# Non-factualism and Evaluative Supervenience [Forthcoming in *Inquiry]*

Nils Franzén

nils.franzen@filosofi.uu.se

**Abstract**

Supervenience in metaethics is the notion that there can be no moral difference between two acts, persons or events without some non-moral difference underlying it. If St. Francis is a good man, there could not be a man exactly like St. Francis in non-evaluative respects that is not good. The phenomenon was first systematically discussed by R. M. Hare (1952), who argued that realists about evaluative properties struggle to account for it.

As is well established, Hare, and following him, Simon Blackburn, mistakenly took the relevant phenomenon to be weak rather than strong supervenience, and the explanations they offered for it are accordingly outdated. In this paper, I present a non-factualist account of strong supervenience of the evaluative and argue that it fares better than competing realist views in explaining the conceptual nature of the phenomenon, as well as in offering an account of the supervenience of the evaluative in general, rather than more narrowly the moral. While Hare and Blackburn were wrong about the specifics, they were right in that non-factualists can offer a plausible account of the supervenience of the evaluative, that in certain respects is superior to competing realist explanations.

## Introduction

Supervenience in metaethics is the notion that there can be no moral difference between two acts, persons or events without some non-moral difference underlying it. R. M. Hare (1952) is usually credited with having coined the term and with having been the first to discuss the phenomenon systematically. For Hare, the observation was instrumental in his argument against various forms of metaethical realism, and as support for his own prescriptivist version of non-cognitivism. Simon Blackburn (1971, 1993) similarly considered supervenience to offer support for expressivist views of the moral, and to constitute a major problem for realist views.

The main point of discussion in the present literature, in contrast, is whether the supervenience of the moral constitutes a problem for metaethical non-naturalism. The general view is that while naturalists have no difficulty in accounting for the supervenience of the moral, non-naturalists struggle to some extent (cf. for instance McPherson, 2012).

The main reason for why expressivists and other non-factualists no longer occupy the centre stage in discussions of supervenience is that both Hare and Blackburn took the relevant phenomenon to be weak supervenience, rather than strong (see §3). This is widely considered to be a mistake. Accordingly, their arguments against realist views are generally considered to be ineffective.

The purpose of this paper is to offer a non-factualist account of strong supervenience. This fills a lacuna in the current literature, since the non-factualist explanations previously offered by Hare and Blackburn target weak supervenience in particular. Moreover, it is argued that the explanation offered here has explanatory advantages over competing realist views, in that it both provides an explanation of the conceptual nature of the phenomenon, and accounts for the supervenience of the evaluative in general, rather than more narrowly the moral.

Non-factualism, as here understood, is a view about the semantics of moral, and more generally, evaluative terms, holding roughly that the extension of evaluative predicates is fixed by how we feel about things. As will be discussed below, this rough characterisation is compatible with several distinct metaethical views, perhaps most importantly semantic relativism and expressivism. As discussed in §5 and §6 below, the explanation of supervenience offered here trades on a specific understanding of semantic content that is ‘not-fully-factual’, developed to a large extent in the discussion of predicates of personal taste during the last 15 years. The ambition is to show that this way of developing the intuitive not-fully-factualness of taste predicates, accounts for the strong supervenience of the corresponding properties. This view can be extended to other evaluative predicates, and can be substantiated along different metaethical lines.

## Supervenience

First, let us get a firm grip on the phenomenon we are trying to explain. The way that R. M. Hare originally phrased his observation is instructive in several respects, so I will quote it here in full:

Suppose that a picture is hanging upon the wall and we are discussing whether it is a good picture; that is to say, we are debating whether to assent to, or dissent from, the judgement ‘P’ is a good picture… First let us notice a very important peculiarity of the word ‘good’ as used in this sentence. Suppose that there is another picture next to P in the gallery (I will call it Q). Suppose that either P is a replica of Q or Q of P, and we do not know which, but do know that both were painted by the same artist at about the same time. Now there is one thing that we cannot say; we cannot say ‘P is exactly like Q in all respects save this one, that P is a good picture and Q not’. If we were to say this, we should invite the comment, ‘But how can one be good and the other not, if they are exactly alike? There must be some further difference between them to make one good and the other not.’ Unless we at least admit the relevance of the question ‘What makes one good and the other not?’ we are bound to puzzle our hearers; they will think that something has gone wrong with our use of the word ‘good’. Sometimes we cannot specify just what it is that makes one good and the other not; but there always must be something. Suppose that in the attempt to explain our meaning we said: ‘I didn’t say that there was any other difference between them; there is just this one difference, that one is good and the other not. Surely you would understand me if I said that one was signed and the other not, but that there was otherwise no difference? So why shouldn’t I say that one was good and the other not, but that there was otherwise no difference?’ The answer to this protest is that the word ‘good’ is not like the word ‘signed’; there is a difference in their logic. (Hare, 1963/1952, pp. 79–80)

The passage makes a number of observations. First, Hare regarded himself as having discovered something significant concerning the ‘logic’, that is the meaning, of the word ‘good’ and other evaluative words. We may express Hare’s point by saying that (2) is infelicitous whereas (1) is not:

(1) P is exactly like Q in all respects save one, namely, P is signed and Q is not.

(2) ?? P is exactly like Q in all respects save one, namely, P is a good picture and Q is not.

It is true that (2) could be read as a hyperbole, roughly meaning that the two paintings are surprisingly similar in respect to their non-evaluative properties. But although this reading would be felicitous, it is obviously not the relevant one in the present context.

