Johannes Schmitt and Mark Schroeder
University of Southern California
February 21, 2009

supervenience arguments under relaxed assumptions

provisionally to be forthcoming in *Philosophical Studies*

1.1 introduction

When it comes to evaluating reductive hypotheses in metaphysics, supervenience arguments are the tools of the trade. Jaegwon Kim and Frank Jackson have argued, respectively, that strong and global supervenience are sufficient for reduction, and others have argued that supervenience theses stand in need of the kind of explanation that reductive hypotheses are particularly suited to provide. Simon Blackburn's arguments about what he claims are the specifically problematic features of the supervenience of the moral on the natural have also been influential. But most discussions of these arguments have proceeded under the strong and restrictive assumptions of the S5 modal logic. In this paper we aim to remedy that defect, by illustrating in an accessible way what happens to these arguments under relaxed assumptions and why.

The occasion is recent work by Ralph Wedgwood [2007], who seeks to defend non-reductive accounts of moral and mental properties together with strong supervenience, but to evade both the arguments of Kim and Jackson and the explanatory challenge by accepting only the weaker, B, modal logic. In addition to drawing general lessons about what happens to supervenience arguments under relaxed assumptions, our goal is therefore to shed some light on both the virtues and costs of Wedgwood's proposal.

1.2 some hasty background

To understand Wedgwood's attempt to explain supervenience within a non-reductivist framework, we need a few tools from modal logic. It is often assumed that necessity and possibility are governed by what is known as the S5 modal logic, which amounts to making the following three assumptions (the letters we use for these assumptions are typical in modal logic[1]):

(K) $\Box(\varphi \supset \psi) \supset (\Box\varphi \supset \Box\psi)$
(T) $\Box\varphi \supset \varphi$
(E) $\Diamond\varphi \supset \Box\Diamond\varphi$

[1] For example, in Hughes and Cresswell [1996].

In the context of the other axioms, axiom (E) is in turn equivalent to, and it is useful to break it down into, the conjunction of (B) and (4):

(B) $\Diamond\Box\varphi \supset \varphi$
(4) $\Box\varphi \supset \Box\Box\varphi$

The intriguing move that Wedgwood makes is to give up on the idea that all of these assumptions are true, and in particular that (4) is mistaken.

The standard semantics for modal logic allows us to interpret what these axioms amount to, by postulating a two-place 'accessibility' relation among worlds, R, such that '◊P' is true when evaluated at some world, w, just in case there is some world, w*, such that R(w,w*), and 'P' is true when evaluated at w*. Given this interpretation, axiom (E) says that the relation R is euclidean, which means that if any world is related by it to each of two other worlds, then they are related by it to each other. Similarly, axiom (B) says that R is symmetric, and axiom (4) says that it is transitive. (E) and (T) together imply that R is an equivalence relation, and it is since R is an equivalence relation under the S5 axioms, that assuming the S5 axioms allows us to ignore R altogether, and talk simply of 'possible worlds', rather than of possible worlds that are 'accessible' from some particular world.

Just to give a picture, which will be useful in a moment, axiom (E) says that whenever the diagram in figure 1 obtains, the situation in figure 2 obtains, as well. To coin a term, it says that *absolutely* possible worlds are possible relative to each other. While axiom (4) says that whenever the situation in figure 2 obtains, the situation in figure 1 obtains, as well. Given our terminology, it says that *relatively* possible worlds are absolutely possible.[2]

fig 1: absolute possibility

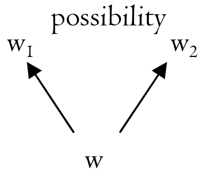
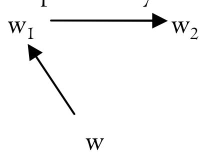


fig 2: world-relative possibility



[2] Point of clarification: in our diagrams, we will follow the convention that the absence of an arrow is the absence of information about accessibility, rather than positive information about non-accessibility – so figure 1 is neutral about whether w_2 is accessible from w_1 and conversely, as well as about whether w is accessible from each of w_1 and w_2, and about whether any world is accessible from itself. To indicate non-accessibility, we will use crossed arrows (as in figure 6, below).

The relationship between these two diagrams is crucially important, because both strong and global supervenience can be and have been formulated in different ways in the literature – ways that correspond to each of the diagrams.

2.1 supervenience theses

Supervenience theses share the feature of claiming that any difference of one kind (an 'A-difference', in the *supervening category*) must be matched by a difference of a certain other kind (a 'B-difference', in the *subvening category*). They differ with respect what counts as an A-difference and a B-difference, and with respect to what goes into the 'must'.

Supervenience theses can be formulated either in natural language, using terms like 'necessarily' and 'possibly', or they can be framed in terms of quantifiers over possible worlds. Characterizations of supervenience claims in terms of possible worlds should not be taken to be committed to the view that 'necessarily' and 'possibly' are object-language quantifiers over possible worlds, or even that they are to be understood in terms of metalanguage quantifiers over possible worlds. Even if 'necessarily' and 'possibly' are primitives, we take it that a suitable notion of 'possible worlds' can be constructed, so that we can use equivalences between natural language claims and possible-worlds talk in order to try to introduce clarity to matters that would otherwise be difficult to sort out.[3] In this and immediately following sections, because we are interested in the logical relationships between supervenience theses, we will follow the convention of framing these theses in terms of quantifiers over possible worlds, as these make these logical relationships clearer. In part 6 we will return to consider what it would take to state equivalent versions of these theses in ordinary English, on the grounds that it is easier to test the *intuitive plausibility* of sentences formulated in ordinary language, which is what will interest us in part 6.

