

DISLODGING BUTTERFLIES FROM THE

SUPERVENIENT*

Toni Rønnow-Rasmussen

Lund University

Toni.Ronnow-Rasmussen@fil.lu.se

Not all areas in value theory are battlegrounds. We find, for instance, a strong consensus when it comes to whether or not values are so-called *supervenient* properties, i.e., properties that accrue to the value bearer in virtue of some or all of its other kinds of (subvenient) properties. These ‘other properties’ are often assumed to belong (at least at some basic level) to the object’s so-called natural properties (in a wide sense of ‘natural’ that would include, for instance, psychological features).¹ Unfortunately, this consensus does not extend to questions concerning the precise nature of this relation. Just how we should best describe the linkage between natural and value properties has been a much-discussed topic since Moore put the matter on the value theorist’s agenda. In this paper I discuss, after some preliminary comments in section 1, an early attempt to explain supervenience, viz., R. M. Hare’s view as it is presented in his article “Supervenience”.²

1. Two Intuitions about Supervenience.

The supervenience thesis regarding value properties has customarily been looked upon as doing at least two kinds of jobs: one being to express what we may call our *dependence-intuition*, that not just any answer to the question ‘what

* I am indebted to David Alm, David Bengtsson, Johan Bränmark, Erik Carlsson, Sven Danielsson, Dan Egonsson, Jonas Olsson, Robert Pulvertaft, Wlodek Rabinowicz, Caj Strandberg, and Daniel Svensson for helpful discussion. The work on this paper was supported by generous grants from The Bank of Sweden Tercentenary Foundation, and the Tornblad Foundation.

¹ Just what we should understand by a natural property is often considered a vexed matter. Obviously we cannot just say that it is the kind of properties that physical things have. However, I believe that if we started ostentively to pick out what we meant by natural properties we would eventually end up agreeing about a very large number of them.

² Hare, R. M. (1984), “Supervenience”, *Aristotelian Society*. Supp. 58. Reprinted in Hare (1989) *Essays in Ethical Theory*, Oxford: Clarendon Press.

makes object x valuable?’ is equally intelligible. Answers such as ‘Oh, it just is valuable! There is nothing that makes it valuable’ are regarded as mysterious, as utterances that defy comprehension: Values are not like butterflies that happen to settle on a flower. Whatever we understand by ‘value’ it surely does not refer to some entity that merely happens to accrue to objects. There is a dependence relation here; the value of the object results from its having the natural properties it has. The value of, say, a painting must have something to do with the kind of painting it is. If it is ugly or beautiful, it is because of something about this painting, some non-evaluative aspect of it. Similarly, a good person is good in virtue of some of his non-evaluative characteristics.

Supervenient properties are ‘dependent properties’. To insist that value always accrues to objects ‘*because of*’ the properties of these objects is to express a belief, by implication, that there is a kind of dependence relation between value and those other subvenient properties.³

The second idea that the supervenience claim customarily is taken to express—the *consistency intuition*—is the following: if you assert that x is valuable and if you agree that y is relevantly similar to x , with regard to natural properties, you must be prepared to assert that y too is valuable. To deny it would leave the listener in a similar kind of perplexity that follows upon hearing someone express something contradictory. Now, I take it that if we do believe that values are so to speak free floating ontological entities—what amounts to the ‘butterfly’ rejection of supervenience—that just land on objects (haphazardly or because they were bestowed on objects according to, say, God’s will), it would be a mistake to conclude that we are actually logically committed to ascribing value to y . It is at least not required by consistency to draw the conclusion that y is valuable, for x and y may differ precisely in that the value has ‘landed’ on x but not on y . Again, does the fact that we can imagine such “value butterflies” not in fact show that the supervenience thesis is not a conceptual thesis but one expressing our deeply rooted beliefs about what and how things are? However, this kind of reasoning is flawed. From the fact that we *seem* to be able to imagine without logical contradiction something such as “free floating entities”, it does not follow that we can claim that we are conceptually able to imagine values as such entities.

The fact that such “butterfly” reasoning seems to be fundamentally confused should not obscure the fact that it is hard to come up with a knock-down argument showing that beliefs in a logical supervenience thesis are not eventually based on certain ontological convictions. I am not sure what such an argument would look like. Whether this mistrust warrants a serious disbelief in

³ This does not seem that obviously true in the case of mental predicates. At least it does not seem implausible that we, *qua* language users, have been mental predicaters (long) before we became aware of the body mind problem.

supervenience is another matter. For my own part, I don't think so. At least, I am certainly prone to say that value terms are for purely conceptual reasons supervenient terms. I will in what follows treat the supervenience thesis as merely a conceptual thesis. Here I am just making it clear that I don't know how to settle beyond question whether my intuitions are purely linguistic or not.

