

How are ethical theories explanatory?

Farbod Akhlaghi
Trinity College, Dublin

Forthcoming in *Synthese*

Abstract

Ethical theories are explanatory. But do ethical theories themselves include explanatory content? The direct model holds that they do. The indirect model denies this, maintaining instead that, if true, ethical theories can be employed to provide explanations of the phenomena they concern. The distinction between these models is left implicit in much of ethics. The choice between them, however, has significant methodological and other consequences. I provide two arguments for the direct model and suggest that ethical theories do contain explanatory content. I then respond to three objections, connecting this neglected issue to others concerning property-identity and the nature of explanation and theory confirmation in ethics.

1. Introduction

Ethical theories aim to be explanatory. A simple utilitarianism, for example, aims not only to show that some action is right if and only if it maximises happiness, but also that x is right *in virtue of* maximising happiness, maximising happiness *makes* an action right, and x is right *because* it maximises happiness.

But in what way are ethical theories explanatory? That is, how are ethical theories supposed to establish explanatory claims about, for example, what makes actions right?

There are two options. To see the first, take our simple utilitarianism:

(A) Necessarily, for any action x , x is right if and only if x maximises happiness.

This theory offers a necessitated universal generalisation about the property of moral rightness and the property of happiness maximisation.¹ On this view, moral rightness is not just co-extensive with happiness maximisation in the actual world, but co-extensive across modal space (that is, rightness and happiness maximisation are claimed to be intensionally equivalent).

¹ There has been substantial recent debate over the nature of the necessity in moral principles (see Fine 2002, Rosen 2020, Hattiangadi 2018). I shall assume for simplicity the orthodox view that they concern metaphysical necessity.

One way in which utilitarianism may be explanatory is to suppose that (A) involves an implicit explanatory clause, turning (A) *itself* into an explanatory claim,

(B) Necessarily, for any action x , x is right if and only if, *and because*, x maximises happiness.

This view suggests that the explanatory nature of utilitarianism is generated by the theory *itself* including an explanatory claim, indicated by the ‘because’ in (B).² On this view, the same is true of other ethical theories, including those that involve no necessitated universal generalisations or moral principles such as particularism, which would only make claims about what is, say, right *and why* in a given situation.

The view, then, is that all ethical theories include as part of their content claims about the explanatory relations between some moral properties and some other features they concern (even if these are not supposed to hold necessarily or universally). Call this model of explanation the *direct* model.³

Another way in which, for example, utilitarianism may be explanatory is that whilst utilitarianism *itself* does not make any explanatory claims, it can still be appealed to in generating explanations of in virtue of what right acts are right.

Suppose (A) is true and we ask *why* right actions are right. Since (A) is a necessitated universal generalisation, it can allow us to provide an explanation of why right acts are right by appeal to the generality and law-like nature of (A): if true, (A) can explain its instances by suggesting that right acts are right because, necessarily, actions are right just when they maximise happiness. By bringing instances of rightness under a general, law-like pattern, (A) can thus explain *why* acts are right, revealing how a theory can be explanatory without itself containing any explanatory claims like ‘ x explains y ’ or ‘ y holds because of x ’.

² The nature of the explanatory relation indicated by this ‘because’ is subject to heated debate, see Berker 2018 and Enoch 2019’s response; also see Leary 2020 and Akhlaghi 2022 for discussion about the significance of this debate. I will remain neutral over this since I am only concerned with whether ethical theories themselves involve explanatory claims or not, as opposed to the precise nature of the explanatory notion they may invoke.

³ Berker 2018: 741 and Kagan 1992: 226 suggest that this is the default view.

One way to see this view is to suppose that we can confirm or falsify ethical theories simply by appeal to their *predictions* about the moral status of actions under certain conditions.⁴ For example, (A) predicts that, in any case, an action is right just when that act maximises overall happiness. And, crucially, a theory's predictions are verifiable *without* explanatory claims being *part of the content* of that theory. Physical theories exemplify this: they provide predictions, often on the basis of mathematics, which are verifiable. But that they are verifiable does not indicate that any part of the theory *itself* involves making an explanatory claim.

The same may be so of ethical theories. Perhaps we can confirm or falsify them through considered verdicts on cases they make predictions about *without* those theories containing explanatory content. To think so has a strong motivation: *anti-exceptionalism* about ethical theories. This anti-exceptionalism suggests that, just as theories that lack explanatory content in other domains can be confirmed or falsified through careful consideration of their predictions, so too can ethical theories. And the lack of such internal explanatory content, as in physics, does not suggest that such theories cannot provide explanations. Rather, the thought goes, they provide them in just the same way our best theories elsewhere do so.