Strong Supervenience

Take strong supervenience first. According to strong supervenience, any two items at any two worlds which share all of their B-properties must also share all of their A-properties. That is, any worlds w_1 and w_2 satisfy the Strong Supervenience Condition[4], or SSC, for short:

---

[3] The issues surrounding quantifying over merely possibly possible worlds, under the assumption that axiom (4) is false, are complex; we propose to finesse those issues here by ignoring them.

[4] $SSC(w_1,w_2) \equiv \forall x \in D(w_1)\ \forall y \in D(w_2)(\text{B-indiscern}(x,y) \rightarrow \text{A-indiscern}(x,y))$, where '$D(w_1)$' picks out the domain of world w_1 – the class of objects existing at that world, and 'B-indiscern(x,y)' is an abbreviated way of saying that x and y share all of their B-properties – i.e., that for each property in B, x has it if and only if y does ($\forall G_{\in B}(Gx \equiv Gy)$) – and similarly for 'A-indiscern(x,y)'. Wedgwood formulates strong supervenience somewhat differently, the main difference being that his formulation requires the

SS $\forall w_1 \forall w_2(SSC(w_1,w_2))$

This is all well and good, if we make the assumptions about necessity and possibility that are codified in S5. But if we deny (E), then it is important to know whether and how the world-quantifiers in SS are restricted. According to one way of precisifying SS, the *absolute* formulation, or SS^{abs}, both world quantifiers range over worlds that are accessible from w – the world with respect to which SS is being evaluated. Whereas according to a second way of precisifying SS, the world-relative formulation, or SS^{WR}, the w_1 quantifier ranges over worlds accessible from w, while the w_2 quantifier ranges over worlds that are accessible from w_1.

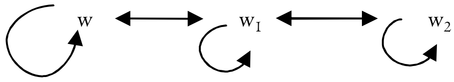
SS^{abs} $\forall w_1{:}R(w,w_1)\forall w_2{:}R(w,w_2)(SSC(w_1,w_2))$
SS^{WR} $\forall w_1{:}R(w,w_1)\forall w_2{:}R(w_1,w_2)(SSC(w_1,w_2))$

So SS^{abs} is true at w just in case for every situation like that depicted in figure 1, w_1 and w_2 satisfy SSC. And SS^{WR} is true at w just in case for every situation like that depicted in figure 2, w_1 and w_2 satisfy SSC. So it follows that since according to (E) every figure 1 situation is a figure 2 situation, whatever goes for figure 2 situations goes for figure 1 situations as well, and hence SS^{WR} together with (E) entails SS^{abs}. Similarly, since according to (4) every figure 2 situation is a figure 1 situation, what goes for figure 1 situations goes for figure 2 situations as well, and hence SS^{abs} together with (4) entails SS^{WR}. That is why so long as we assume S5, it doesn't matter which formulation we use, but as soon as we drop S5, it matters a great deal – one could be true and the other false.

Moreover, one supervenience claim, SS^{Abs}, say, could be true at one world in a given model without being true at another world of that same model. To see this, suppose we have a B-model consisting of three worlds with accessibility-relations as shown in fig. 3 by the arrows. Suppose, moreover, that there is only one subvenient property, G, and only one supervenient property, F, and that in all three worlds w, w_1 and w_2, a is the only object that has G. If a has F in w and w_1 but ~F in w_2, SS^{Abs} will be true at w, but will fail at w_1.

assumption that set of the B-properties is closed under Boolean operations; the formulation used in this note does not require this assumption. In the main text we'll ignore the precise characterization of the Strong Supervenience Condition, the better to make clear how the (E) and (4) assumptions affect the relationship among different kinds of supervenience.

fig 3



In the following, if we talk about a supervenience claim like SS^{WR} being true *tout court*, this is to be understood as an elliptical way of saying that the claim is true at the *actual* world.

Global Supervenience

Like strong supervenience, global supervenience admits of both absolute and world-relative formulations. Intuitively, global supervenience says that any two worlds which are the same in the distribution of their B-properties are the same in the distribution of their A-properties. Or for short, that any two worlds satisfy the Global Supervenience Condition[5] (GSC, for short):

GS $\quad \forall w_1 \forall w_2(GSC(w_1,w_2))$

Again, we can distinguish between whether both quantifiers range over worlds accessible from w, or whether the w_2 quantifier ranges over worlds accessible from w_1, yielding absolute and world-relative versions of global supervenience – GS^{abs} and GS^{WR}:

GS^{abs} $\quad \forall w_1{:}R(w,w_1)\forall w_2{:}R(w,w_2)(GSC(w_1,w_2))$
GS^{WR} $\quad \forall w_1{:}R(w,w_1)\forall w_2{:}R(w_1,w_2)(GSC(w_1,w_2))$

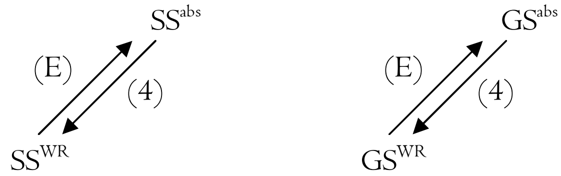
By the same reasoning as before, these bear the same logical relationship to one another. Since GS^{abs} quantifies over figure 1 situations and GS^{WR} quantifies over figure 2 situations, the fact that (E) says that every figure 1 situation is a figure 2 situation means that given (E), what goes for figure 2 situations goes for figure 1 situations, and hence that GS^{WR}, together with (E), entails GS^{abs}. Similarly, the fact that (4) says that every figure 2 situation is a figure 1 situation means that given (4), what goes for figure 1 situations goes for figure 2 situations, and hence that GS^{abs}, along with (4), entails GS^{WR}.