As mentioned earlier, however, there is, or so it has been argued, more to supervenience than that if two things, x and y , are similar with regard to natural properties, then they must—and 'must' here is a conceptual 'must'—have similar value properties. In recognizing that the supervenience thesis expresses a consistency intuition, we must not lose sight of the further important relation that is said to hold between natural properties and value properties, viz., the former necessitates the latter ones. The precise nature of this necessity relation may be moral or metaphysical or it might be some other kind of necessity.⁴

2. Hare's "Subscription" Account of Supervenience.

An influential approach to supervenience is found in the works of R. M. Hare. Hare, who often is credited for introducing the term to the philosophical community,⁵ opposed every attempt to turn moral philosophy into a division of metaphysics. Talk about moral *properties* in a realistic sense was fundamentally alien to his non-cognitivist and irrealist position.⁶ In "Supervenience" (Hare 1984) he asks his reader to consider the following deduction (I have used my own abbreviations rather than Hare's: 'N' refers to a natural property or a set of such properties, and 'V' to a value property).

$$\begin{array}{l} \mathbf{p}: \text{For all } x, \text{ if } Nx \text{ then } Vx \\ \mathbf{q}: Na \\ \hline \text{So: } \mathbf{r}: Va \end{array}$$

According to Hare, when Va ascribes a supervenient property to a the following holds true:

⁴ Cf., Sven Danielsson, in "The Supervenience of Intrinsic Value" (see eds., Erik Carlsson and Rysiek Sliwinski, *Omnium-gatherum. Philosophical essays dedicated to Jan Österberg on the occasion of his sixtieth birthday*, Uppsala Philosophical Studies 50, pp. 93-103, 2001), who suggests that the kind of necessity involved here means "something like 'fact-independently'. If Socrates is good because he is brave and wise, then, whatever the facts that may obtain, anyone brave and wise would be good" (p. 95).

⁵ Hare denied that he was the first to use the term "supervenience"; see Hare, R. M. (1984).

⁶ I suspect that it was this that made him take a generous view of the relation of the supervenience relation. See here Rønnow-Rasmussen *Logic, Facts and Representation: An examination of R. M. Hare's Moral Philosophy*, Lund University Press, 1993, pp. 147-148.

Necessarily, if **r**, then there is a valid inference of the ‘**p**, **q**, so **r**’ form, the two premisses of which hold. (op. cit., p. 70)

There are notably two aspects of Hare’s account that deserves to be highlighted. Perhaps the most striking side to Hare’s approach to value supervenience is that it is indirect; he is not telling us (in terms of necessary and sufficient conditions) what the relation is between natural properties and value properties. Rather, he purports to elucidate this relation *via* a specification of certain conditions, which the speaker who claims Va has to satisfy. His reason for approaching supervenience in this way is intimately connected with his prescriptive analysis of value terms. This brings us to the next salient feature of his explanation of value supervenience—namely the one that concerns the nature of the premise in **p**.

Different metaethical views claim that statements like the one in premise **p** are true as a consequence of the very meaning of the value term in it or because they express something necessarily true (a synthetic truth). A prescriptivist cannot (in any obvious way) avail himself of this option; evaluative judgments such as the one in **p** do not express any truth (or, for that matter, falsity) at all. In fact, we do not even, in Hare’s opinion, assert anything by saying that something is, say, good: we prescribe, and prescriptions, in his view, do not carry truth-values.

So what does Hare replace “truth/falsity” with? Somewhat disappointingly Hare merely says that the premise is held in the sense of being “subscribed to”. The speaker is committed to some such principle. Now, Hare might be right about what we may, on logical grounds, expect of someone who issues **r**. In fact, I do think there is some truth to his claim. However, the pressing question at this moment is to ask whether the “subscription” condition exhausts all there is to the supervenience relation.

3. A Flaw in Hare’s Account.

In what sense is the dependence relation of, say, the beauty of la Gioconda and its colors and forms explained by reference to what the speaker as a matter of consistency has to subscribe to? Of course, we might ask something different, e.g., why does the subject think that la Gioconda is beautiful? A perfectly reasonable reply would be to point to the fact that the subject subscribes to some principle such as the one in **p**. But this reply says nothing about the nature of the dependence that allegedly exists between the beauty of la Gioconda and its non-evaluative properties. The problem here is that the “subscription account” of supervenience leaves out the important dependence aspect of the relation at issue. Hare’s account may or may not in part explain why the *agent* claims or has to claim the conclusion, Vx —the reason being that that is what consistency demands of him when prior to this statement Vx he has claimed Va . But what it

fails to do is to explain that there is a relation in the first place between V and (certain or all of) the natural properties of *a* (N). Consistency can't do that!

