The Southern Journal of Philosophy
Volume 50, Issue 4
December 2012

RELATIVIST DISPOSITIONAL THEORIES OF VALUE

Andy Egan

abstract: Adopting a dispositional theory of value promises to deliver a lot of theoretical goodies. One recurring problem for dispositional theories of value, though, is a problem about *nonconvergence*. If *being a value* is *being disposed to elicit response R in us*, what should we say if it turns out that not everybody is disposed to have R to the same things? One horn of the problem here is a danger of the view collapsing into an error theory—of it turning out, on account of the diversity of agents' relevant dispositions, that nothing is really a value, since nothing is disposed to elicit R in everybody. Alternatively, there is a danger of an objectionable fragmentation of value, according to which there is no such thing as a value *simpliciter*, but only $\text{values}_{\text{me}}$ and $\text{values}_{\text{you}}$, $\text{values}_{\text{us}}$ and $\text{values}_{\text{them}}$. I advocate a *de se* relativist version of a dispositional theory of value. If we go for this sort of *de-se*-ified dispositional theory, we get to keep our theoretical goodies, but we avoid the problem of nonconvergence that leads to a danger of either collapse into an error theory, or else talking-past, and a loss of common subject matter.

1. INTRODUCTION

Consider the account of value offered by David Lewis (1989). According to Lewis's view, for something to be a value is (nearly) for it to be such that we would desire to desire it under conditions of full imaginative acquaintance.[1] This Lewisian view is a nice example of a dispositional theory of value. In general, dispositional theories of value are theories according to which *being a value* is identified with some dispositional property—typically a property of

Andy Egan received a PhD from MIT in 2004. He is Associate Professor at Rutgers University and Professorial Fellow at the Arché Philosophical Research Centre, University of St. Andrews. He works primarily in philosophy of language and philosophy of mind. Some of his other papers relevant to the current topic include "Disputing about Taste" in *Disagreement*, ed. Ted Warfield and Richard Feldman (Oxford University Press, 2010), "Quasi-Realism and Fundamental Moral Error" (*Australasian Journal of Philosophy*, 2007), and "Secondary Qualities and Self-Location" (*Philosophy and Phenomenological Research*, 2006).

[1] Officially: "Something of the appropriate category is a value if and only if we would be disposed, under ideal conditions, to value it" (Lewis 1989, 113).

The Southern Journal of Philosophy, Volume 50, Issue 4 (2012), 557–82.
ISSN 0038-4283, online ISSN 2041-6962. DOI: 10.1111/j.2041-6962.2012.00136.x

the type, *being disposed to elicit response R in us in circumstances C*. A dispositional theory of a particular value F (say, beauty or moral rightness) will have the same structure: it will say that *being F* is *being disposed to elicit R in us in C*. (We will look a bit more at how the details might be spelled out later on.) I will use Lewis's view as my main stalking horse in what follows, when concreteness will be helpful. But my target class of theories and theorists is quite broad.

Adopting a dispositional theory of value promises to deliver a lot of theoretical goodies. Dispositional theories of value are *cognitivist*: evaluative belief is just belief of the paradigmatic sort, with a distinctive sort of subject matter—evaluative beliefs are beliefs about the distribution of certain response-dispositional properties. Dispositional theories of value are *naturalistic*: given their account of the nature of evaluative properties, there is no problem with finding room for the evaluative in a naturalistic metaphysics.[2] And they promise, with appropriately careful selection of the responses and the elicitation conditions, to deliver the kind of tight connection between evaluative judgment and motivation that would make them *internalist*. (I am going to remain as silent, and as neutral, as I can about just what kind of tight connection that is, exactly.[3]) This is (to many of us) an extremely attractive combination. Unsurprisingly, dispositionalist theories—both of value generally and of particular values—have been popular in the literature.[4]

One recurring problem for dispositional theories of value—one that Lewis himself is quite concerned with—is a problem about *nonconvergence*. If *being a value* is *being disposed to elicit response R in us*, what should we say if it turns out that not everybody is disposed to have R to the same things? One horn of the problem here is a danger of the view collapsing into an error theory—of it turning out, on account of the diversity of agents' relevant dispositions, that nothing is really a value, since nothing is disposed to elicit R in everybody. Alternatively, there is a danger of an objectionable fragmentation and relativity of value, according to which there is no such thing as a value *simpliciter*, but only $\text{values}_{\text{me}}$ and $\text{values}_{\text{you}}$, $\text{values}_{\text{us}}$ and $\text{values}_{\text{them}}$, from which various

[2] They are also, as Lewis notes, standardly naturalistic in another sense—they purport to offer analyses of value in other terms.

[3] It is possible, depending on your theoretical preferences about just what kind of connection there ought to be, to tailor your choice of responses and elicitation conditions to deliver quite a wide range of different sorts of such connections. I am, personally, quite sympathetic to Michael Smith's (1989) sort of internalism, where what is wanted is *pro tanto* motivation absent irrationality, but dispositionalism has the resources to underwrite many different sorts and strengths of connections between evaluative belief and motivation, depending on the particular choices of response and elicitation conditions.

[4] In addition to Lewis 1989, see Smith 1989, 1994, 2002, Railton 1986, and Johnston 1989 for some dispositional theories of value.

bad consequences follow. (I will say more about what these bad consequences might be later on.)

One response to this problem is to argue that we really *should* be confident that we are going to get universal convergence, and so there is no danger of sliding into an error theory, and no need to go in for any fragmentation and subscripting of value (Smith 1994, 1997). Another is to argue that the kind of fragmentation and relativity that results when we move away from values *simpliciter* and toward values$_x$ and values$_y$ should not worry us. Perhaps the purported bad consequences do not actually follow, or perhaps they are not really so bad (see, e.g., Lewis 1989, Harman 1975, Dreier 1990, Brogaard 2012).

I want to propose a third way. In what follows, I will advocate a *de se* relativist version of a dispositional theory of value. On this sort of view, evaluative belief is not a matter of believing that *x* is disposed to cause R in Ks in C, but of self-attributing the property, *being someone in whom x is disposed to cause R in C*. (Or if you like your property names to start with 'λ' rather than 'being', *λy.x is disposed to cause R in y in C*.[5]) If we go for this sort of *de-se*-ified dispositional theory, we get to keep our theoretical goodies—the resulting theory is still cognitivist, naturalistic, and (at least as) plausibly internalist—but we avoid the problem of nonconvergence that leads to a danger of either collapse into an error theory, or else talking-past, and a loss of common subject matter.

Sadly, but unsurprisingly, going this way does not just help us avoid the problems about convergence; it also buys us some new problems and challenges. One of these is that, in order to extend the story from evaluative *thought* to evaluative *talk*, we need to say something about how to understand *de se* content in language. Another is that we have to square the analysis according to which the content of evaluative belief and assertion is *de se*, with the clear sense that many differences of opinion (and of assertion) with respect to the evaluative are *disagreements*.

2. THE GENERAL FORM OF A DISPOSITIONAL THEORY OF VALUE

Lewis offers a particular version of a dispositional theory of value, which we will use as our model. His version is (almost) as follows: something is a value just in case we would be disposed, under ideal conditions, to value it.[6] Lewis

[5] To translate back to "being . . ." names, read *λy.BLAH* as *being a y that satisfies BLAH*.

[6] The official version is in note 1. I have left out the bit about being of the correct category, since nothing in our discussion will (as far as I can tell, at least) hinge on it.

notes that there are some follow-up questions one should ask on being presented with this view. I will focus on three of his five: What is this attitude, valuing? What kinds of conditions are ideal? Who are *we*?[7]

Lewis's initial statement is pretty general, but we can be more general still. Dispositional theories of value say:

x is a value iff *we are disposed to have R to x in C*[8]

or

being a value = *being disposed to elicit R from us in C.*

There will be parallel analyses to offer, not just of value generally, but of particular values; *being right*, *being beautiful*, etc., are going to be various properties of the type, *being disposed to elicit R in us in C*, differentiated at least by different choices of R, and perhaps also of C. Lewis's version is what you get when you substitute 'value' for R and 'ideal conditions' for C. (And in particular when you then go on to substitute 'desire to desire' for 'value' and 'conditions of full imaginative acquaintance' for 'ideal conditions'.)