In this section, we have noted that both strong and global supervenience theses admit of both absolute and world-relative formulations, and that their relationship depends on some of the substantive assumptions that are encapsulated in the S5 modal logic. In particular, the absolute versions of these theses

[5] $GSC(w_1,w_2) \equiv (B\text{-indiscern}(w_1,w_2) \rightarrow A\text{-indiscern}(w_1,w_2))$, where 'B-indiscern($w_1$,$w_2$)' is an abbreviated way of saying that w_1 and w_2 have the same distribution of B-properties. There are different ways of precisifying what it takes for two worlds to have the 'same distribution' of B-properties; the same point goes for each of these, and we won't worry about such details, here.

entail their world-relative versions only under the assumption of (4), and the world-relative versions entail the absolute versions only under the assumption of (E). Our progress so far is encapsulated in the following figure:

figure 4



2.2 the relationship between strong and global supervenience

It is often argued or assumed that strong supervenience entails global supervenience, and has sometimes been claimed that global supervenience necessitates some corresponding version of strong supervenience, even though strong supervenience does not follow from global supervenience in every model.[6] In this section we briefly survey the former of these two claims, and how it is affected by relaxing our modal logic.

The idea that strong supervenience entails global supervenience is simple. Global supervenience requires that any two worlds satisfy the global supervenience condition, which says, intuitively, that they differ in their distribution of A properties only if they differ in their distribution of B properties. Strong supervenience says that any two worlds satisfy the strong supervenience condition, which says, intuitively, that any item in the first world differs from any item in the second world in some A property only if they differ in some B property as well. So the argument that strong supervenience entails global supervenience is simple: take two worlds which differ in the distribution of their A properties. Intuitively, there must be something which has some A property in one world but lacks it in the other.[7] If so, then the assumption that these two worlds satisfy the strong supervenience condition tells us that the item which has the A property in one world but not the other must differ in some B property between the two worlds. And intuitively, that suffices to make the two worlds differ in the distribution of their B properties. So on the assumption that any pair of worlds satisfies the strong supervenience condition, we can show that they also satisfy the global supervenience condition.

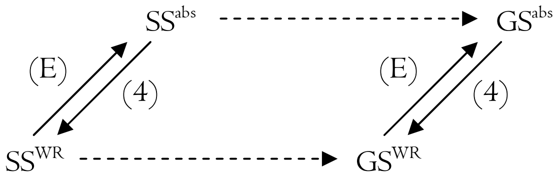
---

[6] Kim [1984] originally argued that strong supervenience did follow from global supervenience in every model; countermodels were subsequently given by Hellman [1985] and others, as discussed in Kim [1987]. In an enlightening paper, Paull and Sider [1992] argue that in interesting cases, these countermodels violate combinatorial constraints on the space of possible worlds, and hence strong supervenience may indeed follow from global supervenience in *intended* models. We won't wade into these questions, here.

[7] One possible complication with the argument may arise if A includes relations as well as one-place properties. At a minimum, this makes a more careful formulation of the argument necessary.

We don't mean to endorse this reasoning, here; just to remind readers of it. Making good on this reasoning requires making good on a precise way of understanding global supervenience's talk of the 'same distribution' of properties in each of two worlds. What we want to observe, is that insofar as the reasoning is sound, it gets us the conclusion that each version of strong supervenience entails its *corresponding* version of global supervenience, and that the reasoning is equally sound for the connection between SS^{abs} and GS^{abs}, as for the connection between SS^{WR} and GS^{WR}. This is because the soundness of the reasoning turns only on whether the strong supervenience condition entails the global supervenience condition, once the global supervenience condition is formulated more precisely. We supplement our picture with dashed arrows to indicate our caution about these entailments:

figure 5



3.1 fancy supervenience arguments from the literature

Both strong and global supervenience have been argued to entail a certain kind of *reductivism*: that if one set of properties supervenes on another, then the first set can be *reduced* in some way to the second. It is best to divide these arguments into a more logically straightforward first step, which argues that supervenience of some kind entails a certain kind of necessary equivalence, and a more philosophically loaded second step, which involves the assumption that necessary equivalences of this kind are a sufficient condition for identity.

So, for example, Jaegwon Kim [1984] has argued that if the set of A properties strongly supervenes on the set of B properties, it follows that for each A property, we can construct a property in the Boolean closure of the B properties with which it is necessarily coextensive. This is the logically straightforward first stage of his argument. Then he assumes that properties can be no more fine-grained than necessary coextensiveness, and hence that the condition that he has established is sufficient for property identity. Hence, he concludes that if the A properties strongly supervene on the B properties, each A property must be identical with some property in the Boolean closure of the B properties, and hence that every A property can be analyzed in terms of the B properties, along with Boolean operations. We'll look at this argument and how it fares under our generalized assumptions in a moment.

