of the alternatives, and more generally, that the acceptance states for normative judgments cannot be *de dicto* (or *de re*) preferences, i.e., preferences defined over states of affairs.[3] The challenge is to say what these states can be if not such preferences.

As Dreier points out, we cannot simply say that the egoist prefers doing what's best for herself. The egoist may, of course, have this selfish preference. But she also holds a *general* normative view that applies to everyone. She thinks that Tom should benefit Tom, and so holds some sort of pro-attitude towards self-interested behavior on Tom's part. The challenge is to characterize *this* pro-attitude—which we may call approval; and if we say no more than that the egoist prefers to benefit herself, we have not done that.

To put the point another way, this proposal would not distinguish the rational egoist, who thinks that everyone should benefit herself, from the egomaniac, who thinks that everyone should benefit *him*, or from the solitary shark, who thinks that he should benefit himself but has no view about what others should do. These characters all prefer doing what's best for themselves, but they are not philosophical egoists. The challenge Dreier extracts from Medlin's paper is to find a coherent set of preferences to go along with the egoist's *universal* prescription of egoistic conduct that would distinguish the egoist from these other characters.

3 Dreier's proposal

Dreier's solution[4] to Medlin's problem retains the idea that normative judgments express preferences, but drops the idea that the relevant preferences are preferences for states of affairs or propositions (sets of worlds). Instead they are *de se* preferences whose objects are *properties* (or sets of *centered worlds*).

Lewis (1979) famously maintains that the contents of ordinary beliefs are properties rather than propositions. When someone believes an ordinary proposition—e.g., that snow is white—he locates himself in logical space. He takes himself to inhabit a world where snow is white. So even if there is no overt indexical or

---

[3] This is a bit quick. This simple case is consistent with taking 'X should φ' to express a *weak preference* for X's φing: a preference ranking according to which no alternative to < X φs > is ranked ahead of <Xφs>. Views of this sort will generally entail that if I am indifferent between going to the movies and staying home, I can express my normative view by saying 'I should go to the movies', and that's quite wrong. See Silk, 2015 for a sophisticated defense of weak preference expressivism for certain deontic modals.

[4] To be clear: Dreier defends his view only as the best option for the noncognitivist. He does not defend noncognitivism against the cognitivist alternatives.

anaphoric pronoun in the sentence we use to specify the content of the belief, still the person who believes that snow is white has a belief about himself, viz., that *he* is such that snow is white. For every proposition p—every set of worlds—there is a corresponding property: the property of being such that p: $\lambda x.\ p$—which Lewis identifies with the set of individuals who inhabit p-worlds. So nothing is lost, formally speaking, if we say that the content of the belief that snow is white is not the proposition that snow is white, but rather the property of being such that snow is white, and more generally, that to believe that p is to self-ascribe the property of being such that p.[5]

The view comes into its own in its account of egocentric or self-locating beliefs—e.g., your belief that *you* are making a mess. In this case your belief self-ascribes a property that may distinguish individuals in a given world: the property of making a mess, $\lambda x.\ x$ *is making a mess*. In this case, the content of the belief cannot be a set of worlds. But it can be a property, a property you self-ascribe. We get a unified theory if we suppose that belief is a matter of self-ascribing a property in every case.

Lewis's framework—which can be modified and reformulated in many ways—has the advantage of explaining how a person who knows exactly which possible world he inhabits may nonetheless be ignorant (Frege, 1918; Perry, 1979; Lewis, 1979).[6] When Lingens the amnesiac wakes up in the Stanford library and wonders whether he is Lingens, he is not wondering which world he inhabits—whether snow is white, or whether Lingens is in the Stanford library. He may know the facts insofar as they are specifiable without indexicals down to the last detail. Instead he is wondering which properties he may accurately *self*-ascribe, and in particular whether he has the property of being Lingens: $\lambda x.\ x = Lingens$.

To self-ascribe a property, for Lewis, is to believe one has it. To self-*pre*scribe a property (Dreier's phrase) is to prefer to have it. If Smith wants to eat an egg, then what he wants is (to instantiate) a property: the property of being an egg eater.[7] If Tom prefers winning to losing, he has a preference ranking defined, not over sets of possible worlds, but over sets of possible individuals, or sets of centered worlds, or the like. He ranks $\lambda x.\ x$ *wins* ahead of $\lambda x.\ x$ *loses*, in the sense that he prefers having the former property to having the latter.

Now return to Egoism. To keep things simple, let's suppose that the egoist prefers whatever would make her happiest. Real eudaimonists have a complex preference schedule. They prefer more happiness to less, and so rank being happy to degree j ahead of being happy to degree k when j is greater than k. But again for

---

[5] In what follows we speak freely of possible worlds and possible individuals à la Lewis. But all of that is dispensable for our purposes. We could identify properties with sets of centered worlds, i.e., pairs of the form $< c,\ w >$ where c is an individual who would exist if w were actual. Or we could adopt an autonomous theory of properties and propositions that makes no foundational reference to worlds. So far as we can tell, these metaphysical issues are orthogonal to our concerns.

[6] For doubts about the framework and Lewis's reasons for accepting it, see Magidor (2015).

[7] If a more eccentric Smith wants to *be* an egg, then he wants to have a different property: *x. x is an egg*. Since this is a property that many things have, the view has the added bonus of giving non-empty content to certain eccentric desires that would have trivial contents if contents had to be sets of possible worlds.

simplicity, let's say that what the egoist prefers is that she make herself as happy as she can. So as far as her preferences for herself go, she has the simple ranking: λx. *x's maximizes x's happiness* $> \lambda x$. *x does not maximize x's happiness*. When we say that the egoist self-prescribes λx. *x maximizes x's happiness* — or that she prefers to benefit herself—that's what we mean.