Importantly, this could be so even of theories that eschew universal generalisations and only make predictions about particular cases. Particularism, say, whilst surrendering some generality (which particularists deny we can have anyway), can still offer verifiable, law-like predictions about the moral status of actions in actual and hypothetical scenarios. For particularists can allow that there are verifiable, law-like predictions about, for example, 'the sort of importance that a property *can* have in suitable circumstances' (Dancy 1993: 67) and then they apply such considerations on a case-by-case basis.⁵

⁴ By 'confirm', I mean that some ethical theory T making the correct prediction about, say, the moral status of some action in a case C provides some degree of support for T. By 'falsify' I mean that an ethical theory T making some incorrect prediction about the moral status of some action in a case C provides at least some degree of support for $\neg T$ (where this incorrect prediction might ultimately constitute a counterexample to T).

⁵ Whilst they typically eschew universal generalisations, it is consistent with particularism to be committed to a necessitated universal generalisation that is *highly* disjunctive, capturing all the various contextual factors that can, in principle, attenuate the normative significance a property can have under certain circumstances. They then predict

Finally, if we think that a theory consistently gives us the right predictions, we might then employ an *abductive* argument from this to the theories themselves and take the theories to help explain, say, why an action is right. Call this the *indirect* model.⁶

Which model is presupposed in normative ethics? This is hard to ascertain as the distinction between them has remained largely implicit to date. Many classic thought experiments are at least couched in terms that suggest acceptance of the direct model. Consider Judith Jarvis Thomson's (1975; 1985) canonical formulation of the trolley problem as one over *what explains* why it seems permissible to kill one to save five in the Trolley case but not in the Transplant case. We often adjudicate between competing ethical theories, however, on abductive grounds that suggests implicit acceptance of the indirect model. For example, when we accept theory T1 over theory T2 because we think T1 gives us the right predictions in more cases than T2 about what acts are right, this suggests acceptance of the indirect model and adjudication between theories *without* consideration of explanatory relations in particular cases.

Despite this unclarity, we might think that the choice between the direct and indirect models is unimportant. This is a significant mistake.

First, the choice between them is crucial in recent debates over the *nature* of ethical theories. For example, Selim Berker (2018) argues that ethical theories are themselves at least partly meta-ethical because the direct model is true and the 'because' ethical theories concern is the notion of metaphysical grounding. Equivalently, Berker maintains that there are no wholly meta-ethically neutral ethical theories. This potentially surprising consequence depends upon the direct model and ethical theories involving more than merely extensional predictions.

that no factor has universal valence and try to show this on a case-by-case basis. I am grateful to an anonymous reviewer for pushing me to clarify this.

⁶ Berker (2019) makes an analogous distinction concerning the explanatory character of moral principles as being either *explanation-involving* or *explanation-serving* respectively, which Fogal & Risberg 2019 call *explanatory in role* or *explanatory in content*. My concerns here, however, go beyond the explanatory character of moral principles, as I am asking how *any* given *ethical theory* could be explanatory, including those that eschew moral principles altogether.

Second, it has important consequences for ethical methodology. The direct model takes the content of ethical theories to include positing explanatory relations whilst the indirect model does not. If the former is correct, then ethicists must establish these explanatory relations directly through whatever is supposed to justify their theories and cannot rest at only correctly predicting which actions have what moral status. The latter does not require this, leaving the ethicists job to only defend claims that generate the correct predictions about the extensional and intensional equivalence between moral and other properties (before turning to suggest that *they* help us explain why actions have their moral properties).

Third, many central ethical debates are helpfully clarified by attention to the direct and indirect models. Consider Brad Hooker's (1996) argument that whilst the Rossian *prima facie* duties are plausible, they can all be explained by his rule-consequentialism. How should we assess whether Hooker's rule-consequentialism provides the better explanation? If the indirect model is true, neither view has internal explanatory content. They can thus only be assessed by appeal to their predictions about various cases and asking whether they provide plausible law-like generalisations about the correct verdicts in these cases. If the direct model is true, however, then the question becomes whether Hooker's rule-consequentialist principle itself provides a *better explanation* of the phenomena than the principles comprising the Rossian duties.⁷

I now provide two arguments for the direct model. First, the indirect model entails that ethical theories which are plausibly distinct from one another are in fact identical because their content is exhausted by claims of intensional equivalence. Second, the indirect model faces a problem that motivates a disanalogy between theory confirmation in the natural sciences and in ethics which threatens a primary motivation for the model.