On the global supervenience side, Frank Jackson [1997] has argued that if the set of A properties globally supervenes on the set of B properties, then for each atomic A proposition, to the effect that some particular thing has one or another of the A properties, it is possible to construct a complex B proposition to which it is necessarily equivalent. This is the logically straightforward part of his argument. The actual dialectic of Jackson's next step is somewhat complicated, because he doesn't directly assume that necessary equivalence is sufficient for proposition identity (though he does think this). What he does instead is to claim that the same argument goes through '*mutatis mutandis*' for properties, to assume that necessary coextensiveness is sufficient for property identity, and to claim that he has established that every A property can be analyzed in terms of the B properties.[8]

It is not clear what Jackson means by saying that his argument goes through *mutatis mutandis* for properties, given that in the text, he has assumed global supervenience but not strong supervenience, and unless Jackson is making the controversial assumption that global supervenience entails strong supervenience (which certainly does not follow given the assumptions we have been making so far), his argument does *not* go through *mutatis mutandis*. Nevertheless, a more careful exponent of the argument from global supervenience to reduction might proceed more directly, and simply assume (what Jackson thinks anyway) that necessary equivalence is sufficient for proposition identity, and hence infer that every A proposition is analyzable as a complex B proposition, staking her claim to have established a reduction on this thesis about the identity of propositions, as opposed to on a claim about the identity of properties.

In the next section we will show that Kim's and Jackson's arguments work only for the absolute versions of strong and global supervenience, SS^{abs} and GS^{abs}. This is the payoff of Wedgwood's rejection of (E); by accepting world-relative but not absolute versions of both strong supervenience and global supervenience, his aim is to keep what is true about supervenience without engendering its commitment to the necessary equivalences which Kim and Jackson argue lead to reduction.

3.2 strong supervenience arguments under relaxed assumptions

Take Kim's argument first. Assume that the set of A properties absolutely strongly supervenes on the set of B properties, at world w. Our first step is to create a partition of possible individuals, by establishing a set of mutually exclusive B-maximal properties, B*. A B-maximal property is a property with the feature that no two individuals with that property can differ in any of their B properties. It is a sufficient

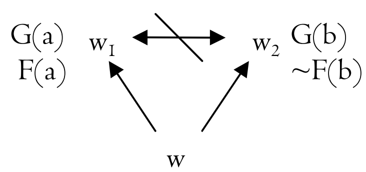
[8] 'The same line of argument can be applied *mutatis mutandis* to ethical and descriptive predicates and open sentences: for any ethical predicate there is a purely descriptive one that is necessarily co-extensive with it.' Jackson [1998, 123].

procedure to construct such a set, to take the set of all conjunctions which has as a conjunct, for each property in B, either it or its negation, yielding a set of B-maximal properties.[9]

To construct a property in the Boolean closure of the B properties that is necessarily coextensive with an arbitrary A property, F, what we do is to partition the members of B* into those which are co-possible with F and those which are not co-possible with F. The disjunction of the former, $\vee\{G\in B^*$:co-possible(G,F)$\}$, is the property that we need. Because it is a disjunction of conjunctions of properties in B and their negations, it is in the Boolean closure of B. It is necessarily sufficient for F, because each of its disjuncts is necessarily sufficient for F. To see why that is so, suppose that some property in B*, say, G, is co-possible with F. So there is a world possible relative to w, where some individual is G and F. Absolute strong supervenience then tells us that any individual *at any world possible relative to w* that is not F cannot be G, because it must differ in some B property, and by the assumption that G is in B*, two things that are G cannot differ in any B property. That is, for any individual at any world possible relative to w, if it is G, then it is F. So G is sufficient for F. Wedgwood calls such sufficiency claims *specific supervenience facts,* and we'll return to discuss them, later. Our disjunction is also necessary for F, because the properties in B* form a partition of possible individuals, and so any individual at any world that is F must have some property in B* – which is consequently co-possible with F, and hence in our disjunction. So $\vee\{G\in B^*$:co-possible(G,F)$\}$ is in the Boolean closure of B, and necessarily coextensive with F, and a similar construction goes for any other property in A.

The reasoning in this argument is impeccable, but it relies on the absolute formulation of strong supervenience at the crucial place that we've indicated with italics. The easiest way to see that the world-relative version of strong supervenience does not by itself get us Kim's conclusion is to observe that if (E) is false, then F and G may be co-possible because they are both instantiated by a at w_1, possible relative to w, but at w_2, also possible relative to w but not relative to w_1, b is G but not F:

figure 6

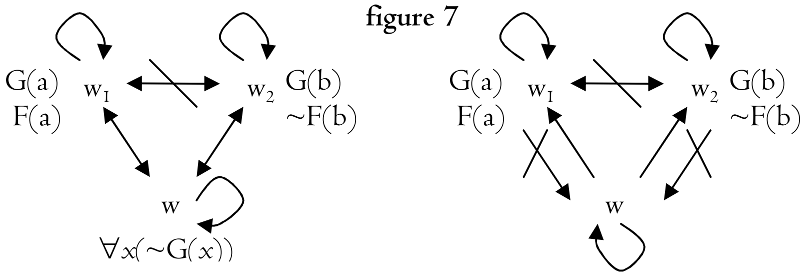


[9] The reasoning here will not rely on the members of B* being mutually exclusive, but this will come into play in one of the proofs in the appendix.

If this is so, then it may be necessary *at* w_1 that everything that is G is F, without being necessary that everything that is G is F.