So far this is just a claim about what the egoist wants for herself. But Rational Egoism is a normative view—a view about what should be done—and a general one at that: a view about what everyone should do. Dreier's most distinctive proposal on behalf of the noncognitivist is that the egoist's acceptance of this general view simply consists in her self-prescription of λx. *x maximizes x's happiness*.

> Speaking in ordinary untechnical English, it is very easy to enunciate a sense in which all egoists have just the same preferences. They all prefer more happiness to less. Agent-centered theories can't be summarized by a single betterness ordering for states of affairs, but they can be summarized by a single betterness ordering for properties. … Accepting an agent-centered theory is taking its betterness ordering as one's preference ordering, where preferences are construed as self-prescriptions. (Dreier, 1996: 416–417)

This gives a sense in which all egoists share a preference. They all self-prescribe the same property. It is of course the somewhat Pickwickian sense in which Snow White and the Wicked Queen share a belief when each judges that *she* is the fairest in the land. But in the case of desire and preference, calling this sort of thing "agreement" is less Pickwickian. Contestants in a race do want the same thing in a sense: they all want *to win*, which is to say, they all self-prescribe λx. *x wins*. On the proposed account, this preference is the acceptance state for the normative judgment *Everyone should win*.

Putting the point more generally, Dreier offers the noncognitivist the following view:

AS The acceptance state for a general normative claim of the form 'Everyone should φ' consists in the preference for, or self-prescription of, the property λx. *φx*.

In the case of an agent-neutral theory, the property will be the sort of property that does not distinguish among world-mates: λx. *p*, e.g., being such that happiness is maximized. In the case of an agent-centered theory, it will be a property in the more familiar sense: λx. *φx*—e.g., being such that one maximizes one's own happiness. According to the proposal under consideration, self-prescribing the latter property is the acceptance state for Rational Egoism.

The account involves an ingredient we should make explicit. It's clear what the egoist is doing when she says 'I should maximize my own happiness'. She is expressing the *de se* preference we've been discussing. But what mental state is she expressing when she says, 'Tom should maximize *his* happiness'? She must say this, given her philosophical view. But in saying it she is not expressing her preference for a state of affairs in which Tom benefits himself, since she may have no such preference. (She may prefer him to be miserable, if that would benefit her.) The

second component of Dreier's package is an account of the content of 'should'-judgments that covers the first-person case and the third-person case in a uniform way.

> SHOULD To judge that X should φ if to self-prescribe a property $\lambda x.\psi x$ and to judge that X can only instantiate $\lambda x.\psi x$ by φing.

When the egoist judges that Tom should win, she doesn't express a preference for Tom's winning. Rather she expresses her now familiar preference for a property— $\lambda x.\ x\ maximizes\ x's\ happiness$—together with the judgment that Tom can only instantiate this property by winning. This is still a noncognitive attitude, since it essentially involves a preference. However, it not a preference for Tom's winning, or maximizing his happiness, but rather a *de se* preference for being such that she maximizes *her own* happiness, together with the judgment that by winning Tom would instantiate the same property.

Dreier's proposal thus allows us to construct an attitude of approving of an action that is not the same thing as preferring that it be done. To approve of X's φing is to approve of X's goal in φing—or better, some property X would come to instantiate by φing—where you approve of a property by self-prescribing it. The noncognitivist can then say that in general, the pro-attitude expressed by 'X should φ' is approval of X's φing in this sense.

The proposal has two important virtues for present purposes. The first is that it renders Egoism and judgments that follow from it—e.g., (1) and (2)—perfectly coherent. There is nothing incoherent about self-prescribing the property of maximizing one's own happiness. Moreover the concrete judgments (1) and (2) now have the following acceptance states:

> P1$_D$ a preference for $\lambda x.x\ maximizes\ x's\ happiness$, together with the judgment that Tom can only instantiate this property by winning
> P2$_D$ a preference for $\lambda x.x\ maximizes\ x's\ happiness$, together with the judgment that Harry can only instantiate this property by winning

Since these attitudes are obviously cotenable, Dreier's proposal solves Medlin's problem.

The second virtue of the proposal is that it nicely distinguishes the egoist, who believes that everyone should benefit himself, from the egomaniac, who believes that everyone should benefit *him*. As Dreier puts it:

> If he is an egoist, Alan will self-prescribe this property: winning. If he is an egomaniac, Alan will self-prescribe this very different property: being such that Alan wins. (Dreier, 1996, 417)

The properties in question are clearly different. In a world in which Alan wins, only Alan has the first ($\lambda x.\ x\ wins$), but everyone (and every*thing*) has the second ($\lambda x.\ Alan\ wins$). The proposal thus assigns distinct acceptance states to these distinct normative views, and that's a plus.

4 The amnesiac egomaniac

The proposal has just the right structure to solve Medlin's problem. But it faces a number of problems, the first of which has to do with Dreier's characterization of egomania.

Suppose Alan is an egomaniac who believes that everyone should do what's best for him. According to the proposal, his egomania consists in a preference for being such that Alan's happiness his maximized. And that sounds right: If Alan is a normal egomaniac, he has this preference. But now consider the sort of complication for which the apparatus of *de se* attitudes was first introduced. Suppose Alan the egomaniac wakes up with no memory of who he is. He is still an egomaniac: he thinks that everyone should do what's best for *him*, whoever he is. But in his amnesiac state he does not prefer that everyone do what's best for *Alan,* in the same sense in which he does not believe that *Alan* is wearing purple sneakers when he looks down and sees that *he* is wearing purple sneakers. So his egomania can't consist in a preference for Alan's happiness—*λx. Alan's happiness is maximized*. And of course it can't consist in a preference for maximizing *his* happiness, since that's the acceptance state for Egoism and Alan is not an egoist. So the view as stated fails to identify an acceptance state for the amnesiac egomaniac's egomania.