It is easy to see what is attractive about this kind of story. It makes the world safe for cognitivism, by providing a clear subject matter for evaluative thought and talk to be about. Evaluative belief is ordinary belief with a distinctive subject matter—it is about which things we are disposed to have distinctive responses to in distinctive kinds of circumstances. Evaluative discourse is discourse about that same distinctive subject matter. It demystifies value talk by giving an analysis in terms that are antecedently better understood. It locates values squarely in the natural world, avoiding commitment either to error theories or to extravagant metaphysics. And it holds out hope of underwriting internalism—if we spell out the responses and the circumstances correctly, there is liable to be a connection of the right (deep but defeasible) kind between evaluative belief and motivation.[9]

There is a lot of room to fight about whether, and how badly, we should really want the (purported) goodies that a dispositional theory of value delivers.

---

[7] His other two questions are "what is the 'appropriate category' of things?" and "what is the modal status of the equivalence?" (Lewis 1989, 113).

[8] In fact we will want something stronger than the biconditional. The property identity below would do the trick. There are other candidates, but I am going to steer clear of going into the details here, since they do not matter for our purposes. It is also worth noting that here, as elsewhere in this paper, I slide between *x is disposed to have R to y* and *y is disposed to elicit R in x* formulations without argument or (further) comment. There is room to worry about this, but everything I say could be rephrased to avoid any sliding back and forth, at the cost of some awkward wordiness.

[9] Depending on just what kind of connection we want, this might require different choices about the response and the elicitation conditions than the particular ones that Lewis makes.

Maybe we should not be trying so hard to secure cognitivism.[10] Maybe an error theory would not be so bad.[11] Maybe a nonnaturalistic metaphysics of value, properly understood, is not as extravagant as its opponents make it out to be.[12] Maybe we should not, after careful consideration, want to vindicate internalism after all.[13] But suppose you *do* want a cognitivist, naturalistic, non-error-theoretic, internalist theory of value. Then something that fits this general schema for a dispositional theory of value is liable to look extremely attractive.

I will be arguing for the following conditional conclusion: *if* you want a cognitivist, naturalistic, non-error-theoretic, internalist theory of value, then your needs will be better served by a *de se* relativist version of a dispositional theory of value than by the versions that have been on the menu to date.[14] But, since the move to relativism brings with it some new costs and theoretical burdens of its own, I will be happy if I can convince you of something a bit more modest: that you can get the goodies offered by standard sorts of dispositional theories of value, and avoid some of their most serious problems, by going relativist. Whether we can get from there to my more ambitious conclusion, that if you are going to be a dispositionalist, you should also be a (*de se*) relativist, will depend on the outcomes of some debates that I will only be able to make a start on here.

3. THE BIG PROBLEM FOR DISPOSITIONAL THEORIES OF VALUE: THE THREAT OF NONCONVERGENCE[15]

Recall the general form of a dispositional theory of value: *x* is a value iff we are disposed to have R to *x* in C. Recall also one of Lewis's follow-up questions: Who are *we*?

If everybody were disposed to have R to the same things in C, then the answer would be straightforward, and things would be easy for the dispositionalist. But securing this sort of universal convergence is a tall order. It is, for

[10] Blackburn 1984, 1993, 1998; Gibbard 1990, 2003; Stevenson 1944, 1963; Hare 1952.

[11] Mackie 1977; Joyce 2001; Daly and Liggins 2010; Streumer, forthcoming.

[12] Parfit 2011, Wedgwood 2007, Enoch 2011, Shafer-Landau 2003.

[13] Railton 1986, 2009, Brink 1989, 1997, Svavarsdóttir 1999, Lillehammer 1997, Cuneo 1999, Dreier 2000, Wedgwood 2007.

[14] The kind of relativism that I am advocating here is, obviously, an application of just one of several species of relativism that have entered the market in the last ten years or so. (It is the one I have advocated in e.g. Egan 2006, 2007, 2010. For other options, see for example Lasersohn 2005, MacFarlane 2005, 2011, Kölbel 2003, 2004, 2008, Stephenson 2005, 2007, Brogaard this volume. Other varieties of relativism will retain some of the features of my view that I will be emphasizing, but they will not all retain all of them. I will leave discussion of the relative merits of different varieties of relativism for another occasion.)

[15] *The* big problem? I think the definite article is warranted, but that is a bit contentious. It does not really matter, though, whether it is *the* big problem or just *a* big problem—so long as being able to avoid it can serve as a motivation for a *de se* relativist version of the theory.

example, far from clear that absolutely everybody is going to be disposed to desire to desire alike under conditions of full imaginative acquaintance. In fact, convergence here seems quite unlikely—it seems far more likely than not that at least some of the diversity in what people *in fact* desire to desire is traceable to simple differences in temperament, taste, and preference, which improved imaginative acquaintance would not do away with. (Lewis is, of course, well aware of this. More on his response below, in note 17.)

This is not, I think, a particularly idiosyncratic feature of Lewis's way of implementing a dispositional theory. For lots of ways of filling in for R and C that look attractive as ways of spelling out a dispositional theory of value, it is far from obvious that absolutely everybody will be alike with respect to what they are disposed to have R to in C. In many cases, something stronger is true: we should, absent some fancy argument, be pretty confident that we *will not* get this kind of convergence.

So if "we" is *everybody*, then there is a danger that there will not be any values, since there will not be anything that is disposed to elicit the relevant response in the relevant circumstances from absolutely everybody. There are two standard moves available to avoid getting signed up for an error theory on account of nonconvergence: (1) give a substantive argument for convergence; (2) go contextualist[16] and say that who "we" is varies across speakers/thinkers, and perhaps also across contexts for a particular speaker/thinker.[17]

---

[16] A terminological note: the way the terminology has developed in the literature to date, the positions labeled "relativist" in metaethics are ones whose counterparts in epistemology or philosophy of (nonnormative) language would be labeled "contextualist." In particular, they would fall on the "contextualist" side of the "contextualism/relativism" debate about, for example, epistemic modals and predicates of personal taste. I will call the views "contextualist," to maintain the parallels with structurally similar debates elsewhere, but I will mostly use "*de se*" or "*de se* relativist" rather than "relativist" *simpliciter*, since there is room to fight over what kind of view or views best deserve the name "relativist" and because so many different kinds of views sail under that flag.

[17] Two other options to note: First, Lewis himself endorses what he calls a "conditionally relative" account: "In making a judgment of value, one makes many claims at once, some stronger than others, some less confidently than others, and waits to see which can be made to stick. I say X is a value; I mean that all mankind are disposed to value X; or anyway all nowadays are; or anyway all nowadays are except maybe some peculiar people on distant islands; or anyway. . .; or anyway you and I, talking here and now are; or anyway I am. How much am I claiming?—as much as I can get away with. . . . So long as I'm not challenged, there's no need to back down in advance; and there's no need to decide in advance how far I'd back down if pressed. What I mean to commit myself to is *conditionally relative*: relative if need be, but absolute otherwise" (Lewis 1989, 129; first ellipsis original, second mine). There is a lot to be said for this proposal. But I think it would be better to have something more straightforward to say: Lewis's conditionally relative story inherits conditional versions of all of the problems with nonconditional contextualist accounts, and it also incurs a difficult-seeming theoretical burden—cashing out the details of just how the conditional relativity works. (Are all of the various-strength claims semantically expressed? Are they all believed? Or are only some

The worry about nonconvergence is, I think, especially acute if we want a view that is going to secure a deep connection between evaluative belief and motivation. One strategy for getting convergence is to build a lot of substantive constraints into C. (At the extreme, cartoonish endpoint, this could be done by saying that ideal conditions are those in which the agent is disposed to respond *thus*. Less cartoonish versions can rule out certain kinds of nonconvergence by, for example, building an "ideal rationality" condition into C and then advancing a very substantive account of what ideal rationality amounts to.[18]) But the more one builds into C, the more vulnerable the view becomes to complaints of the form, "why should we think that there is any kind of tight connection between what we are motivated to do and our beliefs about what we would do/say/think/feel in *those* circumstances?" There is also a danger, depending on what exactly you build into C, of losing the naturalism that was one of the initial attractions of a dispositionalist view. We might, for example, be tempted to secure convergence by building in substantive *normative* or *evaluative* constraints on C. But if these normative or evaluative constraints are not then given their own naturalistically kosher analysis, we will not really have delivered the naturalistic goodies that we were trying to deliver. This leaves the defender of universal convergence—at least, the universalist who wants to underwrite internalism—with a difficult tightrope-walking task: they need to provide a story about C—typically, about what makes for ideal conditions—that is strong enough to guarantee convergence but also weak enough to secure the right kind of universal *pro tanto* motivational force—to keep the story from being vulnerable to the complaint that there is no requirement of the right kind to be *motivated* by what one would do, think, feel, desire, or advise in C. (And also, it must be free of appeal to unreduced normativity, if the story is to be a naturalist one.) This is a tall order—I suspect that there is just no path to walk between these constraints. It is certainly not going to be straightforward to find one.