2. *Two Arguments for the Direct Model*

⁷ This difference generates a challenge for the indirect model that I explain shortly.

It is plausible to think, and seems widely assumed, that establishing extensional convergence amongst ethical theories does not settle which ethical theory is correct. That is why, for example, many think that even if Derek Parfit's Triple Theory correctly identifies a convergence on the necessary extension of various moral terms and the properties they concern amongst some ethical theories, it leaves out a crucial feature of ethical debate: disagreement over *why* moral properties obtain when they do so.

This is some further evidence that the direct model, that ethical theories include explanatory claims as part of their content, is widely presupposed in ethics. But it also suggests an *argument* for the direct model.

For if the indirect model is correct and ethical theories contain no explanatory content – their content being exhausted by claims of necessary co-extension – then if some ethical theories converged on the (necessary) extension of some moral properties, this would entail that the debate between these theories was settled and, in fact, that these theories were *equivalent*. They would now make all the same predictions and have no differing content, since those predictions are not *explanatory* claims about what explains why some acts are right but, rather, about the necessary co-extension of some moral properties with some others.

But it seems false that if utilitarianism and some other theory identified all and only the same acts as right then these theories would be equivalent in every sense that concerns ethicists defending these theories. So, the indirect model looks implausible, since it entails the identity of theories that agree on what acts are right where this looks false. Call this *the problem of theory collapse*.⁸ What looks missing? The *explanatory content* of, for example, utilitarianism and its

⁸ Problems in this vicinity have not been wholly unnoticed in the literature: see, for example, Moberger (2019: 228). Moberger highlights a possible consequence of denying the hyperintensionality of the 'because' relation appealed to in normative explanation. Whilst related to the present debate, I am not concerned with the precise nature of that relation here. Instead, I am highlighting the hitherto-implicit direct/indirect model distinction and offering two arguments for the former. Even if the 'because' relation normative explanation invokes is hyperintensional, I argue in §3 that this does not resolve the problem of theory collapse for the indirect model theorist. I thank an anonymous reviewer for pressing me here.

seemingly rival theories. Once inserted, utilitarianism and any other theory that agrees on which acts across modal space are right remain distinct since they include competing *explanatory* claims.⁹

The indirect model theorist might remain unpersuaded. Perhaps they will think that theories which have all the same predictions about, say, what acts are right just are equivalent theories in every sense that matters to ethical debate despite not seeming so. This is not *obviously* false and appearances to the contrary may be a mere artefact of the prevalence of assuming the direct model amongst ethicists.

But what of the indirect model theorist's suggestion that we might employ ethical theories in support of independent claims about the explanatory relations between moral and other properties? It is here that the indirect model faces a problem that also troubles one of the primary motivations for the view.

Recall the plausible motivation for the indirect model: *anti-exceptionalism* about theories in ethics. This anti-exceptionalism suggests that, just as theories that lack explanatory content in other domains can be confirmed or falsified solely through careful consideration of their predictions, so too can theories in ethics. We saw this earlier via analogy with physical theories.

But there seems to be a relevant disanalogy here. Take (A), supposing that an act x is right if and only if x maximises happiness. Let ' p ' be ' x maximises happiness'. Now suppose that we are considering a case C, where x is right and x maximises happiness. Suppose finally that something which is necessarily co-extensive with p is also present in C – call it q .

You might reasonably ask: does C confirm (A)? For why should we think that it is p as opposed to q that is the condition being met for it to be the case that x is right? Crucially, we cannot now try to assess (A) by appeal to *another* case with the hope of teasing p and q apart, since the latter are *necessarily co-extensive* with one another.

⁹ Why does an analogue not arise in other domains for theories that lack internal explanatory content, like physics? That is because those theories contain substantially more internal content than ethical theories do. A physical theory might have the same empirical predictions as another but differ substantially in their internal mathematical content. And in a case where two physical theories make all the same predictions and are also mathematically identical, then it does not seem problematic to suppose that they are identical in every sense relevant to physical enquiry.