In fact, the scenario just envisaged is fully compatible either with the assumption of (B), or with the assumption of (4), as figure 7 illustrates. The left-hand model in figure 7 is a (B) model in which the Kim argument fails, and the right-hand model is a (4) model in which it fails. Importantly for Wedgwood, the compatibility with (B) requires that G is uninstantiated at w, as the following figure illustrates, and we'll have occasion to comment further on, later:

figure 7



So far, the moral is this: absolute, but not world-relative, strong supervenience entails the constructability thesis for properties: that for each supervening property, there is a property in the Boolean closure of the subvening properties to which it is necessarily coextensive. So if (E) is false, either because (B) is false or because (4) is, then at least one sort of strong supervenience thesis could be true, even though there are no properties in the Boolean closure of the B properties that are necessarily coextensive with any of the A properties.

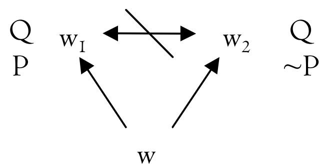
3.3 global supervenience arguments under relaxed assumptions

A similar result goes for the argument from global supervenience to the constructability thesis for propositions. The way that argument works is very similar; instead of constructing a partition of possible individuals by constructing B-maximal properties, we start by constructing a partition of possible worlds by constructing a set, B†, of B-maximal world-descriptions – propositions which describe the entire world in each and every B detail. Global supervenience says, intuitively, that two worlds can differ in the truth of some A proposition only if they differ in which member of B is true at them.

Jackson's construction procedure then asks us to look, for an arbitrary A proposition, P, at each world, w_1, at which P is true. Because the elements of B† form a partition, some member of B†, Q, is true at w_1. And so absolute global supervenience tells us that at every world possible relative to w, if Q is true, then P is true. Now take the disjunction of each such Q, for each world at which P is true: $\bigvee\{Q \in B\dagger$:co-possible(Q,P)}. This disjunctive proposition is sufficient for P, because each of its disjuncts are, and necessary for P, because it includes a disjunct for every world at which P is true.

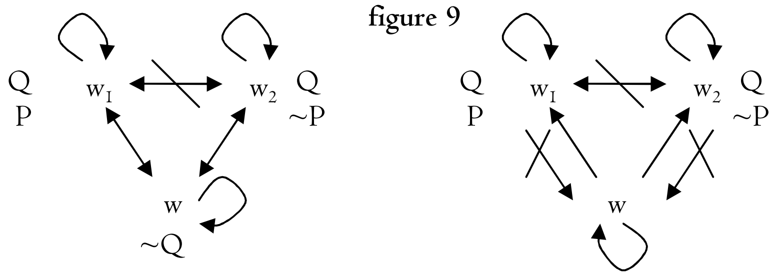
Again, the reasoning in this argument is impeccable, but it relies crucially on the assumption of the absolute version of global supervenience. The easiest way to see that the world-relative version of global supervenience does not get us to Jackson's conclusion is to observe that if (E) is false, then it is compatible with world-relative global supervenience that there may be possible worlds w_1 and w_2, such that P and Q are both true at w_1 and Q and ~P are true at w_2.

figure 8



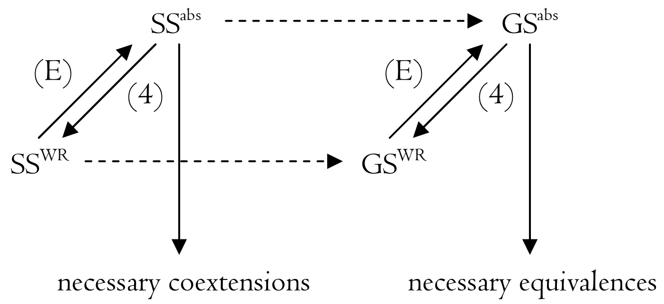
If this is so, then Q⊃P may be necessary *at w_1*, without being necessary. In fact, as before, this scenario is also compatible with either (B) or (4), though the compatibility with (B) requires that Q is false at w:

figure 9



So absolute, but not world-relative, global supervenience allows for the constructability thesis for propositions: that for each supervening proposition, there is a proposition in the Boolean closure of the subvening propositions with which it is necessarily equivalent. We can summarize our progress with figure 10:

figure 10



4.1 specific supervenience facts

So what, then, follows from the world-relative versions of strong supervenience and global supervenience? In the case of world-relative strong supervenience, we can derive, along with the assumption of (T), what Wedgwood calls 'specific supervenience facts'. A specific supervenience fact is a fact of the form: $\forall w{:}R(@,w)(\forall x(Gx \supset Fx))$, where F is a predicate for some A property, G is a predicate for some property in the Boolean closure of the B properties, and @ is the actual world. Equivalently, without explicit quantification over worlds, a specific supervenience fact has the form, $\Box\forall x(Gx \supset Fx)$. Such a fact specifies a modally *sufficient* condition for an A-property in terms of B-properties, without being a *necessary* condition for it.

