

The Hard Problem of Supervenience

Abstract. I argue that the conceptual truth that the normative supervenes on the non-normative has been misunderstood, resulting in sham or incomplete explanations for why it holds. Normative supervenience, correctly understood, poses a significant explanatory challenge even for expressivists, who have long claimed the ability to explain supervenience a primary virtue of expressivism. The conceptual truth that demands explanation is the supervenience of the normative on the *repeatably* non-normative. After rendering the repeatability constraint precise, I show that expressivism is inconsistent with supervenience so defined. This result is derived using the expressivist's preferred semantics for normative terms, together with constitutive facts about the non-cognitive attitudes essentially involved in normative thought and talk. I suggest, ultimately, that the problem is not unique to expressivism: our usage of normative terms seems disciplined by repeatable supervenience, but it turns out to be very hard to say why.

Introduction

Almost everyone writing in metaethics agrees that the normative supervenes on the non-normative.¹ Two things identical in their non-normative features cannot differ in their normative features. Moreover, most philosophers who accept normative supervenience agree that normative supervenience holds as a matter of *conceptual* necessity. Anyone competent with our concept of the normative should see that normative properties cannot vary without some non-normative property varying.

Any metaethical theory that accepts this starting point must explain why normative supervenience holds as a matter of conceptual necessity. There is some question whether metaethical non-naturalists can meet this burden; but most metaethical theories have a credible account.² Expressivists and reductive naturalists, in particular, offer elegant explanations for the conceptual truth of normative supervenience.³ Given this, we can call the challenge of explaining why normative properties supervene on non-normative properties the Easy Problem of Supervenience.

There is a harder problem—the Hard Problem of Supervenience—that demands as much attention, yet has been curiously neglected. It is the problem of explaining the conceptual truth that normative properties supervene on *repeatable* non-normative properties. Roughly, non-normative properties are repeatable when it is possible for numerically distinct individuals to share them. The conceptual truth to be explained is that if two things are alike in all repeatable non-normative respects, they are alike in every normative respect. I argue that the intuitions that

¹ There are a few contrarians as is usually the case in philosophy. See Rosen (MS); Fine (2002).

² Non-naturalist realists are supposed to have a hard time explaining supervenience. See discussion in McPherson (2012) and *infra* Section II.

³ See discussion *infra* Section II.

support ordinary supervenience militate just as strongly in favor of repeatable supervenience. But the two have been conflated in recent discussion, with explanations for ordinary (or unrestricted) supervenience being taken as complete explanations for everything that needed explaining. As will emerge, the problem of explaining repeatable supervenience is hard even for expressivists and reductive naturalists, for whom supervenience was meant to pose no challenge at all.⁴ Partial as I am to expressivism, I will be chiefly concerned in what follows with the problem as it arises for the expressivist.

The problem is this: because repeatable supervenience is a conceptual truth, expressivist semantics for normative terms should entail that a normative judgment inconsistent with repeatable supervenience is incoherent. The judgment's coherence should be ruled out by the correct account of *what it is* to deploy a normative concept. But the semantics don't entail this. According to the expressivist, to make a normative judgment is to express a motivational state: a controlling desire, preference, or plan, or something of the sort.⁵ The incoherence of supervenience-flouting normative judgments thus needs to be explained in terms of constraints on the relevant motivational states. For example, if, *per impossibile*, the judgment that two non-normatively identical individuals are normatively different were coherent, this judgment would amount to the expression of some motivation to treat them differently. So long as this kind of motivation is impossible, this fact rules out the intelligibility of the supervenience-flouting normative judgment. But, I argue, the motivational states we are familiar with are mostly *not* constrained by repeatable supervenience.⁶ Moreover, on the most plausible versions of

⁴ See Blackburn (1971); Smith (2004).

⁵ See e.g. Blackburn (1984); Gibbard (2003).

⁶ I consider and reject a view that concedes that our motivational states are supervenience-flouting but insists that, given the functions of normative language, only those motivational states that respect supervenience count as normative judgments. I show that such accounts that appeal to the functions of normative language do not work unless they illicitly assume the fact that needed explaining. See discussion *infra* Sections III and IV.

expressivism, the motivational states that underwrite normative thought and talk will be supervenience-flouting. So, far from explaining the conceptual truth of repeatable supervenience, expressivism seems inconsistent with the thesis that repeatable supervenience holds as a matter of conceptual necessity.⁷

Since I suspect the Hard Problem to be a general one in metaethics, I do not take this result to be devastating for the expressivist. A claim I will not defend here but do elsewhere is that there is a general puzzle confronting meta-normative theories: our usage of normative terms seems disciplined by repeatable supervenience, and it turns out to be very hard to say why. I suspect that expressivists have made a mistake in thinking that they can provide a deep explanation for this feature of our linguistic practice, and they are not the only ones who have made this mistake.

Section One clarifies the supervenience phenomenon. After stating the repeatability condition on the supervenience base precisely, I offer three arguments in support of the condition. The first is a direct argument for repeatable supervenience that appeals to what is inconceivable. The second and third appeal to the uses to which supervenience has been put in metaethics.

Section Two examines the expressivist's resources for explaining repeatable supervenience. Alan Gibbard's plan-based expressivism provides a useful starting point for the analysis. On Gibbard's view, the non-cognitive attitude that underwrites normative judgments is that of a plan. Further, Gibbard thinks the non-normative properties on which the normative

⁷ Despite some recent criticisms of the expressivist's account of supervenience (Zangwill 1997, 110–11; Sturgeon 2009, 84–87; Dreier 2015), the precise vulnerability I discuss and its general upshot have been missed altogether. These critics fail to press on points where the expressivist's explanation of supervenience seems most vulnerable. Moreover, they appear to be in the minority in expressing any skepticism whatsoever about the expressivist's ability to explain supervenience.

supervenies is the class of *empirically discernable* properties, and one cannot plan to treat empirically indiscernible situations differently.

Section Three demonstrates that the Gibbardian explanation of supervenience fails so long as the class of empirically discernable properties satisfies the repeatability constraint. The explanation only seems compelling on the assumption that there cannot be numerically distinct yet empirically indiscernible situations. But it falters once this assumption is set aside because it is possible to plan to treat numerically distinct yet empirically indiscernible situations differently.

Section Four generalizes the challenge for the expressivist in various ways. I argue that the motivational states essentially involved in normative thought and talk *must* be ones that flout supervenience, on any repeatability-respecting definition of the supervenience base. The result is that it is always meaningful or intelligible, given expressivist semantics, to judge two numerically distinct states of affairs to be normatively different despite their being identical in base non-normative respects.

Section Six takes stock and considers the expressivist's options in light of the Hard Problem. One option is to modify standard expressivist semantics for normative terms by positing a brute fact about such terms—unless one is in the business of expressing motivations that are disciplined by supervenience, one does not count as using such terms intelligibly. The problem with this approach, I argue, is that it takes as brute rather than *explain why* our use of normative terms is disciplined by supervenience.⁸ It goes against the overt explanatory ambitions of expressivists.⁹ A different option, the more attractive one, may be to regard supervenience as an ordinary *normative* truth, rather than a conceptual one. On either way of going expressivists

⁸ I consider and reject attempts to explain this brute fact in terms of the functions of normative language. See discussion *infra* Sections III and IV.

⁹ See discussion *infra* Sections II and IV.

must abandon their claim to a dialectical advantage over their theoretical opponents of having to bite no bullets when it comes to supervenience.¹⁰

1. “The Least Controversial Thesis in Metaethics”¹¹

It would be difficult to find a proposition more widely accepted amongst meta-ethicists than the proposition that if two things share all of their non-normative features then they share all of their normative features. Normative properties—whether of acts, states of affairs, or whatever else—are said to *supervene* on the non-normative properties of their bearer. According to Michael Smith (2004, p8), the truth of supervenience “is accepted by nearly everyone writing on the nature of value” and its denial would be evidence of a very basic conceptual confusion.¹² Arguments for supervenience tend to be rare given the assumption that it is self-evident. One argument—the Argument from Cases—appeals to what seems true in individual cases (we shall see a few more shortly). Consider some act, say an instance of lying, that is judged to be wrong. It seems inconceivable that there could be another act of lying that is identical in *all* non-normative respects—it is told for the same purpose and in the same type of situation, has the same effects, and so on—and yet fails to be wrong.¹³ Fixing the non-normative features of an act of lying seems to settle its normative features. Such cases reflect a more general truth about the character of the normative, or so the argument from cases maintains.

There are different ways of stating precisely what this general truth amounts to, but we can fix on the following as a first approximation:

¹⁰ See Blackburn (1971). See also discussion in Dreier (2015).

¹¹ Rosen (MS).

¹² See e.g. Smith (2004: 208); Blackburn (1985); Enoch (2011). It is easy to multiply examples of theorists across the metaethical spectrum who endorse the thesis.

¹³ See Smith (2004: 225-9) discussing ordinary moral practice; and Hare (1952: 145).