So, if we think that we can confirm or falsify normative ethical theories through verdicts concerning (actual or hypothetical) cases, it seems difficult to see how we could do this without also thinking that such theories include some implicit or explicit explanatory content. Without such content, it is unclear that these theories predictions could be verified through the method of cases. Call this the *problem of necessary co-extensional interference*.¹⁰

However, assuming we do not want to jettison the method of cases – which would be an extreme and implausible result here – *with* explanatory content included in ethical theories we *can* see how we could answer whether C confirms (A). For if (A) is really (B), including an 'and because' after the bi-conditional, we can ask whether C is a case which confirms whether x is right *because of* p or q . In this way, the direct model avoids the problem of necessary co-extensional interference.¹¹

But, again, the indirect model theorist may demur. In motivating this second problem, we said that p and q are intensionally equivalent. But if so, then ' x is right if and only if p ' and ' x is right if and only if q ' are *necessarily equivalent* to one another. That means that ' x is right if and only if x maximises happiness' and ' x is right if and only if q ' are thus true or false in the same situations. If so, why should we think that we must *choose* between p or q when trying to assess whether C confirms or disconfirms utilitarianism? For they will be confirmed in *exactly the same* cases.

Here is why: because without doing so we trouble our ability to appeal to ethical theories to *justify* explanatory claims. To explain, if there is a necessarily co-extensive property for any candidate explainer property that one might appeal to, why should we think that, for example, (A) is offering the best explanation of why x is right? Since p and q are distinct conditions

¹⁰ Moreover, note that, in general, those who use the method of cases tend to do so in *explicitly* explanatory terms, as we saw with Thomson's formulation of the trolley problem.

¹¹ Why is there no physical analogue to the problem of necessarily co-extensional interference? Physical theories do not *themselves* include claims about the metaphysically necessary co-variance of any physical phenomena they concern with any others. Of course, some might argue that the claims they concern *are* physically or metaphysically necessary by abduction or inference to the best explanation on the grounds of their extreme predictive success, generality, and law-likeness.

concerning distinct properties, there is always going to be interference when attempting to provide an *explanation* of x 's being right by appeal to p . For q is also present in every case we might appeal to in trying to use (A) to argue for the claim that an action's maximising happiness explains why it is right.¹²

The direct model, however, avoids this problem altogether. For if (A) is really (B) – ' x is right if and only if, *and because*, x maximises happiness' – then it is built into confirming (B) that we would be acquiring justification for the relevant explanatory claim that x is right because it maximises happiness (and not because of some other co-intensional property) when considering various cases. To confirm (B) so understood, we look for more than just whether x is right and p holds. Rather, we also ask whether x is right because of p or q . Assuming we've confirmed (B), we close the *explanatory gap* that the indirect theorist faces in trying to employ the claim of necessary co-extension in (A) to justify thinking that x is right because of maximising happiness (since the indirect theorist leaves ethical theories unable to tease apart p from its co-intensional q for explanatory purposes as we've seen).

3. Three Objections

I now consider three objections. The first two are to the problem of necessary co-extensional interference, the third to theory collapse and the significance of the direct/indirect model distinction.

3.1. Intensionalism, Necessary Co-Extensional Interference, and Explanation

The problem of necessary co-extensional interference has two parts:

¹² Intensionalists about properties believe that necessarily co-extensive (that is, intensionally equivalent) properties are identical. They might respond here that there is *no* difference between p and q , except at the level of which is more verbally perspicuous or informative, and so this problem is illusory (confirming p *just is* to confirm q). I return to this in §3.

- (i) Without including internal explanatory content, the indirect model leaves it unclear how the method of cases allows (dis-)confirmation of ethical theories.
- (ii) In the presence of relevant necessarily co-extensional properties, the indirect model theorist faces difficulty when appealing to their ethical theory to justify some explanatory claim. For the distinct, necessarily co-extensional property will generate ‘interference’ when attempting to demonstrate that it is, say, happiness-maximisation that explains why some action has the normative property it does.

But, first, why think that necessarily co-extensional properties *are* distinct? My presentation of the problem of necessary co-extensional interference assumed the falsity of:

(Intensionalism about properties) Necessarily, for any property P and any property Q, P and Q are identical if and only if, necessarily, for any entity, Px if and only if Qx.