---

expressed or believed, and the others stand in some more complicated relation to the relevant assertions or doxastic states? Etc.) I do not say that the story-completing burden cannot be discharged. But it would be nice to be able to avoid incurring it in the first place.

Another potential option is to say that the relevant group is invariant across speakers/thinkers, but also a sufficiently (and appropriately) reduced subset of absolutely everybody that, within the relevant group, there *is* universal convergence. I am going to set this option aside, since I think that the project of finding principled grounds for selecting one rather than another of the converging subgroups as *the* group that is relevant to everybody's moral thought and talk—and in particular doing so in a way that preserves internalism (and avoids objections of the form, "why think that Xs should be, even *pro tanto*, motivated by judgments about what Ys would do/want/feel/advise?")—is pretty hopeless. But if I am wrong about that, then there should be a third move on the table. Thanks to the editors for raising this possibility.

[18] See for example Smith 1994.

With that said, for the purposes of this paper, I am going to set aside the option of insisting on a universal reading of "we" and defending the plausibility of convergence. Not because I think it is hopeless, and certainly not because I have in my pocket some knock-down *argument* that it is hopeless.[19] I think, though, that it is pretty clear that securing convergence is going to be a difficult task, and it is far from clear that it will work out. What I am concerned to establish is just that there is a market for a view that does not have to fight this fight. And I think it is clear enough, without going too far into the details, that this is going to be a pretty gnarly fight, which it would be nice to be able to sidestep.

The main alternative strategy on the menu of theoretical options in the literature is to go for a *contextualist* account. This is the sort of view according to which it might turn out that if our responses and the Martians' would not converge under the relevant sort of idealization, *my* evaluative thoughts are about $\text{values}_{\text{human}}$, while J'onn J'onzz's are about $\text{values}_{\text{martian}}$. Or, a bit closer to home, if your idealized responses and mine do not converge, it will turn out that my evaluative thoughts are about the $\text{values}_{\text{me}}$, or the $\text{values}_{\text{my group}}$, while yours are about the $\text{values}_{\text{you}}$ or the $\text{values}_{\text{your group}}$. This gives us a view that says:

x is a value_{K} iff *Ks are disposed to have R to x in C*

or

being a value_{K} = *being disposed to elicit R from Ks in C.*

Different people's evaluative beliefs and assertions will, at least potentially, be about different value_{K} properties.

I will spend a little more time discussing why this is also problematic than I spent on the convergence-defending response. Here too, I am not aiming to give some conclusive argument that the proposal will not work or that the problems are insoluble.[20] I am, again, just aiming to show that it is problematic enough to create a market for a story that does not face the problems.

One worry about this sort of contextualist proposal is that it gives up on the project of identifying a common subject matter for different people's and groups' evaluative thought and talk. On the contextualist picture, we are

[19] For expressions of pessimism about convergence (either about the evaluative or about the normative), see Sobel 1999, Plakias 2011, Robinson 2009, Plunkett 2010, and Prinz 2007.

[20] For arguments that contextualist accounts of normative discourse lead to disastrous results, see Kölbel 2004, as well as Kolodny and MacFarlane 2010; for arguments that things are not so bad for the contextualist, see Finlay and Björnsson 2010 and Plunkett and Sundell MS.

going to get different people, and different groups, thinking and talking about different families of properties when they think evaluative thoughts and make evaluative claims. So we will not have a theory that delivers a distinctively evaluative subject matter for everybody's evaluative thought and talk to be about.

There are two different complaints in the neighborhood here. The first is that we wind up with no common subject matter for everybody's evaluative thought and talk. The second is that the subject matter of my (or any particular person's or group's) evaluative thought and talk is not distinctively evaluative—when you have a thought with the same content as my evaluative thought, your thought need not be an evaluative one.[21] (A cartoon case: when an Australian thinks that Vegemite elicits desire to desire in Australians in ideal circumstances, that is her thinking Vegemite is valuable. But when I think a thought with the same content—that Vegemite elicits desire to desire in Australians in ideal circumstances—that is *not* me thinking that Vegemite is valuable.[22]) The evaluativeness of thoughts (and assertions) is not going to be lodged in their content. Instead, the evaluativeness of a belief is going to depend both on its being a belief with the content that, for example, Australians are disposed to desire to desire Vegemite under conditions of full imaginative acquaintance, plus the fact that *Australian* is the (or perhaps a) category that the believer identifies with in such a way that beliefs about the idealized dispositions of members of *that* category are (defeasibly) motivational in the right kind of internalism-securing way. (Or: that the thought with that content is gotten at via the right kind of evaluative mode of presentation. Then it will be the fact that *Australian* is such a category—one with which the thinker identifies in the right way—that determines that it is the group provided, in the believer's context, to fill in the content of their beliefs and assertions that deploy evaluative concepts, modes of presentation, or expressions.) Another way to put this point: there should be no problem about me having thoughts about which things are $\text{values}_{\text{you}}$. But those thoughts, in my head, will not be evaluative thoughts. This strikes me as an uncomfortable and unsatisfying result. If, for example, we set off on our analysis of value

[21] Dreier (1990, 2009) is admirably explicit about this.

[22] On many ways of implementing the contextualist story, there is also the same kind of variation within an individual—I can have two beliefs with the same content, and one be an evaluative belief and the other not, on account of a difference in the mode of presentation by means of which I am getting at that content. So whether the Australian's belief that vegemite elicits desire to desire in Australians in ideal circumstances is an evaluative belief or not will depend on the answer to some further questions about the belief-state (in Perry's 1979, 1980 sense) or mode(s) of presentation involved.

looking to, in Frank Jackson's (1998) phrase, find the subject matter of evaluative thought and talk in the natural world, it does not look like we have succeeded.

How serious a problem these consequences are is disputable. I think that both the loss of a common subject matter for everybody's evaluative beliefs, and the absence of any distinctively evaluative subject matter for evaluative belief to be belief about, are pretty dissatisfying. But I can understand not being too worried about this, since the concerns are a bit theoretically loaded. (Dreier [1990, 2009] has some things to say about why we should not be so worried, in the case of his relevantly similar metaethical view.)

Here are two slightly more principled reasons to be worried: maybe it does not get you all of the cognitivism, or all of the internalism, that you want. If what you wanted from a cognitivist, internalist account of value was a characterization of a distinctively evaluative subject matter for evaluative thought and talk to be about—a specification of a domain of distinctively evaluative things to believe and assert—such that beliefs about *that* were appropriately tightly tied to motivation, then you will not get what you wanted from this sort of contextualist account.