SUPERVENIENCE: $\Box (\forall N \in \alpha)(\forall x)[N_x \rightarrow (\exists B \in \beta)(B_x \ \& \ \Box(\forall y)(B_y \rightarrow N_y))]$, where α is the class of normative properties, β the class of non-normative or base properties, and \Box is a necessity operator.

In English: *necessarily, for any normative property, if that property holds of some x , then there is some non-normative property that x has such that, necessarily, for any y that has the same non-normative property, y also has the normative property.*¹⁴ The relevant non-normative property might be a complex property: it could be the property of being B_1 or B_2 or¹⁵ There is relative agreement that the outermost necessity operator should be interpreted as conceptual necessity. It is supposed to follow simply as a matter of the concepts involved that the connection between normative and non-normative properties holds. The inner necessity operator should be interpreted as metaphysical necessity. The conceptually necessary truth is that it is metaphysically impossible for two things to be non-normatively identical but normatively distinct.¹⁶ Though SUPERVENIENCE is sometimes interpreted as implying a fact about metaphysical dependence—roughly, that a thing’s normative features obtain *in virtue of* its various non-normative features—we won’t be concerned in what

¹⁴ There are weaker versions of the supervenience thesis, concerning the co-variance of normative and non-normative properties within possible worlds. But we will be occupied with the strong formulation in what follows. On different formulations of supervenience, see Kim (1984).

¹⁵ The formulation assumes that non-normative properties are closed under infinitary Boolean operations. Since reductive naturalists deny this, it’s not an entirely neutral formulation. One way of avoiding this problem is by talking about *sets* of normative properties. See Bader (2016). But that still does not get the naturalist on board who *identifies* a normative property with a simple non-normative property (like *being pleasing*), so long as one thinks a property cannot be both normative and non-normative. For a helpful discussion of how to accommodate such views, see McPherson (2012). Our argument does not turn on these complications.

¹⁶ These assumptions about the types of necessity in play are generally agreed upon. See Dreier (2015); Enoch (2011); Blackburn’s (1984); Smith (2004). More importantly, they are explicitly endorsed by expressivists.

follows with anything stronger than the mere modal covariance of normative and non-normative features.¹⁷

My primary concern is with β —the class of non-normative properties on which the normative supervenes—and whether it needs to be restricted or further characterized. I am interested in the weakest nontrivial supervenience claim that stands a chance of being a conceptual truth. The narrower the class of base properties, the stronger the corresponding supervenience claim. But if the class is too narrow, the truth will not hold simply by virtue of the concepts involved (or the meanings of the terms).¹⁸ For example, it may be true that normative facts supervene on the *physical* facts. But this is not a conceptual truth, since it is conceptually possible that a non-physical thing, God, exists, and makes a difference to the normative truths (Sturgeon 2009).¹⁹

The main aim of this section is to demonstrate that there must be *some* constraints on the supervenience base that follow from our very concept of a normative property. Leaving β unrestricted may serve the aim favored by theorists of stating the supervenience thesis as

¹⁷ See Dancy (1981: 367, 380-2); (2004: 86) on the difference between “in virtue of” claims and supervenience.

¹⁸ Too much precision in specifying the relevant non-normative properties would render the claim *synthetic*: possibly true, but not true in virtue of the concepts involved. As Blackburn (1993) points out, it is not a *conceptual* matter precisely which non-normative properties the normative supervenes on. Consider the debate between deontologists and utilitarians. These theorists disagree about whether an act’s *rightness*, a paradigmatic normative property, covaries with the non-normative property of being happiness maximizing. But their disagreement about does not appear to be a conceptual one. The truth or falsity of utilitarianism does not fall out of the very meaning of “rightness.”

¹⁹ The supervenience base for the normative was traditionally characterized in terms of *natural* properties, where a property counts as natural if it is invoked by the natural sciences (microphysical and causal/functional properties, for instance) or would be invoked by the sciences sufficiently idealized. See e.g. Jackson (1998); Smith (2004). See also exchange between Griffin (1996) and Smith (2004). It is sometimes suggested that all the familiar ways of defining the supervenience base render the thesis vulnerable to counterexample. See e.g. Sturgeon (2009). I take these skeptical arguments to show only that stating the claim precisely is hard; not that supervenience is false.

uncontroversially and modestly as possible (McPherson 2012; Rosen MS).²⁰ But it obscures the precise supervenience phenomenon that needs explaining.

Unrestricted supervenience is true but seems trivial.²¹ Consider the claim that the normative facts supervene on the non-normative facts with no restrictions on the latter. Among the non-normative properties are *haecceitistic* properties that can only be instantiated by particular individuals, like *the property of being Fred*. There may also be properties indexed to possible worlds, like *the property of being Fred in the actual world*. It is certainly true that two things cannot differ in normative respects without differing in some such non-normative respect. But that is because things that are alike in all of these respects are *one and the same thing!* The property of *being Fred in the actual world* cannot be had by anyone other than Fred in the actual world, and, trivially, Fred in the actual world is either good or not good—he cannot be both.

Our aim is to find the weakest conceptually true supervenience thesis that is non-trivial and can serve as an important desideratum in metaethics. I claim that *repeatable* supervenience is the weakest such conceptually true supervenience thesis. There are echoes of the repeatability condition in RM Hare's (1952) original discussion of ethical supervenience.²² However,

²⁰ One reason it has seemed tempting to leave the category of relevant non-normative properties unrestricted is to get the moral particularist on board. Moral particularists deny that there are exceptionless general principles linking the normative to the non-normative (Dancy 2004). It is true that my argument for constraining the supervenience base rejects an extreme kind of moral particularism, according to which even the haecceities of situations and possible worlds could matter from the normative perspective. But if I'm right, we shouldn't have been concerned with accommodating extreme particularism. The supervenience claim I defend does leave room for more modest versions of particularism, on which there are no snappy general principles in ethics, but we know that some kinds of non-normative properties are ethically irrelevant.

²¹ The precise sense in which it is trivial will become apparent in the discussion to follow.

²² In particular, repeatable supervenience (in my sense) is entailed by Hare's claim that moral judgments are necessarily general and universalizable. See Hare (1952, 129):

repeatability as I go on to define it is a far weaker constraint on the supervenience base than any previously suggested or defended.

We can state the repeatability condition precisely. Readers not worried about this can skip to the next paragraph. For each $B \in \beta$, B is repeatable when necessarily if x is B then possibly there is a y distinct from x , such that y is B. Our account of the supervenience base should entail that it is always possible for two distinct individuals to be base-identical. In other words, we should be able to say of the B in β that renders supervenience true for a normative property N that it is the case that $\Box_m \forall x (Bx \rightarrow \Diamond_m \exists y (By \wedge y \neq x))$, where \Diamond_m stands for ‘it is metaphysically possible that.’²³ This constraint is stronger than what we need for present purposes. My overall argument will go through so long as there is *some* normative property, N_1 in

As we shall see, all value-judgements are covertly universal in character, which is the same as to say that they refer to, and express acceptance of, a standard which has an application to other similar instances. If I censure someone for having done something, I envisage the possibility of him, or someone else, or myself, having to make a similar choice again; otherwise there would be no point in censuring him.... When we commend an object, our judgement is not solely about that particular object, but is inescapably about objects like it

Hare thought that for a value judgment to be truly universal, it cannot include proper names. See *id.* at 176-177. Repeatability allows that morally relevant non-normative properties might involve essential reference to particulars using proper names. E.g., the property of *being a kicking of Fred* is repeatable in my sense (there are many acts that are kickings of Fred), though a principle that treated the property as ethically relevant would not be suitably universal on Hare’s view. What repeatability rules out is the property of *being Fred* belonging in the supervenience base for the ethical.

²³ There are two ways to implement this constraint. Either β , the set of base properties, is constrained so that $(\forall B \in \beta) \Box \forall x (Bx \rightarrow \Diamond \exists y (By \wedge y \neq x))$. Alternatively, we can include the constraint in the statement of supervenience: $\Box_c (\forall N \text{ in } \alpha) (\forall x) [Nx \rightarrow (\exists B \text{ in } \beta) (Bx \ \& \ \Box_m (\forall y) (By \rightarrow Ny) \ \& \ \Diamond \exists a (Ba \wedge a \neq x))]$. This affirms inter-world duplication. A stronger requirement would affirm *intra*-world duplication: one might also think (for some normative properties) that $\Diamond_m (\exists a \exists b) (Ba \ \& \ Bb \ \& \ a \neq b)$.

α , such that supervenience holds for *that* property and of the B_1 which renders the supervenience claim true we can say: $\Box_m \forall x (B_1x \rightarrow \Diamond_m \exists y (B_1y \wedge y \neq x))$.²⁴ The supervenience phenomenon that calls for explanation, I shall argue, is the supervenience of the normative on a repeatable non-normative property.