On this view, properties that are necessarily co-extensive (intensionally equivalent) are identical.¹³ But if true, then at least (i) is an illusory problem: confirming *p just is* to confirm *q*. And if so, then there is *no* difference between *p* and *q*, except at the level of which is more verbally perspicuous or informative.¹⁴

Moreover, intensionalism about properties is consistent with thinking that normative explanations are hyperintensional. Suppose happiness-maximisation was identical with some property, X. Even so, explanations of, say, the rightness of an action appealing to happiness-maximisation or to X need not be identical. For example, an explanation could still be more verbally perspicuous, informative, or better meet some practical need(s) of the speakers. Second, then, if intensionalism holds, (ii) also looks, in principle, resolvable for the indirect model theorist. Perhaps some ethical theories, for example, can provide explanations that are more

¹³ This is a coarse-grained criterion of property identity that is sometimes known as ‘Hume’s Dictum’ (not to be confused with another claim that goes by that name, namely that there are no metaphysically necessary connections between (in some sense) distinct entities). For discussion in the meta-ethical context, see Bader 2017; Moberger 2020; Olson 2014: §5.1; Snodgrass 2024; Streumer 2017: 11–19; Williamson 2023.

¹⁴ Strictly speaking, there is no difference between the properties that are being described by the relevant predicates in the propositions *p* and *q* (the intensionalism at issue concerns *properties* not *propositions*). For ease, I gloss over this complication.

perspicuous or informative to hearers, and this is enough to justify some ethical claims made based on such theories. That may be so even if there is some necessarily co-extensional (and so identical) property to that which one's theory appeals to in trying to explain why some act has the normative property that it does.

In response: first, if intensionalism is false, then the first two objections above are unavailable. I cannot demonstrate intensionalism false here. Notice, however, that the view entails that many intuitively distinct properties are identical. A common non-moral example: triangularity and trilaterality are intuitively distinct properties, but since they are necessarily co-extensive, on intensionalism they become identical. A moral example: the necessarily co-extensive properties of wrongly torturing children for fun and the property of torturing children for fun, though intuitively distinct are, on intensionalism, identical. I'm sympathetic to the thought that such consequences ultimately reveal intensionalism implausible, and intensionalism has come under sustained and serious attack in recent years.¹⁵ I concede, however, that (i) and (ii) as presented are conditional upon intensionalism's falsity, and I cannot settle that debate here.

Instead, my response is this: even assuming intensionalism on the indirect model, this will generate a revenge collapse problem. The problem concerns not ethical theories but the explanations they can offer (and thus explanatory roles they can play).

The response to (ii) offered above suggests ethical theories, on the indirect model, can play different explanatory roles by casting the identity conditions of normative explanations they offer as either (a) sensitive to the mode of representation they are offered in or (b) subject to certain linguistic, psychological, or epistemic constraints. It is by appeal to either condition that normative explanations may still be hyperintensional, despite intensionalism entailing identity between the properties of happiness-maximisation and X.

¹⁵ For an overview of this and discussion of alternative hyperintensionalist accounts of properties (and more), see Berto & Nolan 2023.

But both (a) and (b) are implausible. As Victor Moberger (2020: 344) argues, what is normatively relevant to some action being, say, right is ‘what happens out there in the world, independently of our modes of representation’. Normative explanations of rightness, for example, are not sensitive to the conceptual guise that some explaining-property is presented under. As Moberger (2020: 344) puts it, ‘suffering [...] is normatively important (if it is) regardless of how we happen to represent it’.

How a property is represented can, of course, make a difference to the perspicacity and informativeness of an explanation to a hearer. But to insist that this reveals two candidate normative explanations are non-identical, when they appeal to an identical (because necessarily co-extensive) property – under different conceptual guises – is to confuse the *act* of explanation and the *product* of an act of explanation.

Acts of explanation are what occurs when one agent tries to explain to another why *p*. The product of an act of explanation is what one produces during said act: an account of what makes *p* obtain.¹⁶ A successful explanation in the latter sense need not be sensitive to modes of representation or other linguistic, psychological, or epistemic constraints. Compare: why does water boil? A successful explanation in the product-sense just is the correct physical account of why water boils. But a successful act of explanation may require further conditions to be met, such as being understanding-conducive for the hearer, who may not understand a (albeit correct) complex chemical product-explanation.

Explanations in the act-sense are sensitive to modes of representation and subject to various linguistic, epistemic, and other constraints. Not so, however, for explanations in the product sense. And since whether *normative* explanations are correct given ‘what happens out there in the world, independently of our modes of representation’, it is extremely plausible that normative explanations in the product sense are not subject to (a) or (b) above.