To derive a specific supervenience fact for some supervening property F, find some member G of B*, our class of mutually exclusive B-maximal properties, such that F and G are *actually* coinstantiated, i.e. we have F(a)&G(a) for some individual a in the domain of the actual world, @. (T), recall, implies that R(@,@), so from (T) and SS^{WR}, we can derive $\forall w{:}R(@,w)(SSC(@,w)$ by universal instantiation. Now to show that $\Box\forall x(Gx \supset Fx)$, we'll let w be an arbitrary world possible with respect to @, and let x be an arbitrary member of w's domain, and show that $G(x) \supset F(x)$. In order to show that the material conditional holds at w, we assume that the antecedent holds at w and show that the consequent holds at w. By the B-maximality of G, our assumption that G(a) at @ and that $G(x)$ at w, it follows that a and x are B-indiscernable. (If they weren't, then they would differ in some B-property, and hence G wouldn't be B-maximal.) So it follows from SSC(@,w) (which we derive from $\forall w{:}R(@,w)(SSC(@,w)$ by universal generalization) and the rule of detachment that a and x are also A-indiscernable. But since we assumed that F(a) at @, it then follows that $F(x)$ at w. Hence we have shown $G(x) \supset F(x)$ at w and since w and x were arbitrary, we get $\Box\forall x(Gx \supset Fx)$.

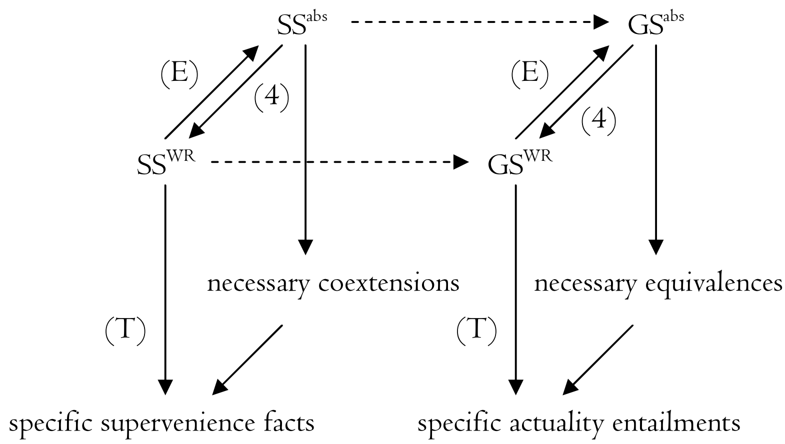
As we noted one paragraph back, this is a one-way entailment. World-relative strong supervenience allows us to construct one such one-way entailment for every B-maximal property with which F is actually coinstantiated. Each of these one-way entailments is a *specific supervenience fact*. We can also take the disjunction of all of the antecedents of these specific supervenience facts, $\vee\{G\in B^*$:actually co-instantiated(G,F)$\}$. This property, too, is sufficient for F. But we cannot get a *two*-way entailment, because nothing forces this disjunction to be *necessary* for F. Any B-maximal property that is not actually instantiated will not be actually co-instantiated with either F or ~F, and this argument works only for B-maximal properties that are actually co-instantiated with F. This should be no surprise; we already saw in section 3.2 that an uninstantiated B-maximal property may be both co-possible with F and co-possible with ~F. So this disjunction gives us a property that is necessarily coextensive with F just in case every B-maximal property is actually instantiated. Hence the constructability thesis for properties holds at all and only worlds where every B-maximal property is instantiated.

In keeping with the tight correspondence between how things work for strong and global supervenience, world-relative global supervenience allows for an analogue of specific supervenience facts, something that we call *specific actuality entailments*. To construct such an entailment, first locate the B-maximal world description, Q, which describes the actual world in every detail. Next choose some A-proposition, P, which is actually true. Since we have R(@,@) (which follows from (T)) GS^{WR} can be simplified to $\forall w$: R(@,w)(GSC(@,w)), which, given the truth of Q & P at @, implies that at every world possible relative to the actual world at which Q is true P is true, as well. But that just means that $Q\supset P$ is necessary at the actual world.

Again, this is a one-way entailment. We can construct similar one-way entailments for each other A-proposition that is actually true, but we can't use world-relative global supervenience to construct any other one-way entailments whose consequents are P, because we don't have any guarantee that what is necessary at other possible worlds will actually be necessary.

Specific actuality entailments are the least controversial sort of evidence for some kind of supervenience of the normative on the non-normative. For example, we take it that you should legitimately be extremely confident that it is impossible for the world to be exactly as it is in every non-normative respect, but different normatively in that the fact that your mother is your mother is a reason for you to torture her. You may not know exactly *how* the facts about the non-normative world guarantee that things could not be different in this normative way without some normative difference, but still legitimately be absolutely certain that they could not – more confident than you are in any particular stronger supervenience thesis. So this is an example of a specific actuality entailment. It makes sense that a thesis like this one should be the most obvious thesis involved with supervenience claims, because as we've seen so far, it is the weakest sort of view that one could take:

figure 11



4.2 explanatory arguments based on supervenience

Kim and Jackson's arguments aspire to show that supervenience *entails* reduction. As we noted, each of these arguments divides into two components. The first is an uncontroversial *constructability* thesis, to the effect that A properties or propositions have necessarily coextensive or necessarily equivalent counterparts from the Boolean closure of the B properties or propositions. The second is a controversial assumption that necessary coextensiveness is a sufficient condition for property identity, or that necessary equivalence is a sufficient condition for proposition identity. So these direct arguments from supervenience to reduction rely on the very strong assumptions, respectively, that properties and propositions are very coarsely individuated. A more cautious set of arguments from supervenience to reduction does not rely on these assumptions.

The key idea of these more cautious arguments is that even if supervenience theses do not *entail* reduction, they still leave something to be *explained*. The idea is that since supervenience theses tell us not just that certain things *don't happen* to occur, but that it is *impossible* for them to occur, they aren't the kind of thing that can just happen by coincidence. If the A properties are genuinely distinct from and irreducible to the B properties, then why should it be *impossible* for something to have such-and-such a B-maximal property and not have so-and-so an A property?