Another way to put the problem is to note that a preference for *Alan's* happiness should really be the acceptance state, not for egomania in Alan's case, but for *Alan fetishism*: the view that everyone should do what's best for Alan. If Josie and Alan both prefer *λx. Alan's happiness is maximized*, they are not both egomaniacs, since you can't be an egomaniac in virtue of preferring the happiness of someone else. The attitude they share is thus not egomania, but a maniacal devotion to Alan. If they are to agree in egomania, that will be at least in part because each prefers that *his* or *her* happiness be maximized. The proposed account of the acceptance state for egomania leaves this out.

So here's the situation. We have three general norms: egoism, egomania and Alan fetishism. Each can be embraced by Alan in his amnesiac state, and he will be taking a different normative view in each case. So we need three distinct acceptance states. In the framework we've been exploring, we know what the answer must be in two cases.

View	Normative formulation	Acceptance state
Egoism	'Everyone should benefit himself'	preference for *λx. x benefits x*
Alan fetishism	'Everyone should benefit Alan'	preference for *λx. x benefits Alan*
Egomania	'Everyone should benefit me'	?

The challenge is to fill in the blank in a way consistent with AS and SHOULD, and in a way that distinguishes the views from one another.

One natural thought is that while the amnesiac egomaniac does prefer to maximize his own happiness—$\lambda x.\ x$ *benefits x*—he also prefers to have a stronger property, viz., that *everyone* act so as to maximize his happiness: $\lambda x.\ \forall y.\ (y$ *benefits x)*. He's an ego*maniac* after all; he thinks that everyone in the universe should benefit him. Given Dreier's apparatus, this normative judgment is constituted by a *de se* preference for a property whose definition embeds a universal quantifier. In this respect it resembles Alan fetishism, which is naturally taken to involve a preference for a property that might better be given as follows: $\lambda x.\ \forall y\ y$ *benefits Alan*. This account of the acceptance state for egomania captures the thought mentioned above that if Alan and Josie are agree in their egomania, they must have a common *de se* preference that everyone do what's best for them.

This is a proposal for what belongs in the unfilled box:

View	Normative formulation	Acceptance state
Egoism	Everyone should benefit himself	preference for $\lambda x.\ x$ *benefits x*
Alan fetishism	Everyone should benefit Alan	preference for $\lambda x.\ \forall y\ y$ *benefits Alan*
Egomania	Everyone should benefit me	preference for $\lambda x.\forall y\ y$ *benefits x*

This satisfies the requirement that distinct normative views be assigned distinct acceptance states. But it won't do. Recall that on Dreier's proposal,

SHOULD To judge that X should φ if to self-prescribe a property $\lambda x.\psi x$ and to judge that X can only instantiate $\lambda x.\psi x$ by φing.

Given this, if Alan the amnesiac egomaniac self-prescribes $\lambda x.\ \forall y\ (y$ *benefits x)*, he thereby judges that *Josie* should be such that everyone maximizes *her* happiness, since he prefers a property she would instantiate if she were in this condition. This means that when Alan the egomaniac prefers that everyone maximize his happiness, he thereby approves of Josie's acting to ensure that everyone benefits *her*. But this mischaracterizes the egomaniac's normative view. The egomaniac does not approve of Josie's acting to ensure that everyone benefit *her*. He approves of everyone, including Josie, acting so to benefit *him*.

5 The solitary shark

Another problem for the proposal is that it cannot distinguish the Rational Egoist—who holds a general normative view—from the *solitary shark*: a blinkered creature who believes that he should benefit himself but has no view about what others should do. The shark may be unopinionated because he's cautious—unwilling to generalize beyond his own case—or because he's uninterested in others and simply hasn't bothered to ask himself what they should do. Either way, the shark self-

prescribes $\lambda x.\ x\ maximizes\ x$'s happiness and so judges that he should benefit himself without accepting the general norm that everyone should benefit himself.

The shark's normative view is intuitively coherent. It is the view of someone who, for whatever reason, holds a view about what he should do but no such view about others. This may be substantively unreasonable; we're sure it is. But Dreier's proposal entails that it is altogether impossible to be cautious in this way. By self-prescribing $\lambda x.\ x\ maximizes\ x$'s happiness, the shark *automatically* takes the general normative view that everyone should act so as to benefit himself. But that is to conflate the shark with the egoist, and so to deny the possibility of the sort of shark we have in mind.

One might contend that the solitary shark, if he is Alan, instead self-prescribes $\lambda x.\ x\ maximizes\ Alan$'s happiness, or perhaps $\lambda x.\ Alan\ maximizes\ Alan$'s happiness. This would avoid the implication that the shark approves of selfish behavior on the part of others. But of course this will not do either. The first preference is the acceptance state for a form of Alan fetishism on the proposed account, and the shark is not an Alan fetishist even if he happens to be Alan. And more importantly, the shark may not have either of these *de re* preferences. He may have amnesia, after all, in which case his self-concern will be essentially *de se*. The only possible content for it would be a preference for $\lambda x.\ x\ benefits\ x$; but that, we have said, is the acceptance state for Egoism.