One may be inclined to object that it is equally obvious that two paintings cannot be qualitatively identical in all respects except for one being signed and the other not. There would have to be some further difference in, for instance, their microphysical properties which underlie this other difference. According to this line of thought, the property of being signed supervenes on other properties in the same way that evaluative facts and properties supervene on the non-evaluative.

This objection brings us to a second aspect of Hare’s observation. Hare, followed by Simon Blackburn (1971, 1993), bases his supervenience argument on a combination of the observation above with the contention that there are no specific conceptual connections between the evaluative and non-evaluative. Directly after the passage quoted above, Hare states that:

A natural response to the discovery that ‘good’ behaves as it does, is to suspect that there is a set of characteristics which together [conceptually] entail a thing being good, and to set out to discover what these characteristics are. (p. 81)

Evaluative supervenience would then supposedly reside in specific conceptual entailments between certain non-evaluative term(s) and ‘good’. Hare relied on Moore’s open question argument to eliminate this option from the discussion. I assume in the discussion below that this argument is successful, also when taken to include evaluative aesthetic terms and predicates of personal taste.[[1]](#footnote-1) The original supervenience observation included that while no particular evaluative doctrine is conceptually necessary, it is nevertheless a conceptual truth that there are some non-evaluative properties that underlie all instantiations of an evaluative property.[[2]](#footnote-2)

Hare thus agrees that it is impossible for a painting to be signed without there being *any* further differences in its descriptive properties. But this is because there are analytic relationships between the property of being signed and other non-evaluative properties. What makes the evaluative case special is that there are no such analytic connections to non-evaluative properties capable of explaining their supervenience.

A third aspect of Hare’s observation concerns the generality of the phenomenon in question. While the vast majority of the current literature focuses on the moral domain, Hare’s original example of the paintings makes it clear that the issue is not limited to morality. Consider:

(3) ?? P is exactly like Q in all respects save one, namely, P is elegant and Q is not.

(4) ?? P is exactly like Q in all respects save one, namely, P is nasty and Q is not.

The phenomenon also extends to what are at times referred to as predicates of personal taste. For example, let P and Q be two cocktails:

(5) ?? P is exactly like Q in all respects save one, namely, P is tasty and Q is not.

Or let P and Q be different performances of a stand-up routine:

(6) ?? P was exactly like Q in all respects save one, namely, P was funny and Q was not.

One could also add Hare’s example from his 1984 paper:

(7) ?? XIII 3 is a nice room, but XIII 2, although similar in all other respects, is not.

The moral case should not be overlooked. For example, let P and Q be actions:

(8) ?? P is exactly like Q in all respects save one, namely, P is virtuous and Q is not.

Supervenience thus unites the domains of morality, aesthetics, and taste.[[3]](#footnote-3) The evaluative in its entirety supervenes on the non-evaluative. The point that I wish to emphasise is that the phenomenon appears to be homogeneous – since all the terms in question are evaluative, and since they all appear to share the feature of not being directly entailed by any non-evaluative description, it seems reasonable to believe that (2) through (8) sound strange for the same reasons.

## Weak and Strong Supervenience

As was stated above, the early discussions of supervenience treated it primarily as a problem for metaethical realism, and as a point in favour of non-cognitivist views in metaethics. To grasp the argument employed by Hare and Blackburn, and to understand its shortcomings, we need to distinguish between weak and strong supervenience:

**Strong Supervenience:** ☐(∀F in α)(∀x)[Fx → (∃G in β)(Gx & ☐(∀y)(Gy → Fy))]

**Weak Supervenience:** ☐(∀F in α)(∀x)[Fx → (∃G in β)(Gx & (∀y)(Gy → Fy)]

In these two formulations, x and y are variables ranging over things and actions; F and G are second-order variables ranging over properties; α is the family of all evaluative properties, and β is the family of all non-evaluative properties. G may be constituted by a conjunction or disjunction of non-evaluative properties. In English these formulations read as follows:

**Strong Supervenience**: Necessarily, if anything x has some evaluative property F, then there is at least one non-evaluative property G such that x has G, and necessarily everything that has G has F.

**Weak Supervenience**: Necessarily, if anything x has some evaluative property F, then there is at least one non-evaluative property G such that x has G, and everything that has G has F (cf. McPherson, 2015).

The two formulations differ in that Strong Supervenience has a second nested necessity operator that is absent in Weak Supervenience. This means that the latter makes it possible for different non-evaluative properties to underlie an evaluative property in different possible worlds. For instance, happiness maximisation could be a non-evaluative property linked to rightness in the actual world, while duty fulfilment could underlie rightness in some other possible world. The nested necessity operator in Strong Supervenience rules this out, which is to say that the non-evaluative property or properties that underlie rightness must be the same in all possible worlds.

A wide-scope necessity operator occurs in both Strong and Weak Supervenience. It corresponds to the observation that, as a matter of conceptual necessity, if anything x instantiates an evaluative property, then there is some non-evaluative property which y also has that is shared by everything that instantiates the evaluative property. This wide-scope operator must be read as conceptual for the rendering to capture the fact that making statements like (2)–(8) reflect a conceptual confusion on the relevant reading.

Both Hare and Blackburn assumed Weak-but-not-Strong Supervenience when they argued that realists struggle to explain the supervenience of the moral. Essentially, the supervenience challenge in Hare’s and Blackburns’ version was to explain how it is possible that, whereas two objects, acts or persons can be non-evaluatively similar while being evaluatively distinct *within the same world*, there can still exist evaluative differences across worlds. Hare and Blackburn thought that realists lacked an explanation of this alleged fact.