¹⁶ See Wilkenfeld (2014).

So, assume, *arguendo*, the indirect model (that ethical theories do not contain explanatory content) and intensionalism (that necessarily co-extensional properties are identical). Assume further that we accept our simple utilitarianism and acknowledge there is some necessarily co-extensional, and therefore identical, property with *maximising happiness*.

Then, we can produce distinct *acts* of explanation when referring to the same property under different conceptual guises. Suppose, for example, that the necessarily co-extensional (and so, identical) property with *maximising happiness* is some physical realiser of happiness-maximisation. Explanations of the rightness of some act that appealed to *maximising happiness*, as opposed to some realiser, may be more perspicuous or informative to some hearers.

But notice that distinct *product* explanations are ruled out by the identity of, in our example, *maximising happiness* and some realiser of it. Compare: if ‘water’ and ‘H₂O’ refer to the same property, because they are necessarily co-extensional, then whilst acts of explanation that refer to said property under different guises can be distinct, any *product* explanation of something by reference to the property ‘water’ and ‘H₂O’ refer to are *identical*.

Ethical theories, then, would be unable to provide distinct *product* explanations just by referring to some property under different conceptual guises. If so, then such theories cannot be distinguished by playing *distinct product* explanation roles. But, crucially, for ethical theories to count as playing distinct explanatory roles, we are looking for them to offer distinct *product* explanations. For the normative explanations ethical theories are trying to provide are product-explanations; explanations of how normative properties behave in the world, analogous in this sense to the scientific product explanation of why water boils.

So, whilst intensionalism allows the indirect model theorist a response to the problem of necessary co-extensional interference, it does so at the cost of generating another collapse problem. The collapse here is not between *theories* but between the (product) *explanations* that such theories can offer. To offer distinct normative explanations, ethical theories devoid of internal explanatory content must be able to play distinct *product* explanatory roles (since that is

the type of explanation that normative explanation concerns). But they cannot. At most, they can offer explanations in different modes of representation or that meet certain practical and related constraints, where such explanations are identical *as product explanations*.

3.2. *The Arbitrariness Objection and the Direct/Indirect Model Distinction*

The above discussion raises a final, important objection. If theories, in general, can be distinguished by appeal to their explanatory roles, even if they appeal to necessarily co-extensional properties, then perhaps little rides on the choice between the direct and indirect model. Either we include explanatory claims in our theories and try to directly confirm or disconfirm explanatory claims, or we do not and instead assess the explanatory power of theories by general criteria for theory choice – such as simplicity, elegance, *inter alia* – applied to explanations the theories can offer. If so, then the problem of theory collapse is confused, and the direct/indirect distinction seems unimportant. Call this the *arbitrariness objection*.

The objection suggests that, *pace* the problem of theory collapse, we can distinguish between two necessarily co-extensional theories by how well they can play an explanatory role *regardless* of whether they contain explanatory content. In §3.1., I argued this is so *only if* they can play distinct explanatory roles *in the product sense* of explanation, where we can then compare distinct product-explanations through various criteria for theory choice. I'll now argue that, regardless of intensionalism, ethical theories cannot play such distinct roles on the indirect model. We cannot, then, distinguish between two necessarily co-extensional theories in the way suggested.

Indirect model theorists have two choices. Either ethical theories can play distinct explanatory roles, in the product sense of explanation, whilst suggesting that the properties they concern are identical and yet can provide distinct product-explanations. Or maintain that the properties in question are distinct and then defend their view's explanation by appeal to the specific property – say, happiness-maximisation – their theory concerns.

The indirect theorist looks unable to take the former. If happiness-maximisation and some property, X, are identical, then ‘x is right because of happiness-maximisation’ and ‘x is right because of X’ *just are*, in the product sense, identical explanations, simply presented under two distinct conceptual guises. They are no more incompatible, as Moberger (2020: 344) puts it, than ‘the claims that glaciers are made of water and the claim that glaciers are made of H₂O’.

Taking the latter, however, whilst maintaining ethical theories lack internal explanatory content, faces (ii) above. That is, the theory faces the challenge of being able to justify why it is the specific normative property they appeal to, and not some *distinct* yet necessarily co-extensional one, that offers the correct product-explanation of the rightness of some action. That is, I’ve argued, the second part of the problem of necessary co-extensional interference, avoided by the direct model.