The problem is that in Dreier's framework, anyone with a *de se* preference automatically holds a universal normative view, since in preferring a property he thereby judges that *everyone* should instantiate it. The view thus leaves no room for the normative outlook of someone like the shark who has *de se* preferences and normative views about his own case, but refuses to hold a universal normative view of any sort.

Some philosophers will see this as feature rather than a bug. There is a long tradition of thinking that normative judgments are somehow essentially universal. Hare famously holds that moral judgments are universal, in the sense that any such judgment "logically commits the speaker to making a similar judgment about anything which is either exactly like the subject of the of the original judgment or like it in relevant respects" (Hare, 1963: p. 139). The idea that moral judgments are universal in this sense is, as Dreier says, "designed to capture the intuitive thought that moral judgments are not supposed to be judgments essentially wedded to any particular, especially not to any particular person" (Dreier, 1996: p. 419). It is plausible that judgments involving the practical 'should' are universal in whatever sense moral judgments are universal. So one might deem it a virtue of Dreier's proposal that it rules out the shark, who refuses to universalize his judgments.

But it is one thing to hold that normative judgments are inevitably universal*able*, quite another to hold that they are automatically universal*ized*. Alan-fetishism may be silly, in the sense that any reason to favor Alan would be a reason to favor anyone qualitatively like Alan. But it's not impossible to be an Alan-fetishist, so it's good that Dreier's theory provides an acceptance state for the position. In our view, the same goes for the shark. The shark may be unreasonable, in the sense that any ground he has for taking himself to be governed by a norm is equally grounds for taking others to be governed by it. But epistemic caution on this

point is not impossible. And that's the problem. On Dreier's view it follows from the nature of normative judgment that the shark who judges that he should benefit himself is not just *committed* to holding a general view tantamount to Rational Egoism; he is already in exactly the same mental state as the egoist. But that's not so, and that's the problem.

6 Problems for Gibbard's plan expressivism

The problems we've been discussing arise for a broader class of expressivist views, the most prominent member of which is Allan Gibbard's plan expressivism (Gibbard, 1990, 2003). Gibbard's leading idea is that normative judgments are constituted, not by preferences, but by plans. When I judge that I should φ, my judgment consists in a plan—roughly, an intention—to φ.[8] When I judge that *Harry* should φ, my judgment consists in a plan to φ 'in Harry's shoes'—a contingency plan for what to do should I find myself in Harry's predicament. The acceptance state for a singular claim of the form '*X* should φ' is thus in general a conditional plan: the plan to φ on the condition that one is in *X*'s situation.

Gibbard's framework, like Dreier's, construes all normative judgments as essentially *de se*, even when the judgment is about what someone else should do. English grammar omits the subject in infinitival clauses, so we speak simply of my plan *to* φ *in C*. But it is clear that every such planning state consists in a system of intentions *for the planner* to do this or that in such and such conditions. So if I judge that Harry should φ, I thereby express *my* plan for *me* to φ should I find myself in Harry's situation, and this sort of plan is essentially *de se*.[9]

This leads to complexities given Gibbard's well-motivated insistence that when I plan for what to do "in Harry's situation" I am planning for what to do in Harry's *total* situation, i.e., the situation in which *I am Harry* in Harry's circumstances (Gibbard, 2003: p. 51). This can seem odd, since modal orthodoxy tells us that if I'm not in fact Harry then there is no possible situation in which I am. The view thus seems to construe run of the mill normative claims like "Harry should get a job" as expressing plans for what to do in metaphysically impossible situations.

But in fact there is nothing odd about such plans, as Gibbard shows. Suppose Anastasia wakes up with amnesia. She's still Anastasia, and necessarily so. But for

---

[8] As Gibbard emphasizes, the states he calls plans are not exactly intentions as we ordinarily understand the notion. Most importantly, Gibbard-style plan for what to do in *C* may *forbid* certain options while expressly *permitting* others (Gibbard 2003, 55). A plan that permits both *A* and *B* is inconsistent with a plan that forbids *A*. By contrast, the intention to do either *A* or *B* is not inconsistent with the intention not to do *A*. This point is important to Gibbard's ingenious solutions to a range of problems. However it will not matter in what follows. For discussion of this aspect of Gibbard's view, see Schroeder, 2008, Silk, 2015 and Ayars, (forthcoming).

[9] Why not say instead that my judgment that Harry should φ consists in a plan *for Harry* to φ? Gibbard rejects this view on the ground that it is conceptually impossible for one person to plan for another in the sense in which a person can plan for herself. For extensive discussion of this aspect of Gibbard's view and a version of plan expressivism that rejects it, see Ayars (forthcoming). This revised version of plan expressivism solves many of the problems we have raised for Gibbard here.

all she knows before she opens her eyes and checks her driver's license, she could be Harry in Harry's situation, or Josie in Josie's situation, etc. In this amnesiac state she can wonder who she is and make contingency plans for the various epistemic possibilities. Her plan to φ if she is Harry is a perfectly intelligible mental state—one that will lead her to φ if she comes to believe that she is Harry, and one whose content is (roughly) the set of centered worlds in which either the center is not Harry or the center φs. Since this is a non-empty set, the plan has a consistent content. According to Gibbard, this essentially *de se* state is the state that we express when we judge that Harry should φ.

Gibbard's view solves Medlin's problem as well as Dreier's does. The egoist's judgments about the competitive race —

(1) Tom should win
(2) Harry should win

are constituted by the following plans:

P1$_G$ a plan to win if one is Tom in Tom's situation
P2$_G$ a plan to win if one is Harry in Harry's situation.