A fundamental problem with this line of reasoning is that it gets the phenomenon wrong. Dreier makes the point succinctly:

Hare observes that there is no man who is just like St. Francis in all descriptive respects but different from him in respect of moral goodness; and we seem to know this without checking. But in the same way, we seem to know that there could not have been anyone just like the actual St. Francis in all descriptive respects but different in his moral goodness. And, indeed, we seem to know that St. Francis himself could not have been different in some moral respect without being different in some non-moral respect. And this is Strong Supervenience. (Dreier, 2019, pp. 1394–1395)

The claim that St. Francis *could not* have been descriptively the same while being morally different has the same intuitive support as the claim that there *is* no one descriptively like St. Francis who possesses different moral properties. Weak Supervenience on its own does not support this insofar as it leaves open the possibility that there are possible worlds that are exactly like this one in non-evaluative respects, but in which St. Francis is not a good person.[[4]](#footnote-4)

Hare and Blackburn are thus right that an account of the evaluative that implies Weak-but-not-Strong Supervenience, with the necessity operator read as conceptual, would provide an explanation for why statements such as (2)–(8) sound odd. However, such an account would not explain why we take evaluative properties to be distributed in the same way on non-evaluative properties even at non-actual worlds.

Why then did Hare and Blackburn think that the phenomenon to be explained was Weak-but-not-Strong Supervenience? The main reason seems to have been their adherence to the once prevalent empiricist doctrine that all necessity is of a conceptual nature (cf. Dreier 2019, p. 1395; see next section for some further considerations brought forward by Hare in particular). Given this presupposition that all necessities are conceptual in nature, strong supervenience, would thus have to be read as follows:

It is a conceptual necessity that, if anything x has some evaluative property F, then there is at least one base property G such that x has G, and it is a conceptual necessity that everything that has G has F.

We have noted that Hare and Blackburn took Moore’s open question argument to have established that there are no conceptual entailments such that some non-evaluative description of an object entails that it has some evaluative property. The assumptions that all necessity is conceptual in nature, in combination with Moore’s open question argument, thus rules out that the evaluative supervenes strongly on the non-evaluative.

Scepticism with regard to non-conceptual necessity has drastically waned in recent decades, and with it the relevance of Hare’s and Blackburn’s supervenience. By mistakenly taking the explanandum to be Weak Supervenience rather than Strong, they not only undermine their own argument against metaethical realists, but they also undercut their own explanations of the phenomenon. With Dreier’s words, what is needed is an explanation of why it is a conceptual truth that if St. Francis is a good man, then there *could not* be a man just like him in non-evaluative respects who is not good. Neither Hare nor Blackburn provides an explanation of that (see Dreier 2015 for further discussion).

So much about the phenomenon that we are seeking to explain. In the following three sections I explain what I think non-factualists should say about the supervenience of the evaluative on the non-evaluative. §4 develops the main idea and §5 presents a generic version of non-factualist semantics for evaluative terms, on the basis of recent discussions concerning predicates of personal taste. §6 then explains how the main idea can be developed within this semantic framework and §7 moves on to compare this account to other explanations of Strong Supervenience in the literature.

## The main idea

To get the discussion started, it is instructive to consider a second, to my mind also misguided, line of thought that Hare offered in support of Weak-but-not-Strong Supervenience. In arguing for why the relevant phenomenon is merely Weak Supervenience, Hare writes:

If I call one room of [a specific] kind nice, there must be some universal, though perhaps highly specific and by me unspecifiable, aesthetic attitude that I have; in other words, I have to be subscribing to some universal premise from which, in conjunction with facts about the room, it follows that the room is nice. But my taste might have been different. (Hare, 1984, p. 5)

The last sentence in this quotation is meant to suggest that in a scenario where our tastes were different, other things would instantiate evaluative aesthetic properties. In other words, Hare is assuming that a non-factualist view of evaluative thought and discourse is committed to the fact that, in a possible world where our sensibilities are different, other things would be nice and beautiful. This is meant to support the claim that evaluative properties supervene only weakly on non-evaluative properties. Since Hare is discussing supervenience of the evaluative in general, he is committed to the same line of reasoning in the moral case. Hare’s contention that non-evaluative properties would be differently distributed in non-actual cases where our sensibilities were different commits him to the truth of counterfactual conditionals like the following:

(9) If we didn’t disapprove of kicking dogs, then doing so wouldn’t be wrong.

This is an unpalatable consequence of Hare’s view. Nearly everyone would agree that (9) is false. Moreover, it is a consequence that non-factualists have painstakingly disavowed. Blackburn, for instance, has at multiple locations insisted that expressivists of his own brand are not committed to the truth of conditionals like (9). Such conditionals, Blackburn maintains, read as first-order moral claims concerning wrongness, and are accordingly obviously to be rejected. Very few of us hold first-order evaluative views on which it is okay to kick dogs, just because we happen to fail to exhibit the appropriate attitude towards such actions (see for instance Blackburn 1984, pp. 217–219 and Blackburn 1993, pp. 152–53, 172–74).

In essence, Blackburn’s view of such conditionals is that we evaluate them through our actual sensibilities, rather than through the sensibilities that we have in the worlds to which the antecedent takes us. Or, to express the same idea in a slightly different manner, think of the non-factualist metaphor of evaluations being glasses through which we look at the world. What (9), on Blackburn’s view, is asking us to do is *not* to look at the world through the lenses of someone who does not disapprove of kicking dogs. What it is asking us to do is to look at a counterfactual version of ourselves in which we do not disapprove of kicking dogs, and at dog kicking, with *our own* (our actual) glasses. The result is that we reject (9), since we do not condone dog kicking.