The dilemma, then, is that either the indirect model theorist endorses intensionalism or not. If they do, then they cannot distinguish between two theories by appeal to distinct product explanation roles, if they appeal to some properties that are necessarily co-extensional (and thus identical). If they do not, then they face the original problem of necessary co-extensional interference as first presented. Either way, this objection does not reveal the choice between the direct/indirect model insignificant, or the problem of theory collapse confused.

I end on a note of clarification. As mentioned earlier, the first and second sections of this paper proceeded on the assumption that necessary co-extension does not entail property identity. We can now more clearly see, then, that there are, in fact, two versions of the direct and indirect models respectively:

(Direct-Hyperintensional Model) Ethical theories include explanatory content and necessarily co-extensional properties need not be identical.¹⁷

¹⁷ ‘Need not be’ is to allow that even if necessary co-extension of properties does not *entail* their identity, some given necessarily co-extensional properties could turn out to be identical, in principle, for other reasons. I thank an anonymous reviewer for this suggestion.

(Direct-Intensionalist Model) Ethical theories include explanatory content and necessarily co-extensional properties are identical.

(Indirect-Hyperintensionalist Model) Ethical theories do not include explanatory content and necessarily co-extensional properties need not be identical.

(Indirect-Intentionalist Model) Ethical theories do not include explanatory content and necessarily co-extensional properties are identical.

Since the first two sections of this paper were presented upon a presumption of hyperintensionalism, the views considered there were (Direct-Hyperintensional Model) and (Indirect-Hyperintensional Model).

I've argued, by appeal to the act/product explanation distinction, that the (Indirect-Intensional Model) cannot provide distinct product-sense normative explanations. So, that view faces a revenge version of the problem of collapse (since necessarily equivalent theories cannot be distinguished by different product explanation roles), revealing that it still faces an explanatory challenge (a version of (ii)). The (Indirect-Hyperintensional Model), I've argued, faces both the problem of theory collapse and necessary co-extensional interference, as originally presented, that the (Direct-Hyperintensional Model) avoids.

This leaves (Direct-Hyperintensional Model) *and* (Direct-Intensional Model). I've noted my sympathy for the former, but nothing argued here settles the choice between them. Nor should it. For the debate over intensionalism is not one that we should expect to be settled by asking in what ways ethical theories can hope to be explanatory.¹⁸ It requires far more than can be done here. Instead, I've focussed upon making this hitherto-unnoticed distinction explicit. My aim has been to clarify that distinction, how to formulate the two models, and defend two arguments against one of those models.¹⁹

¹⁸ I suspect, though, that (Direct-Intensionalist Model) faces the problem of being unable to provide distinct product-explanations. I note this without suggesting it a conclusive reason to reject intensionalism.

¹⁹ I am very grateful to two anonymous reviewers for raising the worries addressed in this section.

4. Conclusion

I have distinguished two models of how ethical theories might be explanatory, the direct and indirect models, and argued that the choice between them is significant. I provided two arguments against the indirect model and for the view that ethical theories are best understood as having internal explanatory content.

These are not the final word. My aim has been to highlight and motivate concern with this important question, distinguish between two views and consider their respective motivations, present and defend some problems for one view, and, in the process, raise more questions than I can answer here.

Settling such questions will require much further work. This will include more reflection on the direct/indirect model debate as it relates to those over property identity, the explanatory relation employed in ethical theories, normative laws, and more. But this paper, finally, has revealed that those latter debates may benefit from greater attention to the *structure* of normative ethical *theories*, and how their structure matters to explanation and theory confirmation in ethics.

Acknowledgements

I am extremely grateful to Roger Crisp, Geoff Keeling, Matt Dougherty, Tim Williamson, and to two anonymous reviewers of this journal for very helpful feedback which much improved the paper. My initial idea for this paper emerged during correspondence with Tim; his entertaining my repeated attempts to convince him of my views on the structure of ethical theories are deeply appreciated.