And there is nothing inconsistent in holding these plans simultaneously.[10]

And yet the view faces versions of the problems we raised for Dreier. Start with the shark, who judges that he should benefit himself but withholds judgment when it comes to others. If he is Alan and he knows it, he judges both that he should benefit himself and that Alan should benefit himself, while taking no view about what Harry and Anastasia should be doing. Now suppose he's hit on the head and contracts amnesia. He still judges, *de se*, that he should benefit himself but now takes no view about what Alan should do. In this state the shark has plans. He plans to benefit himself; that's the plan in which his judgment that he should benefit himself consists. But he also has conditional plans defined over the open epistemic possibilities. He plans to benefit Harry if he is Harry, to benefit Anastasia if he is Anastasia, and so on. And there's the rub. For on Gibbard's view, these conditional plans are the acceptance states for "Harry should benefit Harry" and "Anastasia should benefit Anastasia". The view thus entails that in losing his memory the shark has all of a sudden become opinionated about what other people should do. In planning to benefit Harry if he is Harry, he thereby judges unconditionally that Harry should benefit himself. But that's clearly wrong. In his amnesiac state, he is *uncertain* about whether Harry should benefit himself, since he's uncertain whether he is Harry.

But in fact it's worse than this. The shark's defining characteristic is that she accepts no general norms of the form 'Everyone should φ'. And yet his selfish judgment, 'I should do what's best for me', is grounded, for Gibbard, in what amounts to a general plan. This is clearest if the shark has amnesia, since the

---

[10] The content of a Gibbard-style plan is naturally given by a set of centered worlds, with the plan to φ in C having as its content the set of worlds at which the center φs if the center is in C. A world in which Harry is the center and Harry wins is a world at which both P1$_G$ and P2$_G$ are satisfied, so the plans are jointly satisfiable.

amnesiac shark intends to benefit X if he is X, no matter who X may be. But the shark who knows exactly who he is presumably has the same general plan. When he gets hit on the head he loses his memory and with it his *de re* plans for Alan; but the *de se* plan that remains—to benefit X if he is X—must have been there all along. And the trouble is that this plan is presumably the Gibbard-style acceptance state for the general view that everyone should benefit himself. The proposal thus assigns the same acceptance state to the amnesiac shark's cautious judgment that he should benefit himself and the rational egoist's universal claim that everyone should do this. But these are clearly different judgments, so the view flouts our desideratum that distinct judgments should have distinct acceptance states.

A similar argument shows that Gibbard's view cannot assign a coherent acceptance state to egomania. The egomaniac judges that everyone should benefit him, and if he knows he's Alan, that everyone should benefit Alan. But suppose he loses his memory and so loses the *de re* judgment. In this state he still accepts the *de se* claims 'Harry should benefit me', 'Josie should benefit me', and so on. But now consider the Gibbard-style acceptance states for these claims. To accept the first is to intend to benefit oneself if one is Harry. But to intend to benefit oneself if one is Harry just is to intend to benefit Harry if one is Harry, and this latter plan is the Gibbard-style acceptance state for 'Harry should benefit Harry'. This is bad for two reasons. The first is that the egomaniac in his amnesiac state does not judge that Harry should benefit Harry. He is agnostic on this point, pending information about whether he is Harry. The second is that the view assigns the amnesiac egomaniac a general schematic plan—to benefit X if he is X.[11] But this is presumably the acceptance state for *egoism*. The view thus fails to distinguish the acceptance states for egoism and egomania, at least in the case of the amnesiac. But even in that case, these normative views are clearly different.

7 A fregean response

It would be natural at this point to suspect that these difficulties having to do with amnesia are all artefacts of an excessive reliance on a Lewis-style account of egocentric thought as involving relations to properties (or sets of centered worlds) rather than ordinary propositions (sets of worlds). Take the amnesiac shark who judges that he should benefit himself while taking no view about others. If his judgment consists in a preference for a property that others might instantiate—$\lambda x.\, x$ *benefits* x—Dreier's account saddles him with the general judgment that everyone should benefit himself, thus rendering him indistinguishable from the egoist. But

---

[11] We say "schematic" for the following reason. The amnesiac egomaniac's egomania does not consist in a finite list of plans to benefit Harry if he is Harry, Tom if he is Tom, etc., so it is general in a sense. But it can sound odd to characterize his mental state by saying that *for all x*, he plans to benefit x if he is x. This seems to imply that he plans to benefit Anastasia if he is Anastasia; but he has no such plan if he has never heard of Anastasia and can't think about her. Rather his mental state seems to be one that commits him to forming such a plan should Anastasia come to his attention. For want of a better term, we may call this form of generality *schematic*, by analogy with the sort of generality that attaches to schematic principles in formal logic.

suppose we reject the idea that *de se* preferences are preferences for properties and instead follow Frege in treating egocentric thought as a relation to complete propositions involving special first-person concepts (Frege, 1918). On Lewis's view, when the amnesiac Lingens judges that he is in the library, the content of his belief is $\lambda x.\ x$ *is in the library*; so his belief has the same content as Alice's belief that *she* is in the library. On the alternative Fregean view, Lingens' belief has as its content a Fregean proposition involving a proprietary self-concept that Lingens and Lingens alone uses to think about himself: $< I_{Lingens}$ *am in the library* $>$. If Alice is in a similar predicament, her belief has an altogether different content, viz., $< I_{Alice}$ *am in the library* $>$.