On a contextualist dispositional theory of value, evaluative beliefs turn out to be the usual kind of belief with the usual kind of content, sure enough. But what makes the belief with that content *evaluative*—and what gives it its distinctive motivational force—is something nondoxastic. How come? Well, my evaluative thoughts are thoughts of the type, *x is disposed to cause R in Ks in C*. What fixes which K goes into the content of my evaluative thought? If we want the story to be internalist, it had better be something about my attitude toward K, such that believing that members of *that* kind would, for example, value *x* under conditions of full imaginative acquaintance is going to, with the appropriate qualifications, tend to bring it about that I pursue *x*. (Thinking that psychopaths, or Martians, or white supremacists, would value *x* under conditions of full imaginative acquaintance will not have the relevant connection to motivation for me—at least, not without strong convergence assumptions that I am now assuming that we are not entitled to.) The reason why my thought *that humans/Wisconsinites/philosophy professors/Green Bay Packers fans/etc. would desire to desire x under ideal circumstances* gets to count as evaluative is because the kind *human* (Wisconsinite, philosophy professor, Packers fan, etc.) occupies that special conative place in my heart that makes it the kind that my value-ish modes of presentation hook onto in my (present) context.[23]

[23] Here, and in what follows, there are parallel concerns to have about the sort of metaethical contextualism defended in, e.g., Dreier 1990. Dreier's account is not framed in terms of dispositions of a contextually selected group, but of the verdicts delivered by a contextually

As someone with cognitivist sympathies, I feel some discomfort with this sort of conative intrusion into the fixation of the content of evaluative belief and/or the determination of which beliefs get to count as evaluative. We have, sure enough, ordinary garden-variety beliefs with naturalistically kosher contents to serve as evaluative beliefs. But which of these garden-variety beliefs get to count as evaluative is fixed by a bunch of nondoxastic facts—facts about the conative attitudes of the believer. This is perhaps a less thoroughgoing cognitivism than we might have wanted.

It is not a terribly choate discomfort, and I am a bit concerned that I have not quite properly put my finger on the worry here. But I hope that I have said enough to get the idea across and perhaps to enable someone else to put their finger more squarely on what I think is a real concern in the neighborhood. Since I am not confident that I have quite identified the concern, I am also prepared to be talked out of it and to be convinced that I should not be worried. But as of now, it does give me pause.

Here is a worry that is perhaps clearer, and perhaps more serious: it is not clear that a contextualist dispositional theory will give us all of the internalism that we want. Allow me to distinguish two kinds of internalism. A *content-internalist* view is one according to which there is a tight connection between motivation and having a belief with a certain content. There are some potential objects of belief such that believing one of *those* is tightly tied up with motivation. On the other hand, there is *belief-state internalism*. On this sort of view, the tight connection is not between motivation and beliefs with a certain content but, rather, between motivation and Perry-style (1979, 1980) belief-states with a certain character. One way to think about this is that, on the belief-state internalist model, it is deploying distinctively evaluative or normative modes of presentation, rather than believing distinctively evaluative or normative contents, that is bound up with motivation. The kind of contextualist account now under consideration delivers belief-state internalism, but it does not deliver content-internalism. Again, maybe this is something that we are not really entitled to have pretheoretical preferences about, but I would rather have gotten a distinctively evaluative thing to think, such that thinking *that* has a distinctive connection to motivation, rather than a distinctively evaluative (and distinctively motivating—because affectively/

---

selected moral system. It does, however, share the relevant feature of contextualist dispositionalist accounts: it fragments the subject matter of various people's moral judgments and locates the evaluativeness (in this case, the moral-ness) of the belief, not in its content, but in its mode of presentation. So while we can both have beliefs about what is right-according-to-system-14, only I can have *moral* beliefs with that content, since I have, and you lack, the right sort of conative connection to system 14.

conatively is loaded) way to think thoughts that do not themselves have anything distinctively evaluative, or distinctively motivational, about them.

The preceding worries are, I am afraid, not as fully worked out and nailed down as I would like. I put them forward in large part in the hope that somebody else will pick up on them and either whip them into shape or show why they are misguided and confused. Another, less theoretically loaded worry about contextualism about the evaluative is the one that gets most of the press: the worry that, on account of the way that it fragments the subject matter of evaluative belief and assertion (such that different people's and groups' evaluative thought and talk will be about very different families of properties), a contextualist account is going to prevent us from delivering disagreement, conflict, and incompatibility for incompatible-looking evaluative beliefs and assertions.[24] I say (or believe) *x* is a value, meaning it is a value$_{K1}$. You say (or believe) it is not a value, meaning it is not a value$_{K2}$. It looks as if what we have here is talking-past (or thinking-past), not disagreement. There does not seem to be any conflict, or any incompatibility, between our thoughts and assertions. And this seems wrong—when I say or think *x* is a value, and you say or think it is not, our assertions or thoughts should be in conflict, should be incompatible, and our difference should count as a disagreement.[25]

There is a *lot* more to say at this point. There are, in particular, a number of moves to make in defense of the contextualist picture and its ability to deliver the requisite kind of conflict, incompatibility, and disagreement.[26] But it would be nice not to have to do any fancy footwork here. It would be nice to have a story that just straightforwardly delivered disagreement, conflict, and incompatibility. It would be nice to be able to tell a story that delivers a common subject matter for everybody's evaluative thought and talk and avoids the subject matter fragmentation that comes along with going contextualist. And it would be nice to be able to do that without having to fight the hard and uncertain battle for universal convergence. This is the cluster of

[24] For discussion of this objection to contextualist accounts of normative discourse, see for example Robinson 2009, Dreier 2009, Brogaard 2008, Kolodny and MacFarlane 2010, and Plunkett and Sundell, MS. For arguments that contextualists cannot account for disagreement in other domains, see Lasersohn 2005; Stephenson 2005, 2007; Kölbel 2002, 2003, 2008; and MacFarlane 2011, 2012.

[25] This sort of objection has often been made to contextualist theories in various domains. See, e.g., Wright 2001; Lasersohn 2005; MacFarlane 2011, 2012; Kölbel 2002, 2003, 2008; and Egan 2010.

[26] Some particularly promising examples include Dreier 1990; Sundell 2011; Plunkett and Sundell, MS; López de Sa 2008; DeRose 2004; Marques MS; Finlay and Björnsson 2010; Dowell forthcoming, MS; Cappelen and Hawthorne 2009; Brogaard 2008; and von Fintel and Gillies 2011.

theoretical itches that I think (and will argue) that a *de se* version of a dispositional theory of value is well-suited to scratch.

To sum up, standard kinds of dispositional theories are faced with a hard, "who are *we*?" question. The leading candidate responses are "everybody" and "it depends." Saying "everybody" gives you a universalist, invariantist view and commits you to either accepting an error theory or defending universal convergence. Saying "it depends" leaves you without a common subject matter for everybody's evaluative beliefs and without a distinctively evaluative subject matter for anybody to have beliefs about. As a result of this fragmentation of the subject matter of evaluative thought and talk, you have to do some fancy footwork in order to recover the possibility of genuine disagreement about evaluative matters by people who occupy relevantly different contexts. It would be nice to have an option that avoids the question.

4. A *DE SE* DISPOSITIONALIST ACCOUNT OF EVALUATIVE THOUGHT AND HOW IT HELPS

The standard menu of dispositionalist options considers only possibilities on which the content of evaluative thought and talk falls on the *de dicto* side of Lewis's (1979) distinction between the *de dicto* and the *de se*. We can, I think, avoid the standard problems around convergence by looking to the other side of that divide. Rather than saying that the content of a belief that *x* is a value is a possible worlds proposition of the type, *that x is disposed to elicit R in Ks in C*, we say instead that it is a property, *being disposed to have R in response to x in C*. (In alternative notation: $\lambda y.x$ *is disposed to elicit R in y in C*.) So we analyze the belief that *x* is a value not as a belief, about the particular individual or group, that they are disposed to have R to *x* in C but, rather, as self-attribution of the property of being disposed to have R to *x* in C. It will be helpful to offer a bit of explanation and motivation of the *de se* framework in order to make clearer just what is being proposed.

Start with an initially attractive way of thinking about belief (and other propositional attitudes, but let us focus here on belief): the objects of belief (and other propositional attitudes—this clarification hereafter omitted) are possible-worlds propositions, and the only taxonomy of doxastic states that we need is one that classifies believers based on which possible-worlds propositions they believe. Characterizing a believer's doxastic state is a matter of specifying the doxastic alternatives she leaves open—to believe that P is to be in a doxastic state that leaves open only P-alternatives as live options. One of the central lessons of Lewis (1979) and Perry (1979) is that a taxonomy of states of *believing P* for possible-worlds propositions P (a taxonomy according to which characterizing an agent's doxastic state is a matter of specifying the

doxastic alternatives left open, and the doxastic alternatives that the agent's beliefs distinguish between are possible ways for the world to be) does not allow us to capture all of the doxastic states that we want to capture. Let us quickly run through some examples.