This non-factualist treatment of counterfactual conditionals and other modal contexts holds the key, I want to suggest, to a proper non-factualist account of strong supervenience. To remain within the metaphor of the evaluative being a pair of glasses, the proposal is that the evaluative supervenes strongly on the non-evaluative because, when looking at two things that are similar with respect to their non-evaluative properties they will also appear as similar in their evaluative respects, when looked at through these glasses. Similarly, this will be the case when we are asked to consider the instantiation of evaluative properties in non-actual worlds. Wherever we travel in modal space, we will not find a man who is just like St. Francis in non-evaluative respects but with different evaluative properties, since we bring our evaluative glasses with us on this journey, as it were.

## Non-factualist contents

The next step is the provide a non-factualist semantics in which to develop this idea. Regarding this, we can use the recent discussion of predicates of personal taste as a template (MacFarlane, 2014 provides a thourough discussion). On an intuitive view, often taken for granted in the literature on predicates of personal taste, the extension of such predicates is fixed by the flavours that we appreciate. This idea can be spelled out formally within possible worlds semantics. Within this framework, the semantic values of a sentence are taken to be the set of possible worlds in which that sentence is true. When this model is extended to capture the subjectivity of predicates of personal taste, the semantic value of a sentence in context is taken to be the set of world-taste pairs in which the sentence is true, where a ‘taste’ is analogous to the notion of a possible world. Just as a possible world is ‘maximally decided’ in the sense that all facts in it are settled, a ‘taste’ is maximally decided in the sense that it renders a verdict about every object with respect to whether or not it is tasty.

On this kind of view, the truth conditions of taste-sentences would roughly appear as follows:

(10) ⟦Saltimbocca is tasty⟧c,<w,g> = 1 iff Saltimbocca is tasty according-to-g in w.[[5]](#footnote-5)

In this notation, the double brackets stand for a function that maps an expression to its semantic value. In the case of sentences, this is either truth or falsity (1 or 0). ‘c’ denotes the context of utterance and the angle brackets the points of evaluation, or index, against which the sentence is evaluated for truth and falsity; in the described view this is worlds and gustatory tastes. Non-factualism about ‘tasty’ is a generic view when characterised in this way. The idea that tastiness is settled by the flavours we like is compatible with differing views of the nature of taste predicates. We have so far not been told *whose* taste is relevant for settling the truth of a taste-statement or in *what way*. In addition to these truth conditions, we need what John MacFarlane terms a “post-semantics” that explains what is asserted or expressed by a taste-sentence within a given context. This corresponds to a certain extent to the positions in the literature concerning the predicates of personal taste known as contextualism, relativism and expressivism. Insofar as versions of these views are all compatible with the world-taste-semantics outlined above, they are variations of non-factualism about ‘tasty’ as understood here (cf. MacFarlane, 2014, pp. 167–175; Ninan, 2010; Yalcin, 2014).

To illustrate what the final step might look like, consider MacFarlane’s relativist post-semantics for taste-statements:

**Relativist post-Semantics**: A sentence S is true as used at a context c1 and assessed from context c2 iff S is true at c1, ⟨wc1, gc2⟩, where wc1 is the world of c1 and gc2 is the sensibility of the assessor (Cf. MacFarlane, 2014, pp. 67, 151ff.)

The basic idea of MacFarlane-style relativism is to take taste-statements to be made true or false by the tastes of the individual *assessing* the statement (possibly the speaker). Taste-statements can thus vary in truth value not only over worlds (and possibly times and other parameters), but also in respect to the tastes of individuals who assess them. On this view, the statement:

(11) Saltimbocca is tasty.

is true when assessed by somebody who likes Saltimbocca, but false when assessed by somebody who does not. The same content can thus be true in relation to one person’s taste, but false in relation to another’s, on MacFarlane’s view.

It should be reasonably clear how to define, for instance, contextualist post-semantics along similar lines – just let the point of assessment be determined by the context of utterance. The move to extend the possible worlds framework with a non-factual parameter (a ‘taste’) is also analogous to Gibbard’s expressivist framework in which contents are taken to be pairs of worlds and a formal object named a ‘hyperplan’ (Gibbard, 2003). Exactly how to get to a specifically expressivist view from this more general non-factualist understanding of contents is a contentious issue that I want to avoid in the present context (see MacFarlane, 2014, pp. 167–175 and Yalcin, 2018, for some discussion). Some aspects of Gibbard’s theory are discussed in §7 below.

Thus in summary, several different views about taste predicates are compatible with the general view of contents introduced above. In the present context, I will not focus on these differences. Instead, I will argue that the element that they have in common, the view of contents as sets of pairs of worlds and a non-factualist element, provide the necessary material with which to substantiate the explanation of supervenience that was sketched in the previous section.

As concerns the non-factual element, ‘taste’ is inadequate for our current endeavour, which is meant to account for the supervenience not only of tastiness, but of all evaluative properties. We need something more general, which covers the non-cognitive reaction purportedly associated with moral goodness and badness, virtue, beauty, etc. I propose to use the term ‘sensibilities’ to refer to the extension of gustatory tastes in the formalism above, such that they include maximally decided verdicts concerning cruelty, beauty, goodness, and so forth. Accordingly, we take the semantic values of sentences to be the sets of pairs of worlds and sensibilities in which a given statement is true.

Needless to say, we are making controversial assumptions here. With respect to the gustatory taste-index, it is not hard to understand what type of affective responses it is intended to capture. With respect to the broader notion of sensibilities, however, and terms such as ‘beautiful’, ‘horrible’, ‘cruel’, and ‘morally good’, it is less clear what non-cognitive reaction would play the counterpart to liking and disliking something’s flavour. I am not going to dwell on this question here. We are all familiar with that special feeling we have when finding, say, a sunset beautiful, although most of us would struggle when trying to articulate it. On the hypothesis we are working with here, all evaluative terms are the same; they are tied to specific non-representational states, similar to how ‘tasty’ is tied to the liking of something’s flavour. In contrast to the case of ‘tasty’, however, we often lack specific locutions for these non-representational states beyond ‘finding something beautiful/horrible/cruel/good’.