It might be objected that I am taking a much too realistic approach to this issue. After all, according to prescriptivism there is something fishy about “value property”-statements. The explanation that is missing above seems to set out from the assumption that there is a “real” property relation to be found. And since prescriptivists adhere to irrealism about evaluative properties there is, in their view, no further relation for which they need to account. But this response, as I will argue below, is surely barking up the wrong tree.

4. Value Realism/Irrealism and Supervenience.

The idea that the “dependence” relation between supervenient properties and subvenient ones requires for its meaningfulness some sort of realist ontology is fundamentally a misconception of the issue. But assume for the sake of argument that value irrealism is a correct view, and that some kind of error theory therefore is valid;⁷ our language leads us to believe that we predicate “real” evaluative properties of things when in fact no such properties exist—they are not, so to speak, to be found *in rerum natura*. What light does this assumption then throw on the logical relations that exist between value predications and natural property predications? They would hardly disappear; these logical features obtain, of course, as long as we go on using the “erroneous” way of speaking. We might replace a language containing supervenient terms with one that lacks such terms, and so, if successful, get a language that is *allegedly* more faithful to an irrealist ontology. But I take it that this is not the issue here. In a language containing no supervenient terms there is for obvious reason no need for a supervenience account. But to recognize this fact cannot be taken as a justification for saying that there is no logic governing the relation between “supervenient judgments” and “subvenient judgments” in a language that *does* in fact contain supervenient terms. Hence, unless prescriptivists are eliminativists about supervenient terms, which Hare is not (Rønnow-Rasmussen 1993), their account of supervenience ought to take into

⁷ Even Hare would, I suspect, agree that evaluative language at first sight appears to have more in common with descriptive language than with prescriptive. (“x is good” looks more like “x is yellow” than, say, the prescriptive “Take x!”). We have, in other words, some initial reason to regard evaluative language as a descriptive language. What Hare wanted us to avoid was to commit what Austin referred to as the Descriptive fallacy, i.e., the attempt to analyze all meaning as *descriptive* meaning. Once we have got a more complete picture of what we do in issuing evaluative judgments, it is clear, according to Hare, that such judgments cannot be analyzed along the lines of descriptive language. Evaluative judgments have primarily a prescriptive meaning. The grammatical and syntactical elements of language may be misleading. Today many philosophers do agree that our evaluative language at least appears to be a language of assertions rather than of prescriptions (there is no agreement, though, as to whether appearances deceive us or not).

consideration all the logical features of supervenient terms. As I have argued, one such logical feature of value is that which prevents us from believing that value properties behave like butterflies. However we want to account for the dependence relation between value properties and natural properties, the account is subject to the following test: does the account block the way for a “butterfly interpretation” of value? If it does not, there is something wrong with it.

Here is a further brief consideration regarding the irrealist reply: Instead of talking about properties, we could focus on the relation between the *judgment* (1) that *a* has the evaluative property V, and the *judgment* (2) that *a* has the set of natural properties N. An advocate of “judgmental butterfly” explanations might say that there is no relation between (1) and (2) but the one established purely haphazardly. That there is a dependence relation here of some kind has nothing to do with whether you are a realist or irrealist about value properties. The logical relations would still hold even if one or both of these judgments were false. Acknowledging that ‘good’ and ‘ought’, to mention only two paradigmatic supervenient terms, have the logical dependence features noticed above does not tilt the balance in favor of realism or irrealism.

It might be replied that irrealism is at least a strong incentive for getting rid of certain logical features related to supervenient terms. For sure, we are free to ignore one or more of these features, and go on to reform the language. However, this reformist approach must make it clear that the reformist term does not have the same sense as the original term any longer. For my own part I have nothing against reformist approaches to this matter. However, at present I do think it is too early to throw in the towel when it comes to giving an account of all the features of supervenience, including the “dependence-intuition”.

The Harean “subscription” condition fails the test mentioned above. Complying with this condition certainly explains why a person has to ascribe a certain property to an object, but consistency alone is not sufficient to rule out all “butterflies”. Consistency is actually consistent with what we might refer to as the “addicted butterfly”; such a butterfly settles haphazardly on, say, a red rose, and from then on it only settles on red roses. Consistency alone wouldn’t ban a use of evaluative terms along the lines of the addicted butterfly.