> There is a distinctive doxastic state that Professor Procrastinate gets into when he realizes that the deadline is *tomorrow* (*today*, *two weeks ago*), even though he has believed all along that the deadline is April 1. It has a kind of connection to (urgent, desperate) action that the belief that the deadline is April 1 does not have.
>
> There is a distinctive doxastic state that is in common to all of the people who sincerely assert "my pants are on fire" and go in for that distinctive pattern of jumping, screaming, and looking for fire extinguishers to turn on themselves and/or lakes to jump into. But it is not a state of *believing P* for any possible-worlds proposition P.
>
> When Bob and Judy are out hiking and Bob is attacked by a bear, there is a doxastic difference between them that gives rise to their marked difference in behavior (Bob curls up into a ball; Judy unlimbers her rifle), even though they are alike with respect to the (relevant) possible-worlds propositions that they believe. (Both believe *that Bob is being attacked*, *that Judy has a rifle*, etc.)

In all of these cases, we have people in doxastic states that we want to be able to talk and theorize about but that do not appear in a taxonomy of *believing P* states where P is always a possible-worlds proposition. Lewis proposes a way to accommodate the phenomena: we should revise our picture so that the objects of belief are *properties*, not possible-worlds propositions. And we should take the doxastic alternatives that our beliefs distinguish between to be, not possible worlds, but possible positions, situations, or predicaments within worlds. Belief is, in the first instance, self-attribution—to believe P, where P is a property, is to leave open only P-predicaments as live doxastic alternatives.[27] The right taxonomy is a taxonomy of states of *believing P*, where P is a property (or, equivalently, a centered-worlds proposition).

If we move to this picture, we can still capture all the doxastic states we could capture before—where we used to say that Bob and Judy both believe the possible-worlds proposition *that Bob is being attacked by a bear*, we say that Bob and Judy both believe (i.e., self-attribute) what we can call a *world-occupancy property*: *being in a world in which Bob is being attacked by a bear*, or *being such*

[27] The terminology of "self-ascription" is traditional but potentially misleading. Do not read too much in to it—we can (and should) understand "self-attributes F" talk as meaning just, *leaves open only F predicaments* (only <*w*,*t*,*i*> triples such that *i* is F at *t* in *w*) as doxastic alternatives.

that Bob is being attacked by a bear. But now we can also capture the target doxastic states that evaded the possible-worlds taxonomy: the difference between Bob and Judy is that Bob, but not Judy, self-attributes *being attacked by a bear*. What is in common to the well-informed people with burning pants is that they all self-attribute *having burning pants*. What motivates Professor Procrastinate's spurt of desperate activity is coming to self-attribute *being contemporaneous with the deadline*, or (a bit later) *occupying a temporal position two weeks after the deadline*.

Now, let us apply this apparatus to the task of building a dispositional theory of value. Focusing on the case of evaluative belief: rather than saying that the content of a belief *that x is a value* is a possible-worlds proposition is true iff *x* is disposed to elicit R in Ks in C, say that the content is *being disposed to have R in response to x in C* (i.e., λy.*x is disposed to elicit R in y in C*).

This *de-se*-ified version of a dispositional theory of value retains the good features that were selling points for dispositional theories in the first place. It is still cognitivist: evaluative belief is still paradigmatic belief, with straightforward content of the same kind as other paradigmatic cases of belief. (And on the *de se* account, evaluative belief is marked off by its recognizably evaluative subject matter.) It is still naturalistic: there is no need to wheel any nonnaturalistic properties into your metaphysics to accommodate it. It is still (at least as) suitable for securing internalism. (In fact, it is probably better for securing internalism. In general, there is a worry about getting from my belief *that Ks are disposed to have R to x*, to motivation, without a further *de se* belief that I am a K. There is no such worry for the *de se* version of the dispositional theory.) It is still suitable for avoiding commitment to an error theory: if we choose our Rs and Cs carefully, it is going to be quite plausible that we actually have a lot of the properties that we self-attribute in our evaluative thought.

The main payoff for the move to the *de se* picture is that it avoids the problems about nonconvergence. This kind of view allows us to avoid commitment to an error theory, without relying on universal convergence, and while retaining a single, unified subject matter for everybody's evaluative beliefs. Not only do everybody's evaluative beliefs have a common subject matter, but it is the content of the beliefs that makes them evaluative—there are some things to believe such that any belief with one of those contents is an evaluative belief. It is just that the subject matter is *de se*.

I would like to go on to say that because commonality of subject matter is straightforwardly preserved, so too are disagreement, conflict, and incompatibility in evaluative thought and talk. Sadly, things are not so simple.

Let us look at the issues about evaluative thought first. We can start with some good news: *incompatibility* in thought is straightforwardly preserved. If I think *x* is a value and you think it is not, you cannot come to believe what I

believe without changing your mind. (See Kölbel [2004] for similar ideas, implemented in a different framework.) That is not nothing—it is real progress over the contextualist version of dispositionalism.

But (now the bad news) it is also not everything. Incompatibility of *de se* belief is not sufficient to deliver anything that clearly deserves the name 'disagreement'. There is an incompatibility in thought when I think Sydney is nearby and you think it is far away (i.e., when I self-attribute *being near Sydney* and you self-attribute *being far from Sydney*), when I self-attribute *having burning pants* and you self-attribute *not having burning pants*. You could not come to believe what I believe without changing your mind. But here, it is very clear that there is no *conflict*, no *disagreement* in belief (Dreier 2009). (So, interestingly, the three things I was grouping together before when setting up the puzzle—incompatibility, conflict, and disagreement—can pull apart when we start looking at the *de se*.)

Incompatibility in *de se* belief does not, in general, make for conflict or disagreement. But incompatibility in evaluative belief *does* make for conflict and disagreement. So if we are going to give an account according to which evaluative belief is *de se*, we had better have something to say about what makes *this* kind of difference in *de se* belief so special. This will be the task of section 6.

The first worry about evaluative *talk* is that it is not immediately transparent how to understand the view that the content of evaluative *language* is *de se*. The *de se* has its home, in the first instance, in philosophy of mind. We need some explanation, in particular, of what asserting something *de se* amounts to. And we need an explanation according to which we preserve the phenomenon of conflict and disagreement at the level of evaluative language. It will be convenient to take up the questions about language first, in the next section, before moving on to the question about disagreement in thought.

5. *DE SE* CONTENT IN ASSERTION, AND DISAGREEMENT AND CONFLICT IN EVALUATIVE TALK

If we give a *de se* analysis of the content of evaluative thought, it is important that we also provide a story about evaluative *talk* that tells us how linguistic communication dealing in these kinds of contents works. The account I favor is a package deal: a story about the content of evaluative sentences, together with a particular story about the theoretical significance of attributions of semantic content.

Evaluative sentences get assigned the same kinds of *de se* contents as evaluative thoughts. On the kind of account under consideration, "happiness is a value," for example, will turn out to express, on a *de-se*-ified version of

Lewis's view, *being disposed to value happiness under conditions of full imaginative acquaintance*. (More generally, *λx.happiness is disposed to elicit R in x in C*.)

The accompanying account of the theoretical role of content is broadly Stalnakerian (Stalnaker 1978). Attributions of semantic content to (unembedded, declarative) sentences in context specify the *acceptance conditions* for an assertion of the sentence in that context. The semantic content of a sentence S in context *c* is what is added to the conversational presuppositions by a successful assertion of S in *c* (that is, an assertion that all the other parties to the conversation go along with). Something is presupposed just in case all the parties to the conversation believe it, take each other to believe it, take each other to take each other to believe it, etc. (This is not quite right. Presuppositions need not always be *believed*. Sometimes they are just imagined or accepted for purposes of the conversation. I will mostly talk in terms of belief in what follows because it makes the discussion more straightforward, but everything could be restated in terms of acceptance.) In order to sincerely accept your assertion of P, I need to do my bit toward adding P to the conversational presuppositions. So I have to come to believe it, to believe (absent any signals that other parties to the conversation are dissenting) that all the other parties to the conversation believe it, etc.