Repeatability rules out the inclusion of haecceitistic differences in the supervenience base, of the sort discussed earlier. The base does not include *being Fred* or haecceities of situations or possible worlds, because these cannot be had by distinct individuals.²⁵ Other properties that repeatability rules out include exhaustive qualitative non-normative profiles of individuals of the sort that cannot be shared between distinct individuals, for identity in terms of such particularistic non-normative profiles would entail strict identity.²⁶

The first argument for repeatable supervenience is a direct argument from inconceivability. Given a lie (or a bad person, or any other badness-bearing individual), it seems impossible to conceive of something identical in non-normative terms but different normatively. This imaginative exercise, which is the primary motivation for the conceptual truth of supervenience, standardly involves the failure to imagine that *two* non-normatively identical individuals differ normatively, not the failure to imagine two non-normatively identical individuals. Consider the distinct individuals Fred and Bob who are otherwise qualitatively identical in all non-normative respects. To make their non-normative similarity vivid, suppose that Bob is Fred's clone, both were created at the same time, and they are plugged into a

²⁴ i.e. $\Box_C(\forall x)[N_x \rightarrow (\exists B \text{ in } \beta)(Bx \ \& \ \Box_M(\forall y)(By \rightarrow Ny))] \ \& \ \Box \forall x (B_1x \rightarrow \Diamond \exists y (B_1y \wedge y \neq x))$.

²⁵ One wouldn't be able to say $(\forall B \text{ in } \beta) \Box \forall x (Bx \rightarrow \Diamond \exists y (By \wedge y \neq x))$ because there exists a B—the haecceitistic property of being *this* individual—that can't possibly be had by distinct individuals. Of course, properties that merely involve haecceities may be repeatable, like the property of *being a kicking of Fred*. There are many acts that involve hurting Fred. The repeatability-inconsistent properties are those like *being Fred*.

²⁶ If we assume Leibniz' law: $\forall x \forall y ((\forall F)(Fx \leftrightarrow Fy) \rightarrow x=y)$ then there will always be some qualitative profile that is identity determining (or non-repeatable). But the assumption is not necessary.

computer simulation that feeds them the same experiences. Suppose also that they end up behaving in all the same ways. The only non-normative difference between the two is that Fred is Fred and Bob isn't. Is it conceivable that Fred is a good person, but Bob is not? Surely not. It seems ruled out by the very meaning of the term "good." But this is evidence for the conceptual truth that goodness supervenes on repeatable non-normative properties.²⁷ And, it is sufficient for our purposes that there is some normative property—the ethical property of *being good*—of which it is conceptually true that it supervenes on repeatable non-normative properties.²⁸

The second and third arguments for repeatability appeal to the theoretical significance of supervenience within metaethics. Even though the repeatability constraint has not been made explicit in discussions of supervenience, there are good reasons to think it has been assumed. When we look at the uses to which supervenience has been put and ask which thesis metaethicists had in mind, it looks to be repeatable supervenience in every case.

Consider Simon Blackburn's argument for supervenience. Blackburn (1973) suggests that supervenience should be accepted because it helps explain what he calls the "ban on mixed worlds." Blackburn points out that it is numbingly obvious as anything could be that there

²⁷ Here's another argument for thinking that the type of inconceivability that motivates supervenience involves consideration of two genuinely distinct normative property bearers. Suppose that in trying to imagine individuals that are non-normatively identical we can only imagine a single self-identical individual. This raises the worry that what we really intuit when it seems to us that we cannot normatively distinguish individuals after fixing their non-normative features is not the truth of supervenience but the logical tautology that a single individual cannot both be N and $\neg N$. The imaginative exercise is probative of the truth of supervenience only if it involves imagining two genuinely distinct acts that are non-normatively identical and discovering that these distinct acts must be normatively identical.

²⁸ What about obligations to oneself? If I'm obligated to care about myself simply because of my identity, is that not a counterexample to repeatability? The obligation to care about oneself is not necessarily identity-dependent. The relevant non-normative fact that grounds my obligation to myself is that I am a person with interests, but *being a person with interests* is repeatable.

cannot be non-normative qualitative duplicates in the actual world that differ normatively. And he points out that this is true in the case of *actual* non-normative qualitative duplicates. Two states of affairs in the actual world, one located in the northern hemisphere and the other in the southern, but otherwise non-normatively identical *must* be normatively identical. We can know this without knowing which non-normative features these two states of affairs exemplify. Blackburn argues that the conceptual truth of supervenience explains this fact: the reason *actual* distinct states of affairs that are otherwise non-normatively identical cannot be normatively different is that it is *conceptually impossible* for entities that are identical in this way to be normatively distinct. Blackburn must have had repeatable supervenience in mind, because his explanation of the “ban of mixed worlds” presupposes repeatability. Unrestricted supervenience cannot explain why genuinely distinct but otherwise non-normatively identical states of affairs in the actual world cannot differ normatively, because genuinely distinct states of affairs are not identical in *all* (unrestrictedly) non-normative respects.²⁹ The assumption about repeatability should therefore be common ground between me and my primary interlocutors: expressivists like Blackburn.³⁰

My final argument for repeatability is slightly more involved, in that it relies on *recherché* considerations in modal metaphysics, but is worth the effort. Suppose we read in tomorrow’s newspaper that David Lewis (1968) was right: there is no such thing as strict or literal identity between individuals in different possible worlds. When I consider myself in a different possible world where I am a successful musician, I am not *literally* thinking about myself, but rather a very similar counterpart. Lewisian modal realists are not the only ones who think that the notion of

²⁹ In other words, the actual fact that calls for explanation is $(\forall x)[Nx \rightarrow (\exists B \text{ in } \beta)(Bx \ \& \ (\forall y)(By \rightarrow Ny)) \ \& \ \exists a \exists b (Ba \ \& \ Bb \ \& \ B a \neq b)]$. So, $\Box_C(\forall x)[Nx \rightarrow (\exists B \text{ in } \beta)(Bx \ \& \ \Box_M(\forall y)(By \rightarrow Ny))]$ wouldn’t be explanatory without the repeatability of B.

³⁰ Blackburn obviously accepts it. Gibbard seems to as well as we shall see in a moment.

“trans-world identity” is incoherent.³¹ But the reasons for thinking so are not our concern.³²

What is relevant is that *no one* ever understood the challenge of explaining normative supervenience in a way that turns on a proof that individuals are world-bound. The conceptual truth of supervenience was supposed to function as a dialectical lever that separates good metaethical theories from bad ones. For example, one of the main vulnerabilities of non-naturalist realism is believed to be the view’s struggles when it comes to explaining supervenience.³³ But if *unrestricted* supervenience were the relevant explanandum, then everyone, including the non-naturalist, should be relieved to discover that individuals are world-bound. For we can now explain unrestricted supervenience without assuming anything about the nature of the normative. We can simply appeal to the non-repeatability of the base properties plus the law of non-contradiction.³⁴ The reason it is a conceptual truth that it is metaphysically impossible for two base-identical things to be normatively distinct is that base-identical things are identical *tout court* and $\forall x \forall \phi (\phi x \vee \neg \phi x)$! The problem of explaining the conceptual truth that the normative

³¹ See e.g. Heller (1998), Sider (2001) and Forbes (1982).

³² One might think this because of a commitment to modal realism—one and the same concrete individual cannot be part of two entirely distinct yet concrete worlds. Alternatively, one might accept an *ersatz* view of modality on which only actual individuals have individual haecceities, but denizens of possible worlds (possible individuals) have no such individual essences.

³³ See McPherson (2012).

³⁴ We need to assume that individuals are world-bound—i.e. no strict or literal identity between individuals in different possible worlds—because LNC only applies to individuals at a world. Of course, not everyone rejects trans-world identity, but the point of the argument still stands: supervenience is too easy to explain if B-identity entails strict identity. It can be explained by appeal to (1) B-identity = strict identity; (2) LNC: $\forall x \forall \phi (\phi x \vee \neg \phi x)$; (3) No trans-world identity. Any counter-part theorist or meta-ethicist who denies transworld identity for normative property bearing individuals can explain the phenomenon. Thanks to X for pushing me to make this explicit.

supervenies on the non-normative (a class that includes non-repeatable properties) turns out to be a rather Easy Problem indeed.³⁵

The strength of this final argument lies in its showing not just that unrestricted supervenience is too weak, but that repeatability is precisely what we need to assume to arrive at the weakest non-trivial conceptual truth that can serve as an important desideratum in metaethics.³⁶ It is not like we can give up on the law of non-contradiction. And no one ever thought that assumptions in modal metaphysics concerning identity across possible worlds should bear on the explanation of normative supervenience. Unrestricted supervenience is too easy to explain because it allows non-repeatable properties into the supervenience base, ensuring the literal identity of base-identical individuals, which is the source of the problem.