Hare’s subscription account will therefore not do! His account will tell us why the person who holds (1) and (2) above has to endorse the judgment (3) that *b* has the evaluative property V, if he also holds the judgment (4) that *b* is relevantly similar to *a* with regard to N. According to Hare’s account, if the person refused to endorse (3) we would not understand what he was saying. But again, all this might well be accurate, in fact I think it is true, yet it would nevertheless not explain why the person who “initiated” this consistency standard by claiming (1) and (2) prior to (3) and (4) could not get away with the claim that *a* has V regardless of what other properties it has. The person might

even comply with consistency (as the addicted butterfly) in the sense that he would apply V to objects that are N just in order to comply with this consistency requirement. But doing so would not compel him to take back his initial claim that there is no relation at all between the judgments (1) and (2).

5. How to Block the Way for Butterfly Properties.

The remedy against butterfly properties seems rather obvious. If our aim is to dislodge butterflies from value supervenience, we need to formulate a relation between value properties and natural ones such that if the latter are present, then in every possible world the former ones are too. We need, in other words, to go from intraworld supervenience⁸ to interworld supervenience or, as Kim called it, strong “supervenience” (Recall that ‘V’ refers to value, ‘N’ to natural properties, and that ‘Nec’ is a necessity operator):⁹

Strong Supervenience: $\text{Nec } \forall x (Vx \rightarrow \exists N (Nx \text{ and } \text{Nec } \forall y (Ny \rightarrow Vy)))$

If this strong relation is what the supervenience relation is all about,¹⁰ then it effectively blocks the way for properties that behave in analogy with the addicted butterfly; such properties cannot be supervenient properties. There is a necessary relation (the second one above) of some sort (moral necessity, metaphysical necessity?) between N-properties and V-properties that excludes this possibility.

The remedy against butterfly properties was apparently not that difficult. Notwithstanding, the question remains if Hare actually can take advantage of the strong supervenience thesis, as long as he is faithful to his indirect, prescriptive “speaker” approach.

⁸ So-called weak supervenience— $\text{Nec } \forall x (Vx \rightarrow \exists N (Nx \text{ and } \forall y (Ny \rightarrow Vy)))$ —doesn’t exclude value properties from behaving as butterflies. It only tells us that they do not behave like this in this world.

⁹ See Kim, J. *Supervenience and Mind* (Cambridge University Press, 1993).

¹⁰ The strong supervenience thesis does capture something essential about value supervenience. However, as far as I can see it leaves out one important aspect, namely what we might call the ‘direction-feature’ of the “because of” or “in virtue of” relation. These relations suggest that moral properties and natural ones are connected in a special sense—the latter ones give rise to (or support) the former, and not vice versa. Consider the relation between “x being colored” and “x being extended”; the relation between being colored and being extended is not, in my view, an example of a ‘because of’ or ‘in virtue of’ relation. But suppose we reformulated the Strong Supervenience thesis in terms of ‘Color- and Extension-properties’; in that case the strong supervenience thesis does seem to fit the relation that exists between “color and extension”. In other words, if we do not want to say that color and extension exemplify the same kind of relation that is found between moral properties and natural ones, we need to qualify the strong supervenience thesis in some way.

Suppose we set the prescriptive feature of this *speaker approach* aside for a moment. Then, on such an indirect way of elucidating the supervenience relation, all what is needed—from a descriptivist’s point of view—to block the way for “butterfly” properties is to find some suitable *belief* that is incompatible with a butterfly view, which the value term expresses. What seems necessary is to require of the speaker that he signals by his utterance that he grounds his value judgment in light of the subvenient properties (and not merely in a consistency requirement). More accurately put, the person who asserts that *a* is good must do so believing that the goodness of *a* is dependent upon (at least in the weak sense of being necessarily linked to) some or all of *a*’s natural properties in all possible worlds. On a descriptivist approach (i.e., a non-prescriptive way of understanding the meaning of value terms) we could logically require such a belief from the person if knowing the meaning of “*a* is good” is to know (among other things) that *a* is subsumable under some true thesis such as strong supervenience thesis. The “speaker approach” is, in other words, in itself no obstacle if we want to drive off butterflies from the supervenient.