Here is another, nearly equivalent way to think about acceptance conditions—and in fact the way that I officially endorse, for reasons that we will not be going into here: the semantic content of S in *c* is what hearers are conventionally called upon to take on board (to accept for purposes of the conversation) in order to go along with an assertion of S in *c*. This makes the same predictions as the presupposition-based account for intra-conversational effects—on this view, the way that the contents of assertions get added to the presuppositions is by everybody going along, noticing everybody going along, etc.—but makes different predictions about, for example, cases of inter-conversational assessment (i.e., eavesdropper cases).[28]

So to say that the content of an English sentence "happiness is a value" is some property—call it "F"—is to say that assertions of "happiness is a value" are bids to add F to the conversational presuppositions. That is, they are bids to add F to the stock of things that all of the parties to the conversation believe, take each other to believe, etc. Recall that, in a *de se* framework, to believe a property is to self-attribute it. So if such assertions are successful, we will wind up in a conversational context in which all the parties to the conversation self-attribute F, take each other to self-attribute F, etc. (I will say more about this in a moment, but it is worth flagging now: if this is how we

[28] For more on acceptance-conditions based conceptions of semantics, see Egan 2007, 2010, MS.

are thinking about the theoretical significance of attributions of *de se* content, we are certainly *not* going to want to attribute *de se* content in the first place you might think to look for it—sentences containing first-person indexicals, like "my pants are on fire" or "I am John Malkovich." In the current framework, such attributions of *de se* content make *extremely* bad predictions about the communicative roles of those sentences.[29])

On a *de se* dispositionalist theory of value, the role of evaluative discourse is going to be to get the participants in the conversation on the same page with respect to how they think they would respond to the objects of evaluation under the appropriate conditions. (For example, on the *de se* version of Lewis's view, the role of evaluative assertion will be to get the parties to the conversation aligned with respect to what they think they would desire to desire under conditions of full imaginative acquaintance.)

Given this story about the role of semantic content in a broader theory of linguistic communication, the story about disagreement and conflict in evaluative *talk* is straightforward. When I assert "happiness is a value" and you assert "happiness is not a value," I am trying to add a property F to the common ground, and you are trying to add its complement. We are trying to update the context in incompatible ways—trying to get our interlocutors to accept things that they cannot accept both of.

So, good news for *de se* dispositionalism as a theory of evaluative language: we get a unified, evaluative subject matter for evaluative discourse to be about, and we preserve straightforward disagreement. Those are good results, and better results than we get from contextualist dispositionalist theories.

There are three things to note before moving on to the more problematic case of disagreement in evaluative thought.

First, it is *only* given this Stalnakerian story about the role of semantic content that you get the straightforward story about conflict and disagreement. If, for example, we say that the role of content is to capture *production conditions*—what has to be true of, believed by, known by, etc. the speaker—then we do not get conflict or disagreement. On this kind of picture, all that is required for sincere assertion of F is that the speaker takes herself to have (knows she has, etc.) F, and all that is required for audience members to sincerely accept the assertion is that they accept that the *speaker* has F—not that the audience members self-attribute F themselves. This is not good for our purposes. For one thing, it is very clear that we do not get anything that looks like disagreement, conflict, or incompatibility—there is just no problem at all, on this sort of view, with accepting both my assertion that happiness is a value and your assertion that it is not. All one needs to do is

[29] More discussion coming shortly. See also Egan 2007, 2010.

accept that I have, and you lack, the property that is the content of my assertion. For another thing, it is not a very interesting theoretical option. It is *very* hard to separate the predictions of this sort of view from the predictions of a (solipsistic) contextualist theory, according to which the relevant group is always just the speaker. So the story about semantic content and the story about the role of content in a theory of communication are a package deal. Taken together, they deliver what I take to be some pretty desirable results. But the story about what the semantic content of evaluative sentences *is* does not get you the result without the accompanying story about what semantic contents *do*. (That should not be surprising. In general, stories about what Xs are depend for much of their predictive significance, and hence for many of their virtues, on an accompanying story about the theoretical role that Xs play.)

Second, this is a deviation from the way of explaining disagreement that Dreier (2009) recommends. In Dreier's picture, the way to explain a disagreement between my assertion of some sentence S1 and your assertion of some sentence S2 is to locate a disagreement between the state of mind I express with S1 and the one you express with S2—the story about disagreement in language is going to be grounded in a story about disagreement between the states of mind people use the language to express. This is, obviously, a completely noncrazy picture of disagreement in language. But it is not the *only* picture of disagreement in language, and while I think it is probably the right picture of a lot of disagreements in language, it is not the right picture of *all* of them.

The story about disagreement in evaluative discourse outlined above is emphatically *not* one that fits Dreier's model. It is not a story in which the disagreement in speech is explained by first identifying a disagreement between the states of mind expressed. The story about disagreement here is distinctively linguistic—the conventional uptake conditions for the two sentences are incompatible. All this requires is *incompatibility* (not anything that we would antecedently want to identify as *disagreement*) between the state of mind that one has to get into in order to sincerely accept an assertion of "happiness is a value" and the state of mind one has to get into in order to sincerely accept an assertion of "happiness is not a value." Mere incompatibility between states of mind can then give us conflict, and disagreement (in at least one legitimate sense of "disagreement"), between the assertions and/or between the parties making the assertions. What this shows, I think, is that there are more ways to skin the disagreement-securing cat than we might have thought. (Better: there are more kinds of phenomenon that plausibly deserve the name "disagreement" than we might have thought, and so more ways of skinning the intuitive-sense-of-disagreement-vindicating cat than we

might have thought.) The sort of states-of-mind-first picture that Dreier (2009) advocates is one route to take, but it is not the only one.

The third thing that I want to emphasize at this point is that, given this kind of acceptance conditions-based story about the theoretical role of content in an account of assertion and communication, we definitely *do not* want to go for a semantic theory that assigns *de se* content to indexical sentences. That combination is a big disaster—one of the candidates for the title of "worst possible theory of indexicals."[30]

Such a theory would tell us that the sentence "I am wounded" expresses the property, *being wounded*. This makes very bad predictions about the communicative role of the sentence. The effect of Gustav's assertion of "I am wounded" (if it were accepted) would be to add *being wounded* to the conversation's presuppositions. Part of what would be involved in this would be all of the parties to the conversation self-attributing *being wounded*. But this is emphatically *not* what happens when people assert "I am wounded." The way the English sentence actually works is that Gustav's assertion, if accepted, adds *that Gustav is wounded*[31] to the presupposition, while Gottlob's assertion of "I am wounded" adds *that Gottlob is wounded* to the presuppositions, and so on. Given the conception of the theoretical role of semantic content that we are working with, the right semantic theory for "I" in English is the familiar Kaplanian one.

Other attributions of *de se* content to sentences involving indexicals are similarly disastrous. So we ought *not* to believe that indexical sentences have self-locating content. We ought instead to believe the usual sort of Kaplanian theory.

(How, then, do we explain the tight connection between sincerely asserting "I am wounded" and self-attributing *being wounded*? We can do so by exploiting a very minimal and unambitious sort of two-dimensionalism. Given the standard Kaplanian semantics, there will be some people who are in a position to speak truly by saying, "I am wounded," and some people who are not. If we take the speaker to be sincere and well-informed, we trust that they are among the individuals who are in a position to say something true, rather than something false, with an utterance of "I am wounded." The people in a position to say something true with an assertion of "I am wounded" are all

[30] See Stojanovic, MS, Ninan 2010, Torre 2010, and Kindermann 2012 for advocacy of *de se* content for sentences with indexicals. Crucially, all of these authors are working with pictures of the theoretical role of semantic content that are different in important respects from the one I am assuming here, and so I *do not* think that the package deals that they offer are candidates for being the worst possible theory of indexicals. I do not think that they are going to be correct at the end of the day, but I think they certainly deserve a place at the table.

[31] Or, *being such that Gustav is wounded*.

and only the wounded people. And so it is a prediction of the off-the-shelf Kaplanian semantics for first-person indexicals that it is only people who self-attribute *being wounded* who will sincerely assert "I am wounded," and that audience members will be able to recognize this.)