## Explaining supervenience

The next step is to show how the non-factualist explanation of supervenience outlined in §4 can be substantiated within this semantic framework. Recall, the supervenience relation that we seek to explain is the following:

It is a conceptual necessity that, if anything x has some evaluative property F, then there is at least one non-evaluative property G such that x has G, and necessarily everything that has G has F.

Formally, this is captured by the following formulation, as noted above:

**Strong Supervenience:** ☐ (∀F in α)(∀x)[Fx → (∃G in β)(Gx & ☐ (∀y)(Gy → Fy)]

The argument will proceed in two steps. First, it will be explained how the proposed semantics implies Weak Supervenience with the outer necessity operator being conceptual; second, it will be explained how a natural interpretation of this view also has as a consequence Strong Supervenience with the inner necessity operator having a metaphysical flavour. In the discussion below, for simplicity I will focus on ‘tasty’ and tastiness. In light of the comments provided in the previous sections, it should be clear how the discussion can be generalised to other evaluative terms.

Again, on non-factualism as developed above, the extension of ‘tasty’ is fixed by which flavours are appreciated (by who is determined in the post-semantics). This is a conceptual truth in the sense that one tacitly knows it if one is a competent user of the term. This by itself implies that tastiness supervenes on the assessor’s attitudes and, consequently, on something that is non-evaluative. However, what we wish to show is that tastiness supervenes on various non-evaluative properties of the object that has the tastiness-property. If we show that our attitudes towards objects supervene on the non-evaluative properties of those objects, we will then have shown that tastiness does as well.

This is not difficult to do. For example, when one likes or dislikes the flavour of an object, that attitude is not directed directly at the object – when appreciating a glass of wine, what one really appreciates is its taste, which is a property, rather than an individual. If another glass of wine tastes (exactly) the same, it instantiates the very same (numerically identical) property.[[6]](#footnote-6) Hence, if one appreciates the taste of one of them, one *ipso facto* appreciates the taste of the other – it is not possible to like the taste of one without liking the taste of the other since the attitude is property-oriented.

This line of reasoning shows that the semantics of ‘tasty’ outlined above implies that tastiness weakly supervenes on the non-evaluative properties of tasty objects, specifically on their taste (in the non-evaluative sense of the word). We have seen that it is a conceptual truth that the extension of ‘tasty’ is fixed by the assessor’s taste on this view. It is also a conceptual truth that if one likes the taste of something, then one automatically likes the taste of another thing that tastes the same. This is so because what one directly likes is a property (the taste/flavour). It is therefore conceptually necessary that if a glass of wine is tasty, there is a property of that glass of wine, the way it tastes, such that everything that has this property is also tasty. Consequently, the evaluative property of being tasty weakly supervenes on the non-evaluative with conceptual necessity on the semantics outlined above. This is what we wanted to demonstrate.

Next, consider Strong Supervenience. We want to remind ourselves of the non-factualist metaphor of evaluative properties being lenses through which we look at the actual and non-actual worlds, rather than being part of those worlds. To substantiate this metaphor within the current framework, we want to make it so that when determining what is evaluatively the case in non-actual worlds, we do so with our actual sensibility. This is precisely what is delivered by the current semantic framework. Within the possible-worlds semantics, modal terms are standardly taken to be operators which switch the world of evaluation away from the actual world, to some other specified (set of) possible worlds. With an example from MacFarlane, consider:

(12) If I had not trained my palate on many better wines, Two Buck Chuck[[7]](#footnote-7) would be tasty (MacFarlane, 2014, p. 163).

On the problematic view seemingly endorsed by Hare in the quotation in §4, this should come out as correct. A world in which I have a different taste is a world in which other things are nice and tasty, according to Hare. But as noted by MacFarlane, this is not the result that one gets within the current semantic framework. On the standard view of counterfactual conditionals, these are true if the consequent clause is true in all the closest possible worlds where the antecedent is true (Lewis, 2001/1973). The upshot is that when statements like (12) are evaluated, the taste that is relevant for settling which wines are tasty in the non-actual world to which the antecedent takes me, is the taste that I actually have.

Stated in quasi-technical terms, it is the world-index that is shifted by the antecedent in a counterfactual conditional. We are thus asked to evaluate the truth of the consequent clause in all the closest possible worlds in which the antecedent is true, instead of evaluating it in respect to the actual world. The taste index, g, remains the same. The gustatory taste that is relevant for determining the truth of (12) is not the taste that I have in the relevant alternative possibilities, but rather my taste in the actual world (cf. MacFarlane, 2014, p. 165). All modal operations will deliver the same result within this framework. That is, whenever I look at the extension of ‘tasty’ in a non-actual world, it is my (or whoever the post-semantics fixes as relevant) *actual* attitudes that are relevant for settling this. Consequently, there is no metaphysically possible world in which there is an object such that its flavour is pleasing to my taste (the taste that I have in the actual world) without being tasty. This is what we sought to demonstrate.