The strong and interesting thesis that any acceptable metaethical theory must vindicate is that the normative supervenes on repeatable non-normative base (henceforth, I'll take the repeatability condition for granted unless stated otherwise).³⁷ It is remarkable how difficult the challenge of explaining supervenience becomes once we assume such a minimal and eminently plausible constraint on the supervenience base. The explanatory burden will no doubt vary

³⁵ Can the expressivist invoke LNC given that, strictly speaking, on her view the application of normative predicates does not involve the attribution of properties? The next section demonstrates that there is an expressivist-compatible response to the Easy Problem.

³⁶ Conceptual repeatable supervenience is clearly true in many areas. If two things are exactly alike in all repeatable underlying respects, either both are chairs or neither is. Why is that? It is somehow a conceptual truth that what makes a chair a chair is a matter of its repeatable features, including its shape, its purpose, and so on. What seems true in the case of the chair concept is also true in the case of the concept of goodness. Thanks to X for the comparison.

³⁷ The assumption, quite apart from being reasonable, is a dialectically fair one insofar as expressivists like Blackburn (1984) and Gibbard (2003) take the relevant truth to be the supervenience of the normative on the repeatably non-normative.

depending on the details of one's account of the normative.³⁸ But expressivists, too, face a distinctive explanatory challenge, one that has an important general upshot in metaethics.

2. "The Expressivist Side-Step"³⁹

The expressivist attempts to explain SUPERVENIENCE with a distinctive account of normative concepts. To deploy a normative concept, according to the expressivist, is not to predicate a property of a thing.⁴⁰ On a rough-and-ready understanding of the view, there are *no* normative properties on the expressivist picture, at least not in any substantial sense.⁴¹ Though normative judgments, or utterances involving normative terms, share the syntax of ordinary descriptive judgments, they simply express the agent's non-cognitive attitudes: her desires, preferences, or plans.

To illustrate, on Gibbard's (2003) "plan-based expressivism" the relevant non-cognitive attitude is that of a plan. Roughly, to judge that one ought to give 40% of one's income to charity just is to plan to be that generous with one's money. To *assert* that one ought to do so is to express such a plan. A plan is akin to a dominant preference or controlling desire for a

³⁸ Non-naturalists like David Enoch (2011) who construe normative properties as *sui generis* and entirely distinct from the base non-normative properties must explain why two distinct sets of repeatable properties co-vary as a matter of metaphysical necessity and how it is that our concepts align with this robustly metaphysical truth. We return to non-naturalism in Section V.

³⁹ Dreier (2015).

⁴⁰ Gibbard (1990: 7-8): "Normative talk is part of nature, but it does not describe nature. In particular, a person who calls something rational or irrational is not describing his own state of mind; he is expressing it."

⁴¹ We can ignore quasi-realist wrinkles in what follows for ease of discussion. Quasi-realist expressivists are happy to admit talk of normative facts and properties, with a deflated account of such talk. To judge that there is a normative fact that one ought to help the poor is equivalent to judging that one ought to help the poor. Both judgments get the expressivist treatment. The argument does not turn on whether we assume a quasi-realistic version of expressivism.

specified outcome—an outcome characterized in terms of actions and circumstances.⁴² Let $p\{\Phi, C\}$ be a plan to Φ in a circumstance C (entertained under some description). Φ -ing in C can be described as the plan's aim. There can be grander plans, on Gibbard's view, ones that specify actions for multiple circumstances, and even a universal plan that specifies an action for every possible situation: a plan to Φ_1 in C_1 , Φ_2 in C_2 , Φ_3 in C_3 , and so on.⁴³

The expressivist's key explanatory move is to focus on what it is to *judge* supervenience to be true given the special non-cognitivist semantics for normative terms (Gibbard 2003: 90-2; Compare: Klagge 1988; Blackburn 1985).⁴⁴ The judgment that it is a conceptual truth that two non-normatively identical things cannot possibly differ normatively amounts to the judgment that it follows from *what it is* to deploy a normative concept—namely, to express a certain kind of plan—that one cannot apply the concept differently to two non-normatively identical things.⁴⁵ The only thing that stands in need of explanation is why this constraint on normative concepts holds, an explanation that must be given in terms of the nature of planning.

⁴² While planning is supposed to be an activity we all engage in and on which we have an intuitive grip, plans are, on Gibbard's construal, markedly flexible. We can plan not just for actual and contingent situations, but also for situations that it is impossible for us to be in. It is crucial for plans to be flexible in this way if expressivism is to be able to capture the full range of normative judgments. Technically, Gibbard uses sets of what he calls 'fact-plan worlds' to interpret states of mind expressed, where a fact-plan world can be thought as a pair (f, p) , consisting of a state of the world f and a plan for that state. A fact-plan world (f, p) is compatible with the state of mind expressed when you judge that you ought to ϕ in a given situation if the 'fact' component is consistent with what you believe the world is like, and the second element—the plan—has you ϕ -ing. The state of mind is then characterized in terms of the set of all fact-plan worlds compatible with one's view of the world and one's motivations regarding what to do. These complexities can be ignored in what follows.

⁴³ Gibbard (1990: 168-160). See also (1990: 48–53, 68–71).

⁴⁴ Complications arising on quasi-realistic expressivism are helpfully explored in Dreier (2015).

⁴⁵ Dreier (2015) aptly describes the expressivist's approach to SUPERVENIENCE as the "expressivist side step" for avoiding the metaphysical puzzle,

We can state the conceptual constraint that needs explaining precisely. For ease of discussion, suppose the only normative concept on the scene is that of <being what ought to be done>, and that this concept applies to act-circumstance pairs based on their features, where an act either is or is not *what ought to be done* in the relevant circumstance. SUPERVENIENCE entails that it follows from the nature of the ought-concept that it cannot apply differently to two act-circumstance pairs that are repeatably non-normatively identical (henceforth, base-identical or B-identical).⁴⁶ One cannot coherently judge that one ought to ϕ in C_1 but not in C_2 while recognizing the circumstances as B-identical. It would be akin to judging that Tim is a married bachelor. One can of course *assert* “one ought to ϕ in C_1 but not in C_2 and the two circumstances are B-identical.” But the assertion would be meaningless or unintelligible.⁴⁷ The challenge is to explain why such judgments are unintelligible.⁴⁸

The plan-based expressivist attempts to explain their unintelligibility by appeal to constraints on planning derived from *what it is* to plan (2003: p93-8). A certain kind of plan is impossible: a plan to ϕ in C_1 and not ϕ in C_2 while recognizing that C_1 and C_2 are B-identical. If such a plan *were* possible, given the semantics, one could meaningfully judge that one ought to ϕ in one situation but not in another while recognizing the situations as B-identical. But since the plan is impossible, the corresponding judgment is incoherent.

Why is the supervenience-flouting plan impossible? Gibbard’s own argument provides a helpful starting point for the analysis. First, he assumes the class of base properties is the class of scientifically observable properties or as he sometimes puts it: “prosaically factual properties”

⁴⁶ We can ignore the precise derivation. For discussion see Klagge (1988).

⁴⁷ One can, of course, normatively judge base-identical things differently at different times, just as one can meaningfully judge one and the same person to be married and to be a bachelor at different times.

⁴⁸ There may be a related puzzle concerning norms of belief or judgment revision, but we will get to it later when discussing Sturgeon’s (2009) objection against the expressivist.

(pgs. 32, 98-99). The category includes not just properties currently recognized by the natural sciences, but “spooky properties like the properties of ghosts and gods, so long as these can be *recognized* or *observed* and thus could figure in an empirical science.”⁴⁹ Second, he suggests that a plan is the type of motivational state that cannot treat empirically-indiscernible circumstances differently:

[A] plan can distinguish between situations only in terms of their prosaically factual properties, and it can distinguish between acts only in terms of the prosaically factual properties of those acts. If two acts in two possible situations differ in no prosaically factual way, a plan cannot distinguish them, permitting one and ruling out the other. Either it will permit both or it will rule both out (2003: 91-2).

The argument for the claim goes by quickly, but the basic idea is supposed to be that empirically identical (or E-identical) outcomes are quite simply indistinguishable from the planning perspective.⁵⁰ We cannot tell two E-identical circumstances apart. A plan to ϕ in C_1 but not ϕ in C_2 when C_1 is E-identical to C_2 would be *just like* planning to ϕ and not ϕ in what one recognizes to be one and the same situation. And it should be evident from a moment’s reflection on the nature of planning that a putative plan to ϕ and not ϕ in a single situation would be no plan at all! This is because plans, constitutively, aim to guide action and a plan to ϕ and not ϕ in one and

⁴⁹ The notion of recognizability is not altogether clear. Gibbard sometimes refers to the relevant properties as “prosaically factual properties” which suggests that the notion might be quite strong. For example, it might involve *phenomenological* recognizability, where if a property is recognizable then it is directly available in phenomenal experience or else impinges in some way on our sensory surfaces. But more plausibly, the category includes just those properties whose existence can be taken for granted under the norms of ordinary scientific practice.