This reasoning is reinforced by consideration of natural recombination principles. Toy blocks can be square or round, and red or blue. Even if there are not any red square blocks, there could have been – all that it would take, would be for there to be a block that 'recombines' the square feature that some blue blocks have with the red feature that some round blocks have. Since these are distinct properties that blocks can have, they can be recombined in any combinatorially possible way. Every combination is possible, unless something prevents it. So if A properties are distinct from B-properties, what prevents them from being recombined in any possible way? What makes some recombinations not just non-actual, but *impossible*? Supervenience, it seems, leaves something to be explained.

Reductive hypotheses can explain supervenience. If A properties are reducible to B properties, then we can explain why they supervene on the B properties in the same way as we would explain why the property of being a red square supervenes on the properties of being square and of being red. Since supervenience hypotheses stand in need of explanation, and reductive hypotheses seem particularly well-suited to explain supervenience, there is therefore an explanatory argument from supervenience to reduction. Unlike the Kim-Jackson style arguments, it is only a defeasible argument. But unlike them, it does not rely on controversial assumptions to the effect that properties or propositions are very coarse-grained.[10]

Moreover, explanatory arguments do not require the absolute versions of strong or global supervenience in order to get started. In fact, they don't even require the full-blooded assumption of *any* supervenience thesis to get started. Any necessary connection between A properties and B properties will do, including specific supervenience facts or specific actuality entailments. Even these weaker theses leave something to be explained – something that reductive hypotheses can offer an explanation of.

4.3 where we are

So far we've seen that there are two main kinds of arguments connecting supervenience to reducibility. The first kind, due to Kim and Jackson, seeks to show that supervenience *entails* reducibility, and works by first establishing a constructability result to the effect that every A property (proposition) has a necessarily coextensive (equivalent) counterpart from the Boolean closure of the B properties (propositions). The second, philosophical, step of such arguments, proceeds by assuming that necessary coextensiveness (equivalence) is a sufficient condition for property (proposition) identity. The advantage of this kind of

[10] See, for example, Schroeder [2005].

argument is that if it works, its conclusion is strong: that supervenience entails reducibility. Its weakness is the strong philosophical premise that it requires: that properties (propositions) are so coarse-grained.

The other kind of argument connecting supervenience to reduction is the explanatory argument. According to the explanatory argument, supervenience need not entail reducibility, but it does leave something important to be explained, and reductive hypotheses are the right sort of thing to fill this explanatory gap. The advantage of this kind of argument is that it does not rely on the assumption that properties or propositions are so coarse-grained. Its disadvantage is that it is less conclusive, without some exhaustive classification of the possible strategies for explaining the modal facts that are involved with supervenience.

The good news for nonreductivists who would like to accept supervenience theses is that when we relax the restrictive assumptions of the S5 modal logic, we also strengthen what is required in order for Kim and Jackson's constructability arguments to work. As we saw, those arguments work only for the absolute versions of strong and global supervenience, and not for the world-relative versions, except in S5, in which the two versions are equivalent. This means that relaxing assumption (E) of the S5 modal logic is necessary but not sufficient in order to escape these constructability arguments.[11] If you want to accept

---

[11] To be more precise, relaxing assumption (E) is necessary but not sufficient for evading the first, constructability, step of the Kim-Jackson style arguments. It is not necessary for evading those arguments *tout court*, because it is possible to evade them at the second, philosophical step, at which they assume that necessary coextensive (equivalent) properties (propositions) are identical.

Billy Dunaway has pointed out to us (in discussion) that the claims (i) through (iv) are jointly unsatisfiable

(i) ~(E)
(ii) (B)
(iii) necessary coextensiveness is necessary and sufficient for property identity
(iv) if properties P and Q are identical (at a world), they are necessarily identical at that world.

The proof (by reduction) is very straightforward: We start by constructing a model in which (E) fails (the square brackets indicate which propositions are true at the worlds and the arrows the indicate accessibility-relation):

$$\mathbf{w_1}\,[\,p\,] \leftarrow \mathbf{w_2}\,[\Diamond p\,,\sim p\,] \rightarrow \mathbf{w_3}[\sim\Diamond p]$$

Here, $\Diamond p$ is true at w_2 because w_2 'sees' w_1, but $\Box\Diamond p$ is not, because w_2 also sees w_3; hence we have the antecedent of (E) but not its consequent. Assumption (ii) tells us that accessibility-relations must hold symmetrically across the model:

$$\mathbf{w_1}\,[\,p\,] \rightleftarrows \mathbf{w_2}\,[\Diamond p\,,\sim p\,] \rightleftarrows \mathbf{w_3}[\sim\Diamond p]$$

Assuming the sufficiency direction of (iii) at w_3, the property $\lambda x.\ p$ is identical with the property $\lambda x.\ x \neq x$ at w_3. Moreover, this identity is *necessary* (by (iv)). Since w_2 is accessible from w_3, the identity must hold in w_2, too. So $\lambda x.\ p$ and $\lambda x.\ x \neq x$ are identical at w_2. By (iv) they are *necessarily* identical. Now, w_1 is accessible from w_2, so the identity must hold in w_1, too. But since p is true at w_1, the extension of p at w_1 is non-empty, assuming that w_1 has a non-empty domain (which we now assume by stipulation). Hence by (iii) again, the two properties cannot be identical at w_1. Contradiction!