To reiterate, we have seen that the connection between judging something to be tasty and liking its taste, along with the fact that taste appreciation is directed at properties, makes it the case that it is conceptually necessary that if a particular *a* is tasty, then there is a non-evaluative property such that anything that has this property is necessarily also tasty. We obtain the second ‘necessarily’ since we evaluate with our actual attitudes even when taking a stance on the tastiness of objects in non-actual possibilities. This explanation of supervenience thus relies ultimately on (i) the connection between ‘tasty’ and taste-appreciation (for the outer box) and (ii) the world-index, but not the taste-index, of tasty-statements being shifted in modal contexts (for the inner box). It therefore applies in principle to other “post-semantic” interpretations of the world-taste semantics, such as contextualism and expressivism. These are as well-suited as relativism for explaining the given phenomenon, and we therefore do not have to decide which of these non-factualist views is most plausible when accounting for supervenience.

What about the constraint that no specific evaluative theory should emerge as a conceptual truth? This condition is also met on the current view. Suppose for simplicity that I only like the taste of things which taste like coriander. In that case all and only statements that ascribe tastiness to things which taste like coriander will be true when assessed by me. But on the current view:

(13) Something is tasty if and only if it tastes like coriander.

would still not be a conceptual truth. It would instead be an empirical discovery, not something that follows from the proposed semantics of taste statements.

## Other accounts

This explanation of the supervenience of the evaluative offers some distinct advantages over other views in the metaethical literature. First, as we have already seen, it is superior to Hare’s and Blackburn’s accounts in that it targets Strong Supervenience instead of Weak. It is, to the best of my knowledge, the only non-factualist account that targets Strong Supervenience explicitly. Secondly, the outlined semantics for evaluative terms has the consequence that the supervenience of the evaluative is a conceptual truth, in the sense that it implies that the wide-scope necessity operator has a conceptual flavour. In contrast, many realist accounts of supervenience either fail to accommodate the conceptual nature of the relationship (for instance Leary, 2017) or simply postulate it without explaining it (for instance Enoch, 2011, p. 149).

With this said, there are realist proposals which do seem to capture the right kind of supervenience relationship for moral terms. Dreier argues that appeal to rigid designation, associated with ‘Cornell realism’, advocated by among others Richard Boyd (1988), can do the trick:

Here is what I think the nonreductive naturalist should say: It is a part of the meaning of the word ‘good’ that it names a natural property, just as it is part of the meaning of the word ‘water’ that it names a physical kind. Anyone who thinks two things could be alike in their physical properties while differing in whether they are made of water, reveals ignorance of the meaning of ‘water.’ That is a fact about our linguistic practices. And anyone who thinks two things could be alike in their natural properties while differing in their moral goodness, reveals ignorance of the meaning of ‘good.’ That is a fact about our linguistic practices. (Dreier, 1992, p. 21)

This is a good point – if a term is a rigid designator, this is built into its meaning. This must be the case if we are correct in taking intuitions in such thought experiments as Putnam’s twin-earth at face value. Accordingly, the naturalist can claim that a term like ‘morally good’ rigidly designates some unspecified natural property, and that goodness for this reason supervenes on that property. Cornell realism-style naturalism thus provides an account of why supervenience is a conceptual truth, and, moreover, accommodates the fact that the same (possibly disjunctive) non-evaluative properties underlie moral goodness in every possible world.

An alternative realist explanation of the conceptual relationship, provided by Skarsaune (2015), departs from the thought that it is a conceptual truth that moral properties primarily attaches to *kinds*, and only secondarily to particulars. This, he argues, explains why two individuals sharing the same natural properties, by conceptual necessity share moral properties as well. Skarsaune argues that kinds are the primary bearers of moral properties on the basis of how moral reasoning works:

In descriptive enquiry, we typically go from judgments about particulars to judgments about kinds. So, for example, we might do ornithology in roughly the following fashion:

1. This bird sings in the morning and that bird sings in the morning and yonder bird …—and come to think of it, they are all robins! So it seems robins sing in the morning.

But notice how backwards it would be to try to do ethics in a similar way:

1. This act is wrong and that act is wrong and yonder act …—and by golly, they are all sexual harassments! So it seems sexual harassment is wrong. (Skarsaune 2015, p. 261)

That it is a priori backwards to reason like (B) supports, on Skarsaune’s view, the view that the kind primacy of moral properties is a conceptual truth.

Let us assume that these accounts of the conceptual nature of the supervenience relationship between the moral and non-moral are adequate.[[8]](#footnote-8) They still suffer from a disadvantage of being tailor made for moral cases in particular, rather than for the evaluative as such. When Hare discussed supervenience in his 1952 and 1984 works, he freely moved back and forth between examples from ethics, aesthetics (the example of the painting that I quoted in §1) and taste (the example of the nice room discussed in §4). This is natural, since these kinds of evaluations exhibit the exact same supervenience behaviour in relation to the non-evaluative. This suggests, all things being equal, that they should be given the same kind of explanation. The non-factualist account of Strong Supervenience outlined above achieves precisely that. By contrast, the kind of explanations offered by Dreier and Skarsaune are not plausibly extendable beyond the moral case. While the thought that moral terms work like natural kind terms might seem *prima facie* plausible for someone with realist inclinations regarding the moral, the notion is far-fetched as concerns properties such as being a good painting and being a nice room. Many committed moral realists, would, I think, balk at a view according to which ‘nice room’ is a natural kind term. Extending Skarsaune’s explanation to matters of aesthetics and taste would incur similar ontological woes.

In addition, the epistemological considerations offered by Skarsaune to motivate that moral properties attach primarily to kinds are not plausibly extendable to the other forms of evaluation. By no means do we reason from principles when deciding whether a painting is good or a room is nice. These examples seem more similar to the robin case in the quotation above than to the moral case.