⁵⁰ Dreier (2015) interprets Gibbard similarly: “It is of the nature of plans that the contingencies for which they are plans are features of the situation that are at least in principle recognizable.”

the same situation, would “preclude offering any guidance on what to do on that occasion”
(p56).⁵¹

So, to sum up, the plan-based expressivist doesn’t just think a supervenience-flouting plan is one our community of planners generally avoids or dislikes. The plan is *impossible*, and the fact of its impossibility is supposed to be knowable *a priori* from introspective awareness of one’s own mental states.⁵² The argument for its impossibility drives a tight parallel between the supervenience-flouting plan and a putative plan to ϕ and not ϕ in a single situation. We can make the key moves of the overall explanatory strategy explicit. What the expressivist needs to explain is why a judgment inconsistent with SUPERVENIENCE is unintelligible. The unintelligibility of the relevant judgment follows from the semantics plus key assumptions about the nature of the motivational state expressed through normative judgment:

1. To judge that one ought to ϕ in C_1 but not ϕ in C_2 and that C_1 and C_2 are E-identical *just is* to express a plan with ϕ -ing in C_1 and not ϕ -ing in C_2 as its aim where C_1 is E-identical to C_2 . The judgment is unintelligible if the associated plan is impossible. [Plan-based Expressivism]
2. A plan that cannot guide action is impossible; hence, the impossibility of a plan with ϕ -ing & not ϕ -ing in C_1 as its aim. [Basic Constraint on Planning]
3. A plan that aims at ϕ -ing in C_1 and not ϕ -ing in C_2 cannot guide action. [No Action Guidance].

⁵¹ Gibbard’s intriguing claims about the nature of planning have been curiously neglected. The only one to express any skepticism as far as I know, and that too in passing, is Hawthorne (2002: 173-4)

⁵² I consider later and rule out the possibility that Gibbard might be understood as *stipulating* an artificial notion of a plan with the constraint built in.

From (1), (2), and (3), we conclude that one cannot coherently judge that one ought to ϕ in C_1 and not ϕ in C_2 while recognizing that C_1 is E-identical to C_2 . Call this the Argument from Planning.⁵³

3. The Hard Problem for Plan-based Expressivism

The plan-based expressivist has no trouble at all with the *Easy* Problem of Supervenience. Suppose the notion of empirical detectability is defined so that the class of empirically detectable properties includes all the properties there are, including bare haecceitistic properties like *being this particular circumstance at this time and location...*, and that, accordingly, it is not possible for two numerically distinct circumstances to be empirically identical. In other words, suppose the supervenience base is defined to be non-repeatable. If C_1 is E-identical with C_2 entails that C_1 and C_2 are strictly identical, then the Argument from Planning seems very compelling indeed. A plan with ϕ -ing in C_1 and not ϕ -ing in C_2 as its aim where C_1 is E-identical to C_2 really does amount to a plan with ϕ -ing & not ϕ -ing in C as its aim. And it is very plausible that one cannot plan to ϕ and not ϕ in a single circumstance C . No motivational state can cause one to ϕ and not ϕ in one and the same circumstance. The impossibility of this kind of plan mirrors in the required way the unintelligibility of judging that one ought to ϕ and not ϕ in one

⁵³ Our reconstruction of Gibbard's view suppresses many of its details in the interest of clarity and to emphasize those aspects of his view necessary to explain, specifically, the apparent unintelligibility of normative judgments inconsistent with SUPERVENIENCE. For instance, Gibbard (2003, p89-94) suggests that assenting to SUPERVENIENCE involves having and expressing a kind of meta-plan, one that all planners are committed to: namely the plan *not to* plan to ϕ in C_1 and not ϕ in C_2 while recognizing that C_1 and C_2 are E-identical. Gibbard's discussion of the nature of this commitment is complex and cannot be reproduced here. Nothing in the discussion to follow turns on the issue, however. The critical question for us will be whether Gibbard is entitled to assume, as he does throughout, that a plan to ϕ in C_1 and not ϕ in C_2 while recognizing C_1 as E-identical to C_2 is impossible.

and the same circumstance. Of course, it might be possible, as Sturgeon (2009) and others have pointed out, to independently form distinct plans for a single circumstance that are jointly inconsistent.⁵⁴ After considering one's current situation at different times, one might form the plan to go out for dinner *and* plan not to go out. But it is one thing to have two (cognitively-isolated) plans for a single situation that are incompatible; and quite another to have a *singular* plan that aims at what the agent recognizes as an impossible end: ϕ -ing & not ϕ -ing in C_1 . The expressivist can reasonably maintain that the impossibility of planning to ϕ and not ϕ in C_1 mirrors in the required way the incoherence of a singular judgment that one ought to ϕ and not ϕ in C_1 .⁵⁵

⁵⁴ Sturgeon uses such cases to raise a challenge for Gibbard. Sturgeon suggests that a commitment to SUPERVENIENCE must involve accepting a rational norm of revising pairs of normative judgments jointly incompatible with SUPERVENIENCE. Sturgeon (2009) demands an explanation for why discovering inconsistencies in one's plans should be universally recognized as "a cause for alarm, or reform... rather than just for amusement? Expressivists needs to show (and it is uncertain whether they can) that all planners are committed to a norm of revising pairs of plans that are jointly inconsistent with SUPERVENIENCE. I am skeptical that assent to SUPERVENIENCE necessarily involves accepting the relevant norm governing combinations of plans, but my main purpose in discussing Sturgeon's objection is not to criticize it. It is to note that the cases he describes do not amount to counterexample to the Basic Constraint on Planning.

If the expressivist needs to explain why assent to SUPERVENIENCE operates as a judgment-guiding norm, it suffices to point out that once an agent recognizes that two of her judgments are jointly inconsistent, she will assert, based solely on her competence with normative concepts, that both cannot be true. Gibbard should say, in accordance with his general strategy, that the agent knows—or at least is guided by—an a priori fact about planning that renders unintelligible the judgment that normative judgments that are jointly inconsistent with SUPERVENIENCE are both true. He does not need to show that assent to SUPERVENIENCE involves being motivated to resolve inconsistencies. Admittedly, Gibbard writes as though assent to SUPERVENIENCE involves being committed to a higher-order plan of a certain sort: a plan to avoid having incompatible pairs of plans for E-identical situations.

⁵⁵ That said, one can meaningfully judge that one ought to ϕ in C_1 , and independently judge that one ought *not* to ϕ in C_1 ; just as one can make independently meaningful judgments that Tim is a bachelor and Tim is a married man.

While the Easy Problem is easy for the plan-based expressivist, the Hard Problem is harder. If we assume the supervenience base of empirically recognizable properties is repeatable, then the Argument from Planning becomes far less compelling because its third premise—No Action Guidance—seems false. Whereas a plan with the aim ϕ -ing & not ϕ -ing in C_1 may well be impossible because it cannot guide action, a SUPERVENIENCE-flouting plan with ϕ -ing in C_1 and not ϕ -ing in C_2 as its aim where C_1 and C_2 are E-identical but *distinct* situations seems possible because it *can* guide action.

To see this, we need to consider possible circumstances that are empirically exactly alike but genuinely distinct.⁵⁶ The movie “Groundhog Dog” provides a helpful scenario we can draw on. Bill Murray’s character Phil finds himself reliving the same day over and over—February 2nd, or Groundhog Day. He wakes up at the same time, at the same Punxsutawney bed and breakfast, with the same song playing on the radio, encounters the same people and situations, and so on. In the movie, Phil realizes that he is stuck in a time loop fairly quickly. But, given our purposes, we can add the following twist to the story. Suppose Phil is told beforehand that he will be entering this time loop, and, what’s more, his memory of the previous day will be erased each morning. Each morning he will wake up not remembering the day before. Take two iterations of Groundhog day: Day₁ and Day₂. These are different circumstances, the one occurring before the latter. Their temporal separation is not what is ultimately relevant for the question we are about to pose. What is relevant is that they are *distinct* situations. The question is whether it is possible to plan to treat the situations differently *just because* they are distinct, when their distinctness is

⁵⁶ The argument does not depend on assuming that there are such things as haecceitistic properties. If there are no haecceities of situations, then we can appeal to some qualitative profile of a situation that cannot be shared between distinct situations. Since we’re supposing that the class of empirically detectable properties is repeatable we are entitled to suppose that there is some empirically undetectable non-repeatable qualitative profile of situations. The question becomes: can we treat situations differently based on this non-repeatable qualitative profile? The answer, I argue, is: yes.

undetectable.⁵⁷ The day before he is scheduled to enter the time loop, Phil reflects on what he should do. I claim that Phil *can* plan to act differently on Day₁ and Day₂, *just because they are distinct situations*. We might compare a plan to treat Bob differently from Fred, when Bob and Fred are identical twins and there is no way to distinguish Fred from Bob.