Let us next bring in Hare’s idea that evaluative judgments are prescriptive.¹¹ A prescriptivist such as Hare, who neither recognizes any entailment relations between natural properties and value properties, nor any synthetic true value judgments, would surely reject the idea that we are logically required (in virtue of some descriptive meaning) to hold certain beliefs when we make evaluative judgments. The only fixed meaning that value terms such as, say, ‘good’ carries (a so-called primary evaluative term) is its prescriptive meaning; whatever descriptive meaning value terms have, it is not determined and fixed. The real crux of the matter concerns therefore how a prescriptivist would account for the second necessity operator in the strong supervenience thesis. It is one thing to claim that in endorsing Va we commit ourselves by conceptual necessity to subscribe to a principle like the one in premise p. It is quite another thing to say that endorsing Va commits you, by conceptual necessity, to subscribe to a principle that in part expresses that there holds a *necessity* relation between certain natural properties and a certain value property. The latter claim squares badly with his idea that value terms have no fixed descriptive content. Still, there might be an attenuated sense of ‘necessity’—say, in analogy with the

¹¹ In Rønnow-Rasmussen (1993, chap. II) I defend the idea that Hare explains the speech act of prescribing in terms of the following three theses: **(P)** A speaker *a* performs a prescriptive speech act in uttering the sentence (s), only if *a* by uttering (s) intends to tell the addressee(s) to make something the case—where this something is determined by (s) and where the term ‘tell’ should be understood in its generical sense. **(SC)** A speaker *a* prescribes sincerely only if (P) holds, and only if he intends that the addressee(s) should do what is prescribed. **(I)** For *a* to intend that the addressee(s) should do *x*—what is prescribed (where *a* and the addressee can be the same person)—is for *a* to do something which *a* believes will bring about that the addressee(s) does *x* (if it is in *a*’s physical and psychological power to do so).

“thin”-notion of truth”—that might do the job here for prescriptivists. But even if we granted this, the question remains whether this “thin necessity” actually can do the job of dislodging butterflies from the supervenient. Presently, I don’t see how Hare would account for necessity in such a way that it will block out butterfly uses of evaluative terms, *without* given up the idea that evaluative judgments do not carry truth-values.

6. Prescriptivism and Supervenience.

Although I find it hard to see what meaning ‘necessity’ carries in a prescription that has no alethic value, I do not see why prescriptivists might not, after all, be able to acknowledge that evaluative terms, in virtue of their meaning, do have an additional meaning element besides their prescriptive one? This is not the place to give a detailed answer to how this can be done. However, this much can be said: This further, element must not interfere with the (primary) prescriptive function of value terms, or somehow commit the speaker to any substantive evaluative views. If there is such an element, I do not see why a prescriptivist couldn’t embrace the idea that value terms do have some descriptive meaning. I suspect there is such a further element.

Consider two descriptive expressions such as ‘*a* is a cause (of *b*)’ and ‘*c* is a cause (of *d*)’. It is plausible to argue that *a* and *c* share some properties that made me call them causes. But having said that, obviously it would be wrong to infer that what *a* and *c* have in common is what actually did the causing of *b* and *d*. Surely that does not follow. If this is true then we may conclude that to know at least part of the meaning of ‘cause’ is not to know what actual causes there are. It is not to subscribe to some substantive view on what causes what! By analogy, it could be argued that there is a common element, besides prescriptivity, to expressions such as ‘*a* is good’ and ‘*b* is good’ that I have to know in order to know the meaning of ‘good’. But this common property is not one of the good-making properties of *a* or *b*. In other words, acknowledging such a meaning-element does not commit the speaker to any substantive view about what is good. The precise nature of this meaning element is still unclear to me. Here is a suggestion the details of which have to be left to another work: It is a logical requirement (sincerity condition) that to assert that *a* is good is among other things to *believe* (rather than to express a preference) that *a*’s non-evaluative properties give me a reason to ascribe goodness to the object. To require such a belief of a speaker does not, as far as I can see, commit the speaker to any substantive value judgments. But it does give the prescriptivist a way of responding to a person who regards goodness as a butterfly-entity: such a speaker is not complying with the above condition.

7. Concluding Remarks.

The notion of supervenience expresses our *dependence-intuition*, i.e., our idea that there is some kind of relation between value and the natural properties of the value bearer that is stronger than the one holding between objects that are merely haphazardly related to each other. Supervenience also covers another idea, viz., that if you agree that an object is a value bearer then you have to agree that any other object that is exact similar (but not numerically identical) to it with regard to natural properties will also be a value bearer. Hare's explanation of supervenience gets much right. However, his "subscription explanation" is marred by its failure to account for the dependence aspect of the relation at issue. A remedy is to stress that the speaker signals by his evaluative utterance that the value judgment is *grounded* in light of the subvenient properties and not merely in a consistency requirement.