These considerations are not, however, conclusive. A good way of looking at them is instead as making explicit the commitments incurred by proponents of these ways of explaining the supervenience relationship. They can either bite the bullet and accept what seems like an implausibly strong form of realism for the evaluative across the board, or simply deny that the supervenience of the evaluative calls for a unitary explanation. When compared to a theory of the latter kind, it is a distinct advantage to the view offered here that it provides a unitary explanation of a phenomenon that seems to call for such.

These considerations are also relevant when comparing the proposed explanation of supervenience to that of Gibbard (2003). As noticed in §5, Gibbard operates with a formal framework that is similar to the one outlined here, but with the difference that the non-factual element is called a ‘hyperplan’ instead of a ‘sensibility’. Hyperplans are thought of as modelling maximally decided states of planning, that is, states in which one has decided, with respect to every conceivable action-situation, whether that action is “okay to do” or not. Contents, on Gibbard’s proposal, are thought of as sets of hyperplans and worlds. The set of world-hyperplan pairs which represent my current state of mind are those worlds that are compatible with my current state of information, paired with all the hyperplans that are compatible with my current state of planning with respect to what is okay to do. When I make an ought-claim, the primary target of Gibbard’s semantic, I express that I am in a state of mind represented by all the world-hyperplan pairs which rules out the other courses of action in that situation.

Gibbard’s explanation of the supervenience of ought-claims on matters of fact proceeds as follows. Plans, according to Gibbard, are such that “if two acts in two possible situations differ in no prosaically factual way, a plan can’t distinguish them, permitting one and ruling out the other” (Gibbard 2003, p. 92). Therefore, a perfectly determined planner, an individual whose planning state is represented by only a single hyperplan, will have a plan whose verdicts about what is “okay to do” in any given situation yield the same verdict if the situations are qualitatively similar. For a hyperplanner, which actions are permitted and obligatory will therefore supervene on “prosaically factual” features of action situations. Next, Gibbard argues that even finite planners such as ourselves, who do not have a determinate verdict about what is permitted to do in any conceivable situation, are “committed to” making our planning judgements supervene on the factual situations. The reason for this is as follows. A person, is, according to Gibbard, committed to any claim that holds in all hyperstates which he could arrive at without changing his mind, that is, in all hyperstates that are compatible with his current state of relative uncertainty. Since in all such hyperstates the person’s plan supervenes on the factual, all finite planners are committed to the supervenience claim with respect to their ought-claims (Gibbard 2003, p. 88–94).

There is a concern regarding Gibbard’s explanation of supervenience. It seems at best unclear whether a perfectly determinate planer, a hyperplanner, really needs to treat qualifiedly identical situations in the same way. There seems to be nothing inconsistent in the notion of a planner who treats qualitatively identical situations differently, opting to act in one way in one situation, and in another way in another situation. Hyperplanners, rational as they may be, might enjoy acting on a whim. This point is discussed in some more detail by Sturgeon (2009, p. 84–87).

It is thus open to doubt whether Gibbard has really managed to explain supervenience within his metaethical theory. But it should be noted that even if there are sufficient Gibbardian responses to this concern, the point made in relation to the naturalist and non-naturalist theories above also applies to Gibbard’s proposal. Gibbard provides a semantics for the ‘ought’ in particular, and there is no obvious way to extend his proposal to evaluative words and concepts in general. The current proposal is thus, again, more general. With this said, since Gibbard’s discussion targets ‘ought’ in particular, whereas the current proposal focuses on the evaluative, they are in principle compatible.

In the previous discussion, the ability of the current proposal to account for supervenience in general, in contrast to only the moral, has been presented as an asset in comparison with competing accounts. However, there is a possible concern that one might have in relation to this. This is the concern that the proposed account of supervenience makes different kinds evaluative predicates *too similar*. Is it not, one might worry, obvious that statements to the effect that Saltimbocca is tasty or that a room is nice are very different from serious moral claims. On our proposed view, taste statements have analogous semantics to that of:

(14) Heinrich Himmler’s actions were despicable.

But surely, taste statements carry their subjectivity on their sleeves in a way that (14) does not, one could be inclined to argue. Should that not be somehow manifest in the semantic representation of the respective predicates? This line of thought could be taken to undermine the claim made here, that a unitary explanation should be given of the supervenience of the evaluative on the non-evaluative.

The non-factualist can reply to this by pointing out that one can acknowledge that there is a stark difference between matters of taste on the one hand and moral matters on the other, but that this difference is not necessarily best represented as a matter of the semantics of the relevant predicates. Instead, the non-factualist will want to treat it as a normative or evaluative difference: moral matters are immensely more important than matters of taste. Acknowledging this difference does not, however, commit the non-factualist to holding that there is a difference in their ontology, or that we have to give up on providing a unitary semantic treatment of evaluative terms.

## Concluding Remarks

Contrary to what is held by Hare and Blackburn, there is no “supervenience challenge” for moral realism, in the sense that there is a phenomenon that only non-factualists are in a position to account for. The reason for this is that the challenge issued by Hare and Blackburn mistakenly assumed that the relevant phenomenon was Weak, and not Strong, Supervenience. Still, if what I have argued here is correct, there is a plausible non-factualist explanation of Strong Supervenience which correctly accounts for the conceptual nature of the relationship, and which moreover offers the same explanation for supervenience in aesthetics and matters of taste. As was argued in the previous section, there is no realist explanation of the conceptual nature of the relationship currently on the table, that is plausibly extendable to other forms of evaluation. In this respect, the non-factualist account of supervenience that I have offered here is superior to the alternatives in the literature.[[9]](#footnote-9)

## References

Blackburn, S. (1971). Moral Realism. In J. Casey (Ed.), *Morality and Moral Reasoning*. Methuen.