The Argument from Planning tries to establish such a plan's impossibility by appeal to the action-guiding nature of plans. Necessarily, if x is a plan, then x is a motivational state with some action-involving outcome as its aim *and* x is a state that can possibly cause or bring about the action-involving outcome. That is why a plan that aims at ϕ -ing & not ϕ -ing in C is impossible. No mental state can possibly cause one to both ϕ and not ϕ in C . But there *is* a motivational state that:

- (1) has ϕ -ing if it is Day₁ and not ϕ -ing if it is Day₂ as its aim; and
- (2) can appropriately cause the realization of its action-involving aim.

After all, Phil can dominantly *desire* or prefer to perform incompatible acts in the two situations *just because they are distinct situations*. There is no bar to dominantly desiring incompatible acts for genuinely distinct situations while recognizing the situations as empirically indiscernible. The desire may be whimsical, to be sure. But whimsical desires are not impossible. The posited desire can combine with one's beliefs in a situation to result in action. Day₁ and Day₂ represent distinct situations one could be in, by hypothesis, and so long as one forms a belief either way concerning whether one is in Day₁ or Day₂, one's dominant desire that one ϕ in Day₁ and not ϕ in Day₂ will result in some action (possibly even the desired one so that one ends up ϕ -ing in Day₁ and not Day₂). Even though Phil's memory will be erased, it remains possible that he

⁵⁷ I think that the empirical undetectability of bare identity is perhaps the most plausible (but ultimately mistaken) basis for thinking that our motivations cannot treat two situations differently *just because they are distinct*, which is why Gibbard's discussion of the issues provides a useful starting point and frame for the overall argument.

spontaneously forms a true belief on Day₁ that he is stuck in a time loop and that he is experiencing the iteration of the eternally recurring Groundhog Day that he dubbed Day₁. So, it *is* possible for a mental state to be both action-guiding and have as its aim: ϕ -ing in Day₁ and not ϕ -ing in Day₂.

If one is tempted by the thought that a plan *just is* a dominant preference/desire for an outcome characterized in terms of actions and circumstances, then one should think that the supervenience-flouting plan is possible. The view is very tempting indeed if all we're told about plans is that they are necessarily action-guiding motivational states individuated using action-circumstance pairs.

But perhaps the expressivist can do better than the Argument from Planning. She might insist that a plan is not just a motivational state that can guide action. It constitutively involves a *reasonable expectation* that one will realize the outcome aimed at. The motivation to treat undetectably distinct situations differently must combine with a suitable expectation if it is to count as a plan. And it is this reasonable expectation that is missing in the case of Phil's whimsical desire to treat two situations differently just because they are distinct situations. The only possible worlds where he realizes the desired outcome involve luck—worlds where Phil has a randomly acquired true belief about his situation. Surely, or so the response goes, it is impossible to form a plan that, one recognizes, depends for its success on one's forming unjustified true beliefs. Plans plausibly involve evidentially supported or rational expectations for plan-realization, along with the relevant desires.⁵⁸

⁵⁸ Other ways that theorists have suggested plans might be different from mere dominant desires are not helpful. Bratman (1987), for example, thinks that intentions are distinct from mere desires in essentially involving a kind of resolution not to reconsider one's intention. But, of course, pairing resolve with a dominant desire to realize $\{(\phi, \text{Day}_1) \text{ (not } \phi, \text{Day}_2)\}$ is possible. Even if plans in Bratman's sense must be supervenience-respecting, an argument I

This might seem like a compelling response, but the rational expectations constraint should be rejected. It is vulnerable to two basic objections. First, we must be careful not to confuse the potential oddness or imprudence of the candidate plan for the strong sort of impossibility at issue.⁵⁹ Generally, when we plan to do something we also have some sense of how we will reliably do it, and this seems like good practice. It would be setting oneself up for disappointment to regularly leave the successful implementation of one's plans to chance. One way of guarding against frustration is making sure one has a reliable method for determining whether the situations relevant to one's plans have transpired. A plan to treat two situations differently just because they are distinct and not on the basis of any detectable difference between the two goes against common sense because it leads easily to the frustration of one's ends. The crucial question, however, is not whether Phil's plan is a strange or bad plan, but whether it is impossible or inconceivable. It is neither. There are plenty of impudent planners who are not *failing to plan* on account of their imprudence.⁶⁰

The second and more decisive reason why the 'reasonable expectations' constraint should be rejected is that, setting aside our pre-theoretic intuitions about plans, plans in the sense relevant to normative judging *must* be the kind of motivational states that can treat undetectable

give later concludes that plans in the sense relevant to normative judging must be a kind of supervenience-flouting motivational state.

⁵⁹ Gibbard suggests only that plans necessarily aim to guide the agent's actions (2003: 56). Nowhere is the narrower conception of action-guidance on the basis of evidentially supported beliefs defended or even made explicit.

⁶⁰ There appear to be plenty of ordinary cases of agents purporting to have plans of the sort described – that distinguish situations based on undetectable features. Gnostics believe in a God who exists beyond space and time and is not just unknown but *impossible* to know, at least by ordinary mortals. A world where the Gnostic God exists and a world where he doesn't are plausibly E-identical, but unobservably distinct. A person who follows the faith's practices will treat these worlds very differently with respect to her plans. She might plan to be kind in a world where a Gnostic God exists and not be kind in the world where he doesn't. I think we have good reason to take the agent's reports about her plans at face value.

differences between situations as relevant from the planning perspective. It is true that Phil's purported plan to treat situations differently *just because* they are distinct situations involves distinguishing situations on the basis of undetectable features, and so cannot involve a reasonable expectation of plan-realization. But that *cannot* be the reason for the plan being impossible, given the role that plans play in the expressivist's semantic theory. We must keep in mind that plans, for the expressivist, do significant theoretical work: they undergird meaningful normative thought and talk. Moreover, expressivists are (and indeed should be) committed to vindicating as meaningful those of our ordinary normative judgments that involve so-called *robustly objective* uses of ought. Such usage involves judging that a person ought to do something even when she does not know and could know that it is the thing she ought to do. The person might be constitutionally disposed to doing the wrong thing.⁶¹ Consider:

<Paranoid schizophrenics ought to get themselves some therapy>

According to the expressivist, in asserting this fact, I am, expressing my plan to *get therapy in the circumstance of being a paranoid schizophrenic*. But if I were a paranoid schizophrenic, I might be insensitive to facts about my own paranoia and have no inclination to get therapy. If my extreme psychological propensity to not- ϕ in a situation because of my inability to reliably identify the situation based on my evidence were enough to make a plan to ϕ in that situation impossible, the expressivist would have to give an error-theoretic diagnosis of what seems to be a clearly intelligible and quite likely *true* normative claim: that paranoid schizophrenics *should* get therapy. So, Phil's plan to treat Day₁ and Day₂ differently just because they are distinct situations cannot be impossible on the grounds that it depends, for its success, on the eventuality of his forming true non-evidentially supported beliefs about unrecognizable features of his situation.

⁶¹ See Daskal's (2009) discussion of Gibbard's view in the context of "innocent mistakes of mental constitution," and, in particular, what Gibbard can say about 'Mortis,' the ideally coherent psychopath.

Similarly, it is sometimes argued that there are uses of *ought* where blameless ignorance does not undermine the truth of the ought judgment. Consider:

<I morally ought to turn the light switch on and off repeatedly if doing so will cure cancer, *even if* it is entirely unknowable by me that it would>

While it is controversial as a matter of first order moral theory whether such objective ought claims are true, they are at least meaningful. An expressivist committed to vindicating them as meaningful *has* to think that one can plan to turn the light switch on and off in a circumstance where doing so will cure cancer, even if there is no way for one to detect this feature of the circumstance.

So, we can summarize the critical point (which would apply even if we started our discussion with a different motivational state favored by expressivists, like a stable, higher-order desire): whatever desire-like attitudes plans are supposed to be, they must be attitudes that can underwrite as meaningful robustly objective normative judgments. But this means a plan—where ‘plan’ is a place-holder notion for the *relevant* motivation state—to treat two situations differently *just because* they are distinct situations cannot be impossible for treating as relevant from the planning point of view undetectable features of two situations. Plans cannot constitutively involve rational or reasonable expectations of plan realization.⁶² It must be possible to plan to

⁶² Note that, by contrast, a *non-rational* expectation may be constitutively involved in planning, but that does not help the expressivist. Phil’s desire to ϕ if it is Day₁ and not ϕ if it is Day₂ *can* combine with an expectation that the desired outcome will be realized. For suppose Phil forms the evidentially unsupported belief that:

<On each day, I shall form a correct belief as to whether I am in Day₁ or Day₂.>

The critical question is whether this belief is consistent with Phil’s taking the two situations to be E-identical, which we must hold fixed. The answer is that it is indeed consistent. We can stipulate that Phil doesn’t expect his beliefs about which situation he is in to be correlated in anyway with the truth. The belief he ends up forming might be entirely random. Moreover, Phil’s strong desire to ϕ in Day₁ and not ϕ in Day₂ makes no mention of his beliefs in Day₁ and Day₂. The case as imagined simply involves a combination of a dominant desire with ϕ -ing in Day₁ and

treat E-identical but genuinely distinct situations differently. So, supervenience-flouting plans seem possible, at least if the supervenience base is defined as the category of *repeatable* non-normative properties—a category which does not include bare identities of situations.⁶³

At this point, the plan-based expressivist might be tempted to avoid the challenge by means of stipulation. She might grant that one can indeed plan to treat E-identical but genuinely distinct situations differently while insisting that only supervenience-respecting plans are ones the expression of which results in meaningful normative concept use. We can, after all, define a set of plans* which includes only those plans that do not distinguish between distinct circumstances recognized as E-identical. Normative concept use necessarily involves the expression of plans*, so repeatable supervenience-flouting judgments are unintelligible. Analogously, I could define up the concept of a desire*, which is a lot like the concept of a desire, except that one desires* to ϕ only if ϕ -ing involves maximizing happiness. It is impossible to desire* to not maximize happiness because of *what it is* to desire*.