On a view of this sort, the amnesiac shark's judgment that he should benefit himself might consist in a preference, not for a property others might share, but for a fine-grained proposition: $< I_{Shark}$ *benefit myself* $>$. This preference can't be the acceptance state for Egoism. After all, most egoists can't even entertain this proposition, since they don't have access to the shark's proprietary I-concept. And even if they can, they will not in general prefer that it be the case. So this view is well-placed to avoid the conflation of the shark's narrowly focused selfish ethos (*I should benefit myself*) and the Egoist's general view (*Everyone should benefit himself*) which arises for Dreier and Gibbard.

As things stand, however, the Fregean view is too inchoate to be discussable. We have said what the acceptance state for Egoism is *not* within this framework, but we have not said what it is. If we follow Dreier and Medlin, it must be a preference, and if we follow Frege, the content of this preference must be a proposition (or a ranking defined over propositions). But what is the propositional preference that distinguishes the egoist qua egoist? It's not a preference for $< I_{egoist}$ *benefit myself* $>$, since that won't distinguish the egoist from the shark. And it's not a preference for $<$ *Everyone benefits himself* $>$, since (a) that preference is incoherent given the stipulated facts, since it's satisfied only if Tom beats Harry and Harry beats Tom, which is impossible; and (b) the egoist qua egoist does not prefer that everyone benefit himself. He *approves* of others benefiting themselves; but he does not want this, since he may be selfish. The point of the exercise was to find the non-cognitive state in which this approval consists. Dreier and Gibbard wield the *de se* framework to give us answers to this question that render Egoism coherent, and which are not implausible as accounts of the Egoist's state of mind. The Fregean has not begun this project, and it is quite unclear what she might say. So as things stand, this is not a discussable alternative to the views we have considered.

8 Desire noncognitivism

Medlin's problem arises for any view on which the acceptance states for (1) and (2) involve what Schroeder calls an 'inconsistency-transmitting attitude' towards inconsistent contents. Contents are inconsistent, for present purposes, if they are not jointly satisfiable: Propositional contents are inconsistent if there is no possible world at which all are true; property contents are inconsistent if no possible individual satisfies them all. An attitude-type is inconsistency-transmitting when

anyone who holds it towards obviously inconsistent contents is ipso facto criticizable as irrational or inconsistent (Schroeder, 2008). The paradigmatic inconsistency-transmitting attitude is full belief, *de dicto* or *de se*.[12] But (strong) preference is also inconsistency transmitting. To prefer *p* is to rank *p* ahead of the alternatives, and anyone who ranks *p* ahead of the alternatives to *p* cannot coherently rank an obvious alternative to *p* ahead of *p*. Medlin's version of preference noncognitivism entails the incoherence of Egoism precisely because it has these two features. The view identifies normative judgment with an inconsistency-transmitting attitude-type—*de dicto preference*—while representing the Egoist as holding this attitude towards obviously inconsistent contents (*Harry wins/Tom wins*).

Dreier and Gibbard retain the idea that normative judgments involve an inconsistency-transmitting attitude while finding consistent contents for the attitudes expressed by (1) and (2). For Dreier, the attitude-type is *de se* preference, which is inconsistency-transmitting.[13] But the preferences expressed by (1) and (2) have the same content, *λx. x wins*. Since this property is obviously consistent with itself, the view thus renders the egoist's judgments coherent. In Gibbard's case, the attitudes expressed by (1) and (2) are a plan to win if one is Tom and a plan to win if one is Harry. Planning, like preferring, is inconsistency-transmitting. It is a *de se* attitude towards properties, and anyone who plans to have obviously inconsistent properties is criticizable. However, the properties that Gibbard identifies with the contents of the plans expressed by (1) and (2)—*λx. x wins if x is Harry/Tom* — are consistent. If Tom wins, he has them both. So considered abstractly, Dreier and Gibbard solve Medlin's problem in the same way: by going *de se* and finding consistent contents for the attitudes expressed by (1) and (2). And the trouble is that this gambit runs into trouble when the amnesiac shark and other inconvenient characters are on the scene.[14]

This suggests another strategy for solving the problem, namely, by construing the acceptance state for normative judgment as involving an attitude-type that is not inconsistency-transmitting. As a toy proposal, consider desire noncognitivism: the view that judgments of the form *X should φ* express the desire that *X φ*.[15] Ordinary desire, unlike preference, is not inconsistency-transmitting. If Sasha wants to go to

---

[12] Though the preface paradox may show that even full belief is not inconsistency transmitting without restriction.

[13] More exactly, the attitude expressed by a normative judgment is a mix of *de se* preference and belief, since the judgment that Harry should win expresses a preference for benefitting oneself together with the belief that Harry can only benefit himself by winning. There are several ways to extend the notion of consistency to these hybrid attitudes, but however it is done, Dreier's proposal will entail that normative judgment is inconsistency-transmitting.

[14] The complacent cognitivist solves the problem without going *de se*. For the complacent cognitivist, normative judgments express beliefs, the paradigmatic inconsistency-transmitting attitude. But the contents of the beliefs expressed by (1) and (2) are normative propositions—*Tom should win, Harry should win*; and these propositions are (on the face of it) straightforwardly consistent.

[15] A slightly more plausible version of the view would identify normative judgment, not with ordinary desire, but with some sort of beefed-up desire, e.g., the desire that X φ together with a non-instrumental desire to retain that desire, as in Lewis's (1989) account of valuing.

the movies but also wants to curl up with a book, she is not criticizably irrational; she's just normal. So another way to address Medlin's problem is to see (1) and (2) as expressing, respectively, the desire that Tom win and the desire that Harry win— a state in which the subject is conflicted but not incoherent or inconsistent.