We think that this proof is *prima facie* good evidence that one cannot reject S5 (while accepting B) without giving up one of (iii) and (iv). There are some possible complications with the proof (it assumes unrestricted comprehension for properties within supervenience theses but avoid these arguments, you need to be careful to accept only the world-relative supervenience theses, and to deny the absolute versions.

Consequently, even once we relax assumption (E), much more work remains to be done, for the nonreductivist who would like to accept supervenience but escape the Kim-Jackson constructability arguments. Such a theorist must defend her choice to accept only the world-relative supervenience theses, and deny their absolute versions. For what we've seen is that even if we relax our modal logic, everything will still hang on *which* formulation of supervenience is correct. In the remainder of this paper, we will look at Ralph Wedgwood's proposal for how to exploit the relaxed assumptions involved with denying assumption (4), in order to answer the explanatory argument. Like the Kim-Jackson style of argument, we will see that there is good news for nonreductivists who want to accept supervenience. But also like the Kim-Jackson argument, we will see that everything hangs on *which* formulation of supervenience is correct.

Consequently we will close the paper in part 6 by evaluating which supervenience theses are more compelling – the absolute or the world-relative versions. Peculiarly, though this distinction does so much work in his arguments, Wedgwood never considers this question explicitly, and in fact never explicitly distinguishes between the two formulations of strong supervenience (though he does distinguish between the two formulations of global supervenience). We will argue, however, both that the world-relative theses are hard to formulate clearly and unambiguously in natural language, making it implausible that they are what we find pretheoretically compelling, and that even once we reject S5, denying the absolute versions of the supervenience theses is still particularly unintuitive.

5.1 explaining supervenience given blackburn's point

The explanatory argument challenges the nonreductivist to explain the necessary facts that are involved with the claim of supervenience. Ralph Wedgwood's attempt to do so consists in three basic parts, which we'll explain in the next three sections, respectively. The first part of the explanation is to explain *some* of the necessities involved with supervenience by appeal to claims about essence. It is an important fact about supervenience, however, emphasized in Simon Blackburn's arguments about supervenience, that some

a second-order framework) but we think they may not be decisive. It seems to us that denying (iv) is mind-boggling, so we conclude that (iii) has to go (notice that in Dunaway's proof both directions of (iii) – the sufficiency and the necessity - are appealed to).

This basic consequence should really be no surprise; in any ~(4)-model, properties can be necessarily coextensive without being necessarily necessarily coextensive, and necessarily necessarily coextensive without being necessarily necessarily necessarily coextensive, and so on. So intuitively, it is not plausible that necessary coextensiveness is sufficient for property identity after all; a better candidate is *hyper*necessary coextensiveness, where 'hypernecessarily *p*' is defined as the limit of '$\Box p$', '$\Box\Box p$', '$\Box\Box\Box p$'.... But then the constructability argument would need to establish properties that are hypernecessarily equivalent, not just necessarily equivalent.

necessities involved with supervenience cannot be explained in this way, and this leads to the second part of Wedgwood's account, which is to explain the remaining necessities emphasized in Blackburn's point by appeal to contingent explananda – a resourceful move made available by giving up on (4). Finally, the third part of Wedgwood's account confronts what he takes to be a remaining problem arising from global supervenience. Our goal in these three sections will be not only to explain the nature of Wedgwood's account, but to illustrate how it hangs on the choice between SS^{abs} and SS^{WR}, as well as on the choice between GS^{abs} and GS^{WR}, in addition to hanging on the denial of (4).

The first part of Wedgwood's account of supervenience is to propose that the normative supervenes on the non-normative because it is in the *nature* of normative properties to satisfy strong supervenience. Just as it is in the essence of squares to have four sides, it is in the essence of the normative to supervene. This view adopts the way of thinking, familiar from thinkers like Kit Fine [1994], that essence is prior to, and explanatory of, modality, and things are possible because they are compatible with essence (though this becomes slightly more complicated at the next step of Wedgwood's account).

Wedgwood is not the first to suppose that it is part of the nature of the normative to supervene, though he is the first to articulate this thought so clearly in the framework of thinking about essence. It faces an obvious problem, of which he is keenly aware, and it is that the normative cannot supervene, without supervening in some *particular way*. As we observed in section 4.1, given (either) version of strong supervenience, we can derive *specific supervenience facts*, which tell us *how* the normative is supervening. For example, nothing about the bare fact of strong supervenience tells us whether the property of being good is compatible or not with, say, the property of causing much more deep unhappiness than it prevents. But if strong supervenience is true, then these properties *may* be incompatible – indeed, there are guaranteed to be some properties that are incompatible, if strong supervenience is true.

Simon Blackburn emphasized this important fact about strong supervenience in three different versions of an argument against moral realism that he presented in his [1973], [1984], and [1985]. We won't rehearse all of the parts of that problematic argument here, but a central feature of the argument was Blackburn's point that *bare* supervenience is impossible – in order for there to be strong supervenience, there must be some specific supervenience facts or other, even though no particular specific supervenience facts are required for strong supervenience. (Blackburn himself tried to argue that there are no specific supervenience facts of the relevant kind, and used this to argue that strong supervenience should be replaced with *weak* supervenience, which he in turn argued that only non-realists are in a position to do.)

The relevant point for Wedgwood's argument, however, is that it is not enough, to say that it is in the essence of normative properties to supervene, in order to