6. DISAGREEMENT IN THOUGHT

Difference of opinion with respect to value is disagreement—even unvoiced difference of opinion, which never manifests itself in assertion or argument. If I think happiness is a value and you do not, we thereby disagree. But difference of opinion with respect to *de se* matters is not, in general, disagreement. If I think Sydney is nearby and you think it is far away (if I self-attribute *being near Sydney* and you self attribute *being far from Sydney*), we do not thereby disagree. So if we want difference of opinion about value to be disagreement, while analyzing it as a difference of opinion with respect to a *de se* subject matter, we need a story about why *this* sort of incompatibility in *de se* belief makes for disagreement. There are at least three options.

The first is to take a hard line and insist that disagreement in assertion, and mere incompatibility in thought, really is enough. I am not terribly happy with this response. But I do not think it should be dismissed out of hand. Even if all we get from the move from a contextualist to a *de se* version of a dispositional theory of value is a common subject matter, straightforward disagreement in conversation, and straightforward incompatibility in thought, that is not an insignificant bunch of advantages. (Though the fight about whether the contextualist story delivers disagreement in conversation is more complicated than I have let on. See references from endnote 22.)

The other two options are ways of trying to deliver (enough) disagreement in thought, rather than denying that it is something we need to deliver. The first such strategy is to note that mere incompatibility in *de se* thought *does* make for straightforward disagreement when it is combined with an assumption, on either side, of relevant similarity (i.e., of similarity with respect to the property in question). If you self-attribute F, and I both self-attribute the complement of F and *also* think that we are alike with respect to F-ness, then we are plainly disagreeing about which properties you have. We might also—following a suggestion by Mark Richard (MS)—count as disagreeing because, while neither of us believes that we are now similar in the relevant respect, we *ought* to be, or believe that we ought to be, similar in the relevant respect.

It is plausible that many cases of evaluative disagreement will be like this—at least one party will be assuming relevant similarity, or assuming that

there *ought* to be relevant similarity.[32] It is going to be contentious whether this gets us all of the disagreement that we want—it does not get us disagreement in cases of mutually known, and mutually endorsed, nonconvergence, for example. How badly we should want to preserve disagreement in these cases is, I think, debatable, so I think we should take seriously the possibility that this kind of account will deliver all of the disagreement that we ought to want, or at least all that we are entitled to insist on. But I confess to not being completely comfortable with this reply, and I think that it is also plausible that this kind of story will not really get us everything that we want by way of disagreement.

The second strategy for securing disagreement is to borrow a page from expressivist metaethics. Expressivists like Stevenson (1944, 1963) and Gibbard (2003) emphasize that disagreement need not always be cognitive. There is also a phenomenon of *disagreement in attitude* (or, in Gibbard's 2003 framework, *disagreement in plan*). At least on many ways of spelling out a *de se* dispositional theory of value, incompatibility in evaluative belief *will* reliably give rise to a disagreement in attitude (or a disagreement in plan). And (the proposal has it) that is a suitable way of marking off a boundary between incompatibilities in *de se* belief that make for disagreement and those that do not.[33] Lots of incompatibilities in *de se* belief—for example, incompatibilities in geographical *de se* belief—do not give rise to the relevant kinds of divergences in attitude or plan. What marks off incompatibilities in evaluative belief as special, on this account, is that they *do* (predictably and regularly, but probably not invariably) bring such disagreements in attitude in their wake.

7. CONCLUSION

Dispositional theories of value have a lot of features that make them very attractive to a pretty broad philosophical audience. But they have a big problem: picking the relevant group whose dispositions are being tracked by (a given person's or group's) evaluative thought and talk.

Going universalist—saying that it is *everybody*—threatens to force us into an error theory. It is not going to be straightforward to secure the convergence of absolutely everybody's relevant dispositions, while retaining the theory's other attractive features (e.g., naturalism and internalism). The other standard option that has been on the table to date is to go contextualist, selecting

[32] For discussions of presuppositions of relevant similarity, see Egan 2007, 2010 and de Sa 2008. For some criticism of de Sa's appeal to this sort of move, see Baker 2012.

[33] For expressivist advocacy of disagreement in attitude/disagreement in plan, see, e.g., Stevenson 1944, 1963, Gibbard 2003, and Yalcin, MS.

a different, narrower group, among the members of which there *is* convergence, on different occasions of evaluative thought or talk. One worrying question about this is how to select the group such that there will be the right kind of local convergence. Another worry is the loss of a common subject matter, and a distinctively evaluative subject matter, for evaluative thought and talk. And this fragmentation of the subject matter of different thinkers' and talkers' evaluative beliefs and assertions gives rise to a worry that the theory will not be able to explain the fact that difference in evaluative belief gives rise to (or just is) conflict and disagreement.

We can avoid this cluster of problems if we adopt the sort of *de se* relativist account I have sketched above. On that sort of picture, the content of evaluative judgment and assertion is not about any particular individual or group. (The content is just the property of having a certain disposition, not the proposition that any particular individual or group has the disposition.) But we buy ourselves another problem: making sense of disagreement and conflict over the *de se*.

This is straightforward for disagreement and conflict in evaluative *talk*. Makers of incompatible evaluative assertions are attempting to update the context in incompatible ways. It is less straightforward for evaluative *thought*. I have surveyed some responses above that I think are promising.

One complicating fact is that many of these disagreement-recovering moves have parallels that are available to the contextualist. This complicates the issues as far as weighing the two stories against each other. There are some advantages to a *de se* that are worth emphasizing, though. First, the *de se* account *does* straightforwardly deliver incompatible contents for the evaluative thoughts that we wanted incompatibility between, in a way that the contextualist just does not, and cannot. Second, the *de se* account delivers a very clean, completely straightforward account of disagreement in assertion.[34] It is also possible that the advocate of a *de se* dispositionalist story over a contextualist one will, in the end, need to rest more weight on the concerns, which I initially tried to resist putting too much weight on, about whether or not we have secured a common subject matter, or whether we have secured content-internalism rather than just belief-state internalism. (And the accompanying worry about whether belief-state internalism really gets us all of the cognitivism we wanted.)

I am not at all confident that a *de se* dispositionalist theory of value is correct. I am more confident (but still not *terribly* confident) that if any dispositional theory of value is correct, it is going to be a *de se* one. But I am

[34] Also a more general account of disagreement in assertion, which also applies to assertions that do not take place in the same conversation. More on this in Egan, MS.

very confident—and I hope that I have convinced you—that *de se* versions of a dispositional theory of value have enough going for them that they deserve a seat at the table alongside their universalist and contextualist siblings.[35]

REFERENCES

Baker, Carl. 2012. Indexical contextualism and the challenges from disagreement. *Philosophical Studies* 157: 1–17.

Blackburn, Simon. 1984. *Spreading the word*. Oxford: Clarendon Press.

———. 1993. *Essays in quasi-realism*. Oxford: Oxford University Press.

———. 1998. *Ruling passions*. Oxford: Oxford University Press.

Brink, David O. 1989. *Moral realism and the foundations of ethics*. Cambridge Studies in Philosophy. Cambridge: Cambridge University Press.

———. 1997. Moral motivation. *Ethics* 108: 4–32.

Brogaard, Berit. 2008. Moral contextualism and moral relativism. *Philosophical Quarterly* 58: 385–409.

———. 2012. Moral relativism and moral expressivism. *Southern Journal of Philosophy* 50: 538–56.

Cappelen, Herman, and John Hawthorne. 2009. *Relativism and monadic truth*. Oxford: Oxford University Press.

Cuneo, Terrence. 1999. An externalist solution to the "moral problem." *Philosophy and Phenomenological Research* 59: 359–80.

Daly, Chris, and David Liggins. 2010. In defence of error theory. *Philosophical Studies* 149: 209–30.

DeRose, Keith. 2004. Single scoreboard semantics. *Philosophical Studies* 119(1–2): 1–21.

de Sa, Dan López. 2008. Presuppositions of commonality: An indexical relativist account of disagreement. In *Relative truth*, ed. Manuel García-Carpintero and Max Kölbel, 297–313. Oxford: Oxford University Press.