References

- Akhlaghi, Farbod. (2022). Does the Unity of Grounding Matter? *Mind*, Vol. 131, No. 523, pp. 828–837. <https://doi.org/10.1093/mind/fzab008>
- Bader, Ralf. (2017). The Grounding Argument against Non-reductive Moral Realism. In *Oxford Studies in Meta-ethics*, Vol. 12, edited by Russ Shafer-Landau, Oxford: Oxford University Press, pp. 106–134.
- Berker, Selim. (2018). The Unity of Grounding. *Mind*, Vol. 127, No. 507, pp. 729–777. <https://doi.org/10.1093/mind/fzw069>
- Berker, Selim. (2019). The Explanatory Ambitions of Moral Principles. *Nous*, Vol. 53, No. 4, pp. 904–936. <https://doi.org/10.1111/nous.12246>
- Berto, Francesco & Nolan, Daniel. (2023). Hyperintensionality. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta.
- Enoch, David. (2019). How Principles Ground. In *Oxford Studies in Metaethics: Volume 14*, edited by Russ Shafer-Landau (Oxford, UK: Oxford University Press), pp. 1–18.
- Fine, Kit. (2002). The Varieties of Necessity. In *Conceivability and Possibility*, edited by T. Gendler and J. Hawthorne (Oxford, UK: Oxford University Press), pp. 253–281.
- Fogal, Daniel. & Risberg, Olle. The Metaphysics of Moral Explanations. *Oxford Studies in Metaethics: Volume 15*, edited by Russ Shafer-Landau (Oxford, UK: Oxford University Press), pp. 170–194.
- Hattiangadi, Anandi. (2018). Moral Supervenience. *Canadian Journal of Philosophy*, Vol. 58, No. 3-4, 592–615. <https://doi.org/10.1080/00455091.2018.1436034>
- Hooker, Brad. (1996). Ross-style pluralism versus rule-consequentialism. *Mind*, Vol. 105, No. 420: 531–552. <https://doi.org/10.1093/mind/105.420.531>
- Kagan, Shelly. (1992). The Structure of Normative Ethics. *Philosophical Perspectives* Vol. 6, pp. 223–242.
- Leary, Stephanie. (2020). ‘Normativity’. In Raven, Michael J (ed.) *The Routledge Handbook of Metaphysical Grounding* (London, UK: Routledge Publishing), pp. 472–483.
- Moberger, Victor. (2019). The Mackian Supervenience Challenge. *Ethical Theory and Moral Practice*, Vol. 22, 219–236.
- Moberger, Victor. (2020). Hume’s Dictum and Metaethics. *The Philosophical Quarterly*, Vol. 70, No. 279, 328–349.
- Olson, Jonas. (2014). *Moral Error Theory: History, Critique, Defence*. Oxford: Oxford University Press.

- Rosen, Gideon. (2017a). Metaphysical Relations in Metaethics. In *The Routledge Handbook of Metaethics*, edited by T. McPherson & D. Plunkett (Abingdon, UK: Routledge Publishing), pp. 151–169.
- Rosen, Gideon. (2017b). What is a Moral Law?. In *Oxford Studies in Meta-ethics: Volume 12*, edited by Russ Shafer-Landau (Oxford, UK: Oxford University Press), pp. 135–159.
- Rosen, Gideon. (2020). Normative Necessity. In *Metaphysics, Meaning, and Modality: Themes from Kit Fine*, edited by M. Dumitru (Oxford, UK: Oxford University Press), pp. 205–233.
- Snodgrass, J. J. (2024). Multiplying Co-intensional Properties: A Reply to Streumer. *Inquiry: An Interdisciplinary Journal of Philosophy*. Online First, pp. 1–10.
- Streumer, Bart. (2017). *Unbelievable Errors: An Error Theory about All Normative Judgments*. Oxford: Oxford University Press.
- Thomson, Judith Jarvis. (1976). Killing, Letting Die, and the Trolley Problem. *The Monist*, Vol. 59, No. 2, pp. 204–217.
- Thomson, Judith Jarvis. (1985). The Trolley Problem. *The Yale Law Journal*, Vol. 94, No. 6, pp. 1395–1415.
- Vayrynen, Pekka. (2013). Normative Explanation and Grounding. *Proceedings of the Aristotelian Society Supplementary Volume*, Vol. 87, pp. 155–178.
- Wilkenfeld, Daniel A. (2014). Functional explaining: A new approach to the philosophy of explanation. *Synthese* 191 (14): 3367-3391.
- Williamson, Timothy. (2023). Moral Anti-Exceptionalism. In Bloomfield, P. & Copp, D. (Eds), *The Oxford Handbook of Moral Realism*. (Oxford, UK: Oxford University Press), pp. 554–576.