This would not only solve Medlin's problem. It would solve many of the problems we have identified for Dreier and Gibbard. The desire non-cognitivist can say that Egoism consists in the desire that everyone benefit himself; that Alan fetishism consists in the desire that everyone benefit Alan; that the amnesiac's egomania consists in a de se desire that everyone benefit *him*, and that the amnesiac shark's limited egoism consists in a *de se* desire whose content is *λx. x benefits x*. The approach avoids the conflations that plagued the other approaches by rejecting SHOULD, the principle that turns every *de se* preference into a universal normative judgment, in favor of a simpler idea:

SHOULD* To judge that X should φ is to desire that X φ.

Of all the views we have discussed, this one stands the best chance of assigning distinct acceptance states to normative views that are distinct from one another, and assigning coherent acceptance states to views that appear to be coherent, including Egoism.[16]

But this version of expressivism is untenable for other reasons. To cite only the most obvious: the rational egoist who judges that everyone should do what's best for himself may be a selfish bastard who only *wants* what's best for himself, or a cool economist who judges that rational actors like Tom and Harry should benefit themselves while taking no motivationally efficacious interest in whether they do what they should. The chief motivation for noncognitivism down through the ages has been that first-person normative judgments play a motivational role that points to a desiderative nature. But—as these examples show—*third-personal* normative judgments involving the all things considered 'should' display no such internal connection to motivation on the part of the judging subject. The views of Dreier and Gibbard respect this asymmetry.[17] The present view does not.

---

[16] It is not obvious that desire noncognitivism renders Egoism coherent if it identifies Egoism with the desire that everyone benefit himself. After all, in the cases that generate Medlin's problem it is impossible for everyone to do what would most benefit himself, and it is arguably incoherent to desire a manifestly impossible state of affairs. The most interesting response to this worry is to identify Egoism, not with a desire for the universally quantified state of affairs—*Everyone benefits himself*—but rather with a schematic desire: a state that commits the subject to a desire that x benefit himself for each individual x who comes to his attention, without committing the subject to a desire for the conjunction of these states (n.10 above). A desire of this sort is general *in sensu diviso* in Abelard's sense, as distinct from the universally quantified desire, which is general *in* sensu *composito*. Since it is not incoherent to desire two states of affairs that cannot be realized together, this would save the egoist from the charge of incoherence.

[17] In Dreier's case the judgment that X should φ involves a preference for $\lambda x.\varphi x$ together with the belief that that X can only by φing. The preference for $\lambda x.\varphi x$ may motivate the subject to see to it that she has this property; but it will not motivate her to see to it that others have it. So Dreier's view respects the asymmetry. In Gibbard's case, the judgment that X should φ expresses (roughly) the plan to φ if one is X. This planning will motivate one to φ provided one believes that one is X; but it will not motivate one to see to it that others φ. So Gibbard's view likewise respects the asymmetry.

And then there are the familiar problems for any view on which the state that underlies normative judgment fails to be inconsistency-transmitting. It had better be incoherent to judge that X should φ and also that X should ψ when it is obviously impossible for X to do both (where X is a single agent).[18] But if these judgments are constituted by the *desire* that X φ and the desire that X ψ, then the judgments are not inconsistent, but merely reflect run of the mill ambivalence.

Here is a way to state the larger upshot of this point. To judge that X should φ, we have said, is to *approve* of X's φing, For the noncognitivist, approval is a noncognitive pro-attitude, and the challenge throughout has been to characterize this attitude in illuminating terms. If the view is to solve Medlin's problem, painting Rational Egoism as a coherent view, then approval cannot be interpersonally inconsistency-transmitting: It must be coherent to approve of X's φing and Ys ψing even when it is impossible for X to φ and Y to ψ, as in (1) and (2). And yet approval must be *intrapersonally* inconsistency-transmitting. It must be incoherent to approve of X's φing and X's ψing when it is obviously impossible that X do both. The views of Dreier and Gibbard satisfy this condition, to their credit.[19] The desire view and every more plausible variant of it do not, so those views too are unacceptable.

9 Conclusion

We began with two constraints on noncognitivist accounts of normative thought and language. A tenable view must provide.

(a) coherent acceptance states for views that are clearly coherent; and
(b) distinct acceptance states for views that are clearly distinct.

If we think Rational Egoism is coherent, it follows from (a) that we need a coherent acceptance state for it. Medlin's problem shows that *de dicto* preference non-cognitivism fails this desideratum. Gibbard and Dreier provide versions of *de se* noncognitivism that do well with (a); but these views fail with respect to (b), as we have seen.

We have not canvassed every possibility, so we cannot say that there is no view that satisfies both desiderata. But it seems safe to say that it will not be easy to thread the needle on the assumption that Rational Egoism and similar agent-centered norms are in fact coherent. So we should ask: What should we conclude if the problem cannot be solved?

---

[18] It is sometimes said that the analogous principle in the moral case is falsified by dilemmas, in which the agent is morally required to perform each of two incompatible actions. But whatever one says about this, it is harder to believe that the practical 'should' can give rise to such dilemmas. One standard gloss brings this out: If to say that X should φ is to say that X has *most reason* to φ, then it clearly can't be the case that X should φ and that X should do something incompatible with her φing.

[19] The pertinent accounts are as follows:
   APPROVAL$_D$ S approves of X's φing iff S prefers to and believes that X can only by φing.
   APPROVAL$_G$ S approves of X's φing iff S plans to φ if he is X in X's situation.
   It is easy to verify that these attitudes are inter- but not intra-personally inconsistency-transmitting.