Dowell, Janice. Forthcoming. A flexible contextualist account of epistemic modals. *Philosophers' Imprint*.

———. MS. Flexible contextualism about "ought." Available at: http://unlcms.unl.edu/philosophycollegeofartssciences/departmentofphilosophy/FCOJacksonCasesNB.pdf.

Dreier, James. 1990. Internalism and speaker relativism. *Ethics* 101: 6–26.

———. 2000. Dispositions and fetishes: Externalist models of moral motivation. *Philosophy and Phenomenological Research* 61: 619–38.

———. 2009. Relativism (and expressivism) and the problem of disagreement. *Philosophical Perspectives* 23: 79–110.

Egan, Andy. 2006. Secondary qualities and self-location. *Philosophy and Phenomenological Research* 72: 97–119.

———. 2007. Epistemic modals, relativism, and assertion. *Philosophical Studies* 133: 1–22.

[35] Thanks to Tyler Doggett, Alexandra Plakias, Ivan Milic, David Plunkett, the editors and referee for this volume, and the members of the Rutgers metaethics reading group for feedback on drafts of this paper. Special thanks to Bob Beddor for editorial assistance and extremely helpful comments.

———. 2010. Disputing about taste. In *Disagreement*, ed. Ted Warfield and Richard Feldman, 247–92. Oxford: Oxford University Press.
———. MS. Two Euthyphro questions in semantics.
Enoch, David. 2011. *Taking morality seriously: A defence of robust realism*. Oxford: Oxford University Press.
Finlay, Stephen, and Gunnar Björnsson. 2010. Metaethical contextualism defended. *Ethics* 121: 7–36.
Gibbard, Allan. 1990. *Wise choices, apt feelings: A theory of normative judgment*. Cambridge, MA: Harvard University Press.
———. 2003. *Thinking how to live*. Cambridge, MA: Harvard University Press.
Hare, R. M. 1952. *The language of morals*. Oxford: Oxford University Press.
Harman, Gilbert. 1975. Moral relativism defended. *Philosophical Review* 84: 3–22.
Jackson, F. 1998. *From metaphysics to ethics: A defence of conceptual analysis*. Oxford: Oxford University Press.
Johnston, Mark. 1989. Dispositional theories of value. *Proceedings of the Aristotelian Society* 63 (Supplement): 139–74.
Joyce, Richard. 2001. *The myth of morality*. Cambridge: Cambridge University Press.
Kindermann, Dirk. 2012. Perspective in context: relative truth, knowledge, and the first person. Dissertation, University of St Andrews.
Kölbel, Max. 2002. *Truth without objectivity*. London: Routledge.
———. 2003. Faultless disagreement. *Proceedings of the Aristotelian Society* 104: 53–73.
———. 2004. Indexical relativism vs. genuine relativism. *International Journal of Philosophical Studies* 12: 297–313.
———. 2008. Motivations for relativism. In *Relative truth*, ed. Manuel García-Carpintero and Max Kölbel, 1–38. Oxford: Oxford University Press.
Kolodny, Nico, and John MacFarlane. 2010. Ifs and oughts. *Journal of Philosophy* 107: 115–43.
Lasersohn, Peter. 2005. Context dependence, disagreement, and predicates of personal taste. *Linguistics and Philosophy* 28: 643–86.
Lewis, David. 1979. Attitudes *de Dicto* and *de Se*. *Philosophical Review* 88: 513–43.
———. 1989. Dispositional theories of value. *Proceedings of the Aristotelian Society* 63 (Supplement): 113–37.
Lillehammer, Hallvard. 1997. Smith on moral fetishism. *Analysis* 57: 187–95.
MacFarlane, John. 2005. Making sense of relative truth. *Proceedings of the Aristotelian Society* 105: 321–39.
———. 2011. Epistemic modals are assessment-sensitive. In *Epistemic modality*, ed. Andy Egan and Brian Weatherson, 144–78. Oxford: Oxford University Press.
———. 2012. Relativism. In *The Routledge companion to the philosophy of language*, ed. Delia Graff Fara and Gillian Russell. New York: Routledge.
Mackie, J. L. 1977. *Ethics: Inventing right and wrong*. London: Penguin.
Marques, Teresa. MS. Disagreement in context: A coordinated proposal.
Ninan, Dilip. 2010. *De se* attitudes: Ascription and communication. Philosophy Compass 551–67.
Parfit, Derek. 2011. *On what matters*. Oxford: Oxford University Press.
Perry, John. 1979. The problem of the essential indexical. *Noûs* 13: 3–21.

———. 1980. A problem about continued belief. *Pacific Philosophical Quarterly* 61: 317–81.
Plakias, Alexandra. 2011. *The good and the gross: Essays in metaethics and moral psychology*. Dissertation, University of Michigan. http://deepblue.lib.umich.edu/bitstream/2027.42/86268/1/aplakias_1.pdf.
Plunkett, David. 2010. *Locating practical normativity*. Dissertation, University of Michigan.
Plunkett and Sundell. MS. Ethical Disagreements.
Prinz, Jesse. 2007. *The emotional construction of morals*. Oxford: Oxford University Press.
Railton, Peter. 1986. Moral realism. *Philosophical Review* 95: 163–207.
———. 2009. Internalism for externalists. *Philosophical Issues* 19: 166–81.
Richard, Mark. MS. What is disagreement? http://markrichardphilosophy.wordpress.com/work-in-progress/
Robinson, Denis. 2009. Moral functionalism, ethical quasi-relativism, and the Canberra plan. In *Conceptual analysis and philosophical naturalism*, ed. D. Braddon-Mitchell and R. Nola, 315–48. Cambridge, MA: MIT Press.
Shafer-Landau, Russ. 2003. *Moral realism: A defence*. Oxford: Clarendon.
Smith, Michael. 1989. Dispositional theories of value. *Proceedings of the Aristotelian Society* 63 (Supplement): 89–111.
———. 1994. *The moral problem*. Oxford: Blackwell.
———. 1997. In defense of "the moral problem": A reply to Brink, Copp, and Sayre-McCord. *Ethics* 108: 84–119.
———. 2002. Exploring the implications of the dispositional theory of value. *Noûs* 36: 329–47.
Sobel, David. 1999. Do the desires of rational agents converge? *Analysis* 59: 137–47.
Stalnaker, Robert. 1978. Assertion. *Syntax and Semantics* 9: 315–32.
Stephenson, Tamina. 2005. Assessor sensitivity: Epistemic modals and predicates of personal taste. In *New work on modality*, ed. Jon Gajewski et al, MIT Working Papers in Linguistics, vol. 51.
———. 2007. Judge dependence, epistemic modals, and predicates of personal taste. *Linguistics and Philosophy* 30: 487–525.
Stevenson, C. L. 1944. *Ethics and language*. New Haven: Yale University Press.
———. 1963. *Facts and values: Studies in ethical analysis*. New Haven: Yale University Press.
Stojanovic, Isidora. MS. *De Se* assertion. http://jeannicod.ccsd.cnrs.fr/docs/00/60/75/81/PDF/deseassertion.pdf.
Streumer, Bart. Forthcoming. Can we believe the error theory? *Journal of Philosophy*.
Sundell, Timothy. 2011. Disagreements about taste. *Philosophical Studies* 155: 267–88.
Svavarsdóttir, Sigrun. 1999. Moral cognitivism and motivation. *Philosophical Review* 108: 161–219.
Torre, Stephan. 2010. Centered assertion. *Philosophical Studies* 150: 97–114.
von Fintel, Kai, and Anthony Gillies. 2011. 'Might' made right. In *Epistemic Modality*, ed. Andy Egan and Brian Weatherson, 108–30. Oxford: Oxford University Press.
Wedgwood, Ralph. 2007. *The nature of normativity*. Oxford: Oxford University Press.
Wright, Crispin. 2001. On being in a quandary. Relativism vagueness logical revisionism. *Mind* 110: 45–98.
Yalcin, Seth. MS. Comments on Schroeder, *Being For*. http://yalcin.cc/resources/yalcin.2012.schroeder.pdf.