There are several reasons why brute stipulation cannot save the expressivist. Firstly, the move would render normative concepts and whether one succeeds at deploying them arbitrary. After all, while I can't plan* to act differently in situations that are E-identical, I can do something that looks a lot like planning*—namely, I can plan to act differently. It seems hardly

not ϕ -ing in Day₂ as its aim and Phil's belief, entirely unsupported by the evidence, that he will believe correctly which situation he is in. This seems a coherent combination of attitudes that is consistent with thinking the situations to be E-identical.

⁶³ As it turns out, it wasn't even necessary for us to establish decisively that supervenience-flouting plans are possible. We have successfully raised a serious problem for the expressivist by marshaling enough reason to *doubt* and *suspend judgment* on the impossibility of supervenience-flouting plans. But suspending judgment on whether planning is constitutively constrained to bar supervenience-flouting normative judgments entails suspending judgment on whether supervenience holds as a conceptual matter if expressivism is true.

plausible that whether I succeed at meaningfully deploying normative concepts can hang on so tenuous a thread.

More importantly, the expressivist cannot simply stipulate that normative concepts involve plans* and not plans, no more than one can stipulate a meaning for the ordinary concept of redness. The expressivist must *motivate* the claim that normative concepts involve planning* and not planning. Crucially, she cannot appeal to our commitment to SUPERVENIENCE as proof that we are planners* because an explanation of our commitment to SUPERVENIENCE was supposed to emerge from reflecting on the activity we are engaged in when we deploy normative concepts. That is, we were supposed to have a more explanatorily basic grasp on the nature of the non-cognitive attitude expressed when we use normative terms like 'good' and 'ought' than on why it is that we are committed to SUPERVENIENCE. And if there is no other justification she can give for restricting meaningful normative judgments to those that involve the expression of a plan*, the restriction will seem *ad hoc* and designed to save the view from objection.

It is worth emphasizing this final point as a kind of summary of the dialectic so far. We must keep in mind the order of explanation: our verdict on supervenience is supposed to be *explained* by our understanding of normative concepts and not the other way around. Since, on the expressivist picture, an understanding of normative concepts turns out to be an understanding, perhaps implicit, of the non-cognitive mental state we express when we deploy normative concepts, it had better be the case that we know we are planners* and not planners (when we deploy normative concepts) independently of our knowledge of SUPERVENIENCE. I have argued that expressivists have offered us no reason for thinking that we are planners* rather than planners when we deploy normative concepts.

4. A General Challenge

There are two other ways expressivists might try to deflect our challenge. One would be to concede that the motivational state involved in normative thought and talk *can* be supervenience-flouting, but then provide some reason for further complicating the semantics. The expressivist might try to explain *why* we only recognize normative judgments as meaningful when the motivational state expressed is supervenience-respecting (instead of simply stipulating that we do).

Compare, for example, how a non-naturalist robust realist about normativity might go about explaining features of our normative concept. The robust realist might argue that our concept of a normative property is that of a property that covaries with a repeatable non-normative property precisely because it lies in the nature of the objective property in normative reality that our concept tracks that it covaries with a repeatable non-normative property, and normative reality has in some way shaped the concepts we use. This at the very least has the form of a genuine explanation, whatever one thinks of its assumptions. The expressivist obviously cannot appeal to objective normative reality to explain why meaningful normative concept deployment involves the expression of supervenience-respecting plans.⁶⁴ Instead, she might appeal to facts about our usage and the purposes that normative language is designed to serve. As argued earlier, what she cannot do is appeal to the truth of SUPERVENIENCE.

The problem is that no expressivist has offered a remotely plausible explanation of why, independently of our commitment to SUPERVENIENCE, only supervenience-respecting non-cognitive attitudes can underwrite normative thought and talk. Blackburn (1986), for whom the relevant non-cognitive attitude is a kind of stable, higher-order desire, suggests that we use normative vocabulary to coordinate with one another. We engage in normative discussion, and the associated expression of our motivational states, to invite others to share our higher-order desires and plans so we can harmonize our actions. If normative talk aims at coordination, this

⁶⁴ On this point, see Dreier (2015).

might explain our disposition to negatively evaluate those who express supervenience-flouting motivational states. That is, the coordination account would predict that we would be inclined to use normative language to express our disagreement with those who express supervenience-flouting motivations insofar as our own motivations are supervenience-respecting. But it is hard to see why the coordination aim of normative language should entail that normative claims underwritten by supervenience-flouting non-cognitive attitudes would be rejected as *meaningless*. To put it differently, if normative language was introduced to facilitate coordination amongst agents, the expression of motivational attitudes that *can* possibly violate supervenience serves just as well as the expression of those that constitutively cannot.⁶⁵

An alternative means of deflecting the challenge would be to suggest that plans cannot treat haecceitistic features of situations—bare identity—as relevant from the planning point of view for some reason other than the empirical undetectability of bare identity. While our previous arguments show that the fact (if it is one) that bare identity is empirically undetectable cannot be the reason for treating supervenience-flouting plans as impossible, we have not shown that there might not be some other reason.

⁶⁵ So, when Blackburn (1993) writes:

it seems to be a conceptual matter that moral claims supervene on natural ones. Anyone failing to realize this or to obey the constraint would indeed lack something constitutive of competence in the moral practice. And there is good reason for this: it would betray the whole purpose for which we moralize, which is to choose, rank, approve, for bid, things on the basis of their natural properties. (p66)

he presupposes the fact that needs to be explained—namely, that the purpose behind using normative terms (or moralizing) is to choose, rank, approve and so on *specifically on the basis of the natural properties of things*. As we have seen, the coordination account cannot explain why a supervenience-flouting normative judgment would so betray the aim of coordinating to render the judgment unintelligible. Without a plausible explanation of why it must be supervenience-respecting motivational attitudes that underwrite normative claims, the expressivist should suspend judgment on whether supervenience is a conceptual truth, at least if the supervenience base is defined in terms of ‘natural’ or empirically recognizable features.

There are two things to say about this attempted deflection. First, the burden is on the party claiming the impossibility of a kind of mental state to provide reasons for thinking it impossible. In fact, our reason for using Gibbard's discussion as a starting point was precisely that his account seemed most likely of any to produce a reason for thinking it is impossible to plan based on haecceitistic differences. But we've seen that the empirical undetectability rationale does not work. Secondly, we have strong reasons, based on the previous discussion, for supposing that the motivational states relevant for normative judging *must* be able to treat individuals— persons, acts, situations, or even possible worlds (construed as maximally described situations)—differently based purely on their identity. It is certainly true of a dominant preference or desire: I can dominantly prefer Fred over Bob even if Fred and Bob are indistinguishable twins just because Fred is Fred and Bob isn't! The only constraint on such mental states derivable from their nature that *is* plausible is a rather trivial constraint: to treat *numerically identical* individuals alike. So, perhaps a singular preference that has its object ϕ -ing and not ϕ -ing in a single circumstance C is impossible. This trivial constraint on our motivational states ensures that the expressivist can straightforwardly address the *Easy* Problem of Supervenience. But she has no answer for the *Hard* Problem of explaining why it is a conceptual truth that the normative supervenes on the repeatably non-normative.

5. Ways of Biting the Bullet

The expressivist has two options in light of the Hard Problem, neither one of which leaves the conventional wisdom about supervenience and expressivism intact. One is to concede that the conceptual fact of repeatable supervenience has no explanation. This would amount to embracing the view I described earlier as stipulative, but without pretending that it solves the Hard Problem. The expressivist can complicate her semantic story and accept as a *brute fact* about our interpretive practices *vis à vis* normative discourse that we only recognize normative

judgments as meaningful if they involve the expression of motivations consistent with supervenience. There is no deeper explanation for why normative talk works this way. As far as brute facts go, this one needn't be all that mysterious. In general, there are presumably many arbitrary features of our linguistic practices that lack deep explanations. Notably, expressivism, unlike other meta-normative theories, at least renders it *plausible* to take the relevant conceptual fact to be brute. Contrast the unattractiveness of an analogous move in the case of the non-naturalist. If the non-naturalist takes it to be a brute fact about our normative practice that we do not recognize normative judgments inconsistent with supervenience as meaningful, she confronts the question: how is it that an arbitrarily determined feature of our practice mirrors a robustly metaphysical fact in normative reality—namely, the supervenience of robustly real normative properties on the repeatably non-normative?⁶⁶ So, perhaps the right thing for expressivists to say is not that expressivism *explains why* it is a conceptual truth that the normative supervenes on the non-normative. She can say that expressivism explains why it is unproblematic to take as *unexplained* the conceptual truth that the normative supervenes on the non-normative.