We could conclude that since Egoism is clearly coherent, the argument amounts to a refutation of noncognitivism. The cognitivist, after all, has no trouble identifying a coherent acceptance state for Egoism. For her it is simply the *belief* that everyone should do what's best for himself, a belief whose content is a normative proposition that looks to be consistent in every relevant sense. Alternatively we could follow Medlin and say that insofar as noncognitivism remains a serious contender in metaethics, the argument amounts an unrebutted challenge to the Egoist.

Needless to say, we cannot adjudicate this issue here. But we will conclude with one small point in favor of the latter option. Return one last time to the Egoist's verdicts:

(1)   Tom should win (because that's what's best for him)
(2)   Harry should win (because that's what's best for *him*)

Egoism is coherent only if the acceptance states for these judgments cohere with one another. But note that on reflection, even in the absence of theory, there is something odd about these judgments. It would be fine to say that Tom and Harry should both *try* to win. But it sounds odd to say outright that both *should* win in a context where it is clear that that's impossible. Suppose Tom and Harry are friends and come to you before the race to ask how what they should do this afternoon. It would be odd for you in your capacity as advisor say: 'Well, everyone should do what's for himself, so you, Harry, should win, and you, Tom, should win.' That sort of advice sounds not just false but incoherent.[20]

The non-cognitivist of a certain sort has an explanation for this. If her account of the acceptance state for agent-centered norms entails that it is incoherent to approve of incompatible actions—if approval turns out to be interpersonally inconsistency-transmitting—that would explain why it sounds so odd to affirm (1) and (2) together, and why we feel compelled to retreat from Egoism as standardly formulated to the weaker (and rather different) view that everyone should *try* to do what's best for himself. That view is still an agent-centered norm. But it is an agent-centered norm with the special property that (very odd cases aside), it is always possible for everyone to do what the norm requires.[21] Medlin's problem for Egoism arises because Egoism sometimes requires incompatible courses of action. But any view with this feature will sound odd when applied to cases in just the way that Egoism does. If normative judgments express an intrapersonally inconsistency-transmitting attitude, the noncognitivist can explain the oddness.

Note that the cognitivist has no similar explanation for the felt oddness of asserting (1) and (2) together, or for the pressure to retreat to the "trying" version of Egoism as a formulation of the only coherent position in the vicinity. Why shouldn't

---

[20] The explanation for this cannot be that 'should' implies 'can'. For we can stipulate that in every relevant sense, Tom can beat Harry and Harry can beat Tom. If this requires that the race be set in an indeterministic world, so be it.

[21] The odd cases are cases in which two people—say conjoined twins—share the circuitry necessary for choosing or intending, so that if the one tries to do what's best for her, the other cannot even try to do what's best for him.

the normative propositions expressed by (1) and (2) both be true in the circumstances? Of course she can take the substantive view that one can only be required to φ if it is entirely within one's power to φ. But then it does not flow from the nature of normative judgment itself that (1) and (2) can't both be true, but rather from a substantive and highly revisionary view about what the true practical norms must say.

This amounts to running Medlin's argument in reverse. Medlin assumes noncognitivism and argues for the incoherence of Egoism. We note the incoherence (or at least the oddness) of Egoism and argue for noncognitivism as an intriguing explanation for this fact. But of course this can only be a suggestion at this point. What the noncognitivist needs is an independently plausible account of normative judgment which has as a consequence that approval is intrapersonally inconsistency-transmitting. Preference noncognitivism the relevant feature, but it is *not* independently plausible: As we have seen, the Egoist does *not* prefer that Tom do what's best for himself, even though he does approve of it. The problem is that preference non-cognitivism does not respect the asymmetry noted earlier, viz., that while approving of one's own actions is motivationally efficacious in a distinctive way, approving of Tom's actions is not motivationally efficacious in this way, even if it is a non-cognitive state of some sort.

As we see it, this sets a new problem for metaethical noncognitivism: to provide an account of practical approval—the pro-attitude one takes towards *X's* φing when one judges that *X* should φ—according to which approval is inconsistency-transmitting across the board, even in the interpersonal case, but motivationally efficacious only in the first-person case. None of the views we have canvassed solves this problem. Whether it can be solved is a matter for another time.

References

Ayars, A. Deciding for Others: A non-cognitivist account of third-personal normative judgments. Forthcoming in *Philosophy and Phenomenological Research*.

Dreier, J. (1996). Accepting agent centered norms: A problem for non-cognitivists and a suggestion for solving it. *Australasian Journal of Philosophy, 74*(3), 409–422.

Frege, G (1918–1919). Der Gedanke. *Beiträge Zur Philosophie des Deutschen Idealismus* 2:58–77.

Gibbard, A. (2003). *Thinking how to live*. Harvard University Press.

Gibbard, A. (1990). *Wise choices, apt feelings: A theory of normative judgment*. Harvard University Press.

Hare, R. M. (1963). *The language of morals*. Oxford University Press.

Lewis, D. (1979). Attitudes de dicto and de se. *Philosophical Review, 88*(4), 513–543.

Lewis, D. (1989). Dispositional theories of value. *Aristotelian Society Supplementary, 63*(1), 89–174.

Magidor, O. (2015). The myth of the de se. *Philosophical Perspectives, 29*, 249–283.

Medlin, B. (1957). Ultimate principles and ethical egoism. *Australasian Journal of Philosophy, 35*(2), 111–118.

Parfit, D. (1984). *Reasons and persons*. Oxford University Press.

Rosen, G. (1998). Blackburn's essays in quasi-realism. *Noûs, 32*(3), 386–405.

Schroeder, M. (2008). *Being for: Evaluating the semantic program of expressivism*. Oxford University Press.

Silk, A. (2015). How to be an ethical expressivist. *Philosophy and Phenomenological Research, 91*(1), 47–81.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.