Blackburn, S. (1984). *Spreading the word: Groundings in the philosophy of language*. Clarendon Press ; Oxford University Press.

Blackburn, S. (1993). *Essays in quasi-realism*. New York: Oxford University Press.

Boyd, R. (1988). How to Be a Moral Realist. In G. Sayre-McCord (Ed.), *Essays on Moral Realism* (pp. 181–228). Cornell University Press.

Dreier, J. (1992). The Supervenience Argument against Moral Realism. *The Southern Journal of Philosophy*, *30*(3), 13–38.

Dreier, J. (2015). Explaining the Quasi-Real. In R. Shafer-Landau (Ed.), Oxford Studies in Metaethics, Volume 10 (pp. 273–298). Oxford University Press.

Dreier, J. (2019). Is there a supervenience problem for robust moral realism? Philosophical Studies, 176(6), 1391–1408.

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

Foot, P. (2001). *Natural goodness*. Oxford University Press.

Gibbard, A. (2003). *Thinking how to live* (1. paperback ed). Cambridge, Mass.: Harvard Univ. Press.

Hare, R. M. (1963). *The Language of Morals*. Oxford University Press.

Hare, R. M. (1984). Supervenience. *Aristotelian Society Supplementary Volume*, *58*(1), 1–16.

Horgan, T., & Timmons, M. (1992). Troubles on moral twin earth: Moral queerness revived. *Synthese*, *92*(2), 221–260.

Leary, S. (2017). Non-naturalism and Normative Necessities. In R. Shafer-Landau (Ed.), *Oxford Studies in Metaethics* (Vol. 12). Oxford University Press.

Lewis, D. K. (2001). *Counterfactuals* (Rev. ed.). Malden, Mass: Blackwell Publishers.

MacFarlane, J. (2014). *Assessment sensitivity: relative truth and its applications* (First edition). Oxford: Oxford University Press.

McPherson, T. (2012). Ethical Non-Naturalism and the Metaphysics of Supervenience. In R. Shafer-Landau (Ed.), *Oxford Studies in Metaethics Volume 7* (pp. 205–234). Oxford University Press.

McPherson, T. (2015). Supervenience in Ethics. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2015). Metaphysics Research Lab, Stanford University. Retrieved from https://plato.stanford.edu/archives/win2015/entries/supervenience-ethics/

Ninan, D. (2010). Semantics and the objects of assertion. *Linguistics and Philosophy*, *33*(5), 355–380.

Sibley, F. (1965). Aesthetic and Nonaesthetic. *The Philosophical Review*, *74*(2), 135–159.

Skarsaune, K. O. (2015). How to Be a Moral Platonist. In R. Shafer-Landau (Ed.), *Oxford Studies in Metaethics, Volume 10* (pp. 245–272). Oxford University Press.

Sturgeon, N. (2009). Doubts about the Supervenience of the Ethical. In R. Shafer-Landau (Ed.), *Oxford Studies in Metaethics: Volume Four*. Oxford University Press.

Yalcin, S. (2014). Semantics and Metasemantics in the Context of Generative Grammar. In A. Burgess & B. Sherman (Eds.), *Metasemantics* (pp. 17–54). Oxford University Press.

Yalcin, S. (2018). Expressivism by Force. In D. Fogal, D. W. Harris, & M. Moss (Eds.), *New Work on Speech Acts*. Oxford University Press.

1. This article does not engage with analytical naturalists, such as Philippa Foot (2001), who believe that there are conceptual entailments from the descriptive to the evaluative. [↑](#footnote-ref-1)
2. The open question argument in fact targets only the moral in particular. Frank Sibley (1965) makes a similar point about evaluative aesthetic terms. [↑](#footnote-ref-2)
3. It is customary within aesthetics to distinguish between evaluative and non-evaluative aesthetic terms and properties. My use of ‘aesthetical’ in the discussion below refers specifically to evaluative aesthetic terms and properties. [↑](#footnote-ref-3)
4. However, if Dreier’s argument is to justify ascribing a metaphysical flavour to the inner box in the formulation of Strong Supervenience, he must assume that ‘could’ quantifies over *all* metaphysically possible worlds and not, for instance, merely over the closest ones. This appears to be potentially questionable. If the quantification of ‘could’ is restricted to close worlds, then the argument is compatible with the subscript of the inner box having a more restricted flavour than metaphysical necessity. [↑](#footnote-ref-4)
5. For the sake of simplicity, I do not take into consideration other parameters, such as time and variable assignments. [↑](#footnote-ref-5)
6. If properties are tropes, and hence not numerically identical across objects, one would have to speak of attitudes as directed at what an object is qualitatively like, rather than at properties, to arrive at the same conclusion. [↑](#footnote-ref-6)
7. A very cheap wine. [↑](#footnote-ref-7)
8. However, see (Mitchell, 2016, p. 2919–2922) for an argument to the effect that appeal to rigid designation fails to capture what distinguishes moral inquiry from scientific inquiry. Moreover, the moral twin-earth considerations offered by Horgan and Timmons (1992) have been taken by many to speak against the assimilation of moral terms to natural kind terms. [↑](#footnote-ref-8)
9. This research is supported by The Swedish Research Council grant number **2019-02905.** Thanks to Karl Bergman, Matti Eklund, Anandi Hattiangadi, Jens Johansson, Andrew Reisner, Olle Risberg, Henrik Rydéhn, Andreas Stokke, Pekka Väyrynen and the participant at the Uppsala higher seminar in practical philosophy for providing comments on earlier versions of this article. [↑](#footnote-ref-9)