But, of course, expressivists have long claimed the ability to *explain* the conceptual truth of SUPERVENIENCE.⁶⁷ I doubt that expressivists have merely been speaking loosely in putting forward their explanatory claims. Expressivist explanatory ambitions have presumably been guided by the attractive assumption that our concept of the normative is not arbitrary and that a feature like the supervenience constraint calls out for explanation. While it may be true that the conceptual fact is more plausibly taken as brute given expressivism rather than, say, non-naturalism about the normative, the move remains less than satisfying for anyone who took it to

⁶⁶ Incidentally, the non-naturalist does not have to take the relevant conceptual fact as brute, and in this non-naturalists may find basis for claiming a theoretical advantage over expressivists (and reductive naturalists), but that is a discussion for a different paper.

⁶⁷ Blackburn (1970: 120; 1985: 64): “[Supervenience] is *explained* by the anti-realist nature of moralizing.” See Gibbard (2003: 96) and the discussion of Gibbard's view in Section II.

be an important pre-theoretical datum that the nature of the concept demands a deeper explanation. Another problem with taking the conceptual truth as brute is that it is in tension with the fact that, by the expressivist's own lights, we find no further support for SUPERVENIENCE through greater insight into the nature of normative concepts. What we learn about the nature of normative thought and talk—namely, its motivation-expressive function—is consistent with SUPERVENIENCE being false.

There is an alternative reaction to the Hard Problem that seems more attractive than treating SUPERVENIENCE as a brute conceptual truth. The expressivist might reject the claim that it is a *conceptual* truth that the normative supervenes on the repeatably non-normative. She might argue, instead, that supervenience is a *normative* truth rather than a conceptual one.⁶⁸ As discussed, the expressivist needn't face any significant challenge explaining why we might shun plans to treat empirically indiscernible situations differently. Such plans certainly seem like bad ones insofar as they reliably lead to the frustration of one's ends and are possibly grounded in unreasonable expectations. Embracing supervenience as a kind of normative truth would allow the expressivist to explain why it might have *seemed* to be a conceptual truth. For one, the epistemology of normative and conceptual beliefs is similar. Both types of belief appear to be based on *a priori* reasoning.⁶⁹ Moreover, this approach allows the expressivist to give a deeper explanation for the truth of supervenience, albeit a normative one, in terms of why one ought not to have supervenience-flouting motivations.

While some of the costs of adopting either one of the two non-ideal (or “unhappy-face”⁷⁰) solutions to the Hard Problem may be perfectly general and concern the enduring

⁶⁸ There is a minority view that takes precisely this view of supervenience. See Kramer (2009); Rosen (MS).

⁶⁹ For a priorism in ethics, see Audi (2015: 61).

⁷⁰ Schiffer (2003: 5-6) calls a solution to a philosophical puzzle a “happy-face” solution if it can identify which amongst a set of incompatible intuitions is false and explain why it is false. Unhappy-face solutions simply point out that there isn't a happy face solution to the puzzle. There is instead an irreconcilable tension resulting from our pre-

plausibility of supervenience conventionally understood as a conceptual truth that demands explanation, there is a theoretical price that expressivists necessarily incur. Expressivists have long regarded it a central virtue of their account of the nature of the normative that it is both consistent with and explains the conceptual truth of normative supervenience.⁷¹ For example, Simon Blackburn (1993: 137) writes that it is a principle virtue of expressivism that it fully explains why “it seems to be a conceptual matter that moral claims supervene. . . . Anyone failing to realize this or to obey the constraint would indeed lack something constitutive of competence in the moral practice.” Once we render the relevant supervenience phenomenon precise—in particular, when we see that that it involves a *repeatable* supervenience base—we find that expressivist explanations of the conceptual truth falter. Having to take supervenience to be a normative truth or an unexplained conceptual truth undercuts the expressivist’s right to claim a dialectical advantage that has long been taken for granted on all sides: the advantage of having no revisionary or counter-intuitive implications with respect to supervenience. The problem of explaining the conceptual truth of normative supervenience, precisely described, turns out to be a more general problem in metaethics than conventional wisdom would have us believe.

References

Audi, R. (2015) “Intuition and Its Place in Ethics”, *Journal of the American Philosophical Association*, 1: 57–77

theoretical commitments. The problem of squaring expressivism with our intuitions about supervenience admits only of an unhappy-face solution.

⁷¹ Blackburn (1984); Gibbard (2003).

Bader, R. (2016) "The Grounding Argument Against Non-reductive Moral Realism." Oxford Studies in Metaethics.

Blackburn, S. (1971) "Moral Realism," in *Essays on Quasi-realism*, Oxford: Oxford University Press.

_____ (1984) *Spreading the Word*. Oxford: Oxford University Press.

_____ (1993) "Supervenience Revisited," *Essays on Quasi-realism*, Oxford: Oxford University Press.

Bratman, M. (1987) *Intentions Plans, & Reasons*. Center for the Study of Language and Information.

Daskal, S. (2009) "Plan-Based Expressivism and Innocent Mistakes," *Ethics* 119: 310–35.

Dancy, J. (1981) "On Moral Properties," *Mind*, 1981, 90: 367—385.

_____ (2004) *Ethics Without Principles*, Oxford: Oxford University Press.

Dreier, J. (2006) "Negation for Expressivists: A Collection of Problems with a Suggestion for their Solution," in (ed. Shafer-Landau, R.) *Oxford Studies in Metaethics*, Vol. 1

_____ (2015) "Explaining the Quasi-Real," in (ed. Shafer-Landau, R.) *Oxford Studies in Metaethics*, Vol. 10. Oxford University Press. Oxford.

Enoch, D. (2011) *Taking Morality Seriously*. Oxford UK. Oxford University Press.

Fine, K. (2002) "Varieties of Necessity," in (Tamar Szabo Gendler & John Hawthorne, eds.) *Conceivability and Possibility*. Oxford University Press. pp. 253-281.

Gibbard, A. (1990) *Wise Choices, Apt Feelings*, Cambridge, Mass: Harvard University Press.

_____ (2003) *Thinking How to Live*. Cambridge, MA. Harvard University Press.

- Griffin, J. (1996) *Value Judgment*, Oxford: Clarendon Press.
- Hawthorne, J. (2002) "Practical Realism?," *Philosophy and Phenomenological Research*, 64: 169-178.
- Heller, M. (1998). "Property Counterparts in Ersatz Worlds," *Journal of Philosophy*, 95: 293-316.
- Jackson, F. (1998) *From Metaphysics to Ethics*, Oxford: Clarendon Press.
- , Pettit, P, & Smith, M (2000). "Ethical Particularism and Patterns", in Hooker and Little (eds.) *Moral Particularism*, Oxford: Oxford University Press.
- Kim, J. (1984) "Concepts of Supervenience," *Philosophy and Phenomenological Research* 45: 153–76.
- Klagge, J.C. (1988) "Supervenience: Ontological and Ascriptive," *Australasian Journal of Philosophy* 66: 461-70
- Kramer, M. H. (2009) "Supervenience as an Ethical Phenomenon," in Kramer, *Moral Realism as a Moral Doctrine*. Oxford: Wiley-Blackwell, 304–64.
- Lewis, D. (1986) *On the Plurality of Worlds*, Blackwell: Oxford.
- McPherson, T. (2012) "Ethical Non-Naturalism and the Metaphysics of Supervenience." In *Oxford Studies in Metaethics* Vol 7.
- Parfit, D. (2011) *On What Matters*, Oxford, UK: Oxford University Press
- Rosen, G. (MS). "What is Normative Necessity?," available at:
https://www.academia.edu/9159728/Normative_Necessity.
- Schiffer, S. (2003) *The Things We Mean*, Oxford University Press. Oxford UK.
- Smith, M. (1994). *The Moral Problem*. Oxford: Blackwell.
- (2004) "Does the Evaluative Supervene on the Natural?" reprinted in *Ethics and the A Priori*, Cambridge: Cambridge University Press.

Sturgeon, N.L. (2009) "Doubts about the Supervenience of the Evaluative," In *Oxford Studies in Metaethics*, Vol. 4

Zangwill, N. (1995) "Moral Supervenience". *Midwest Studies in Philosophy*, 20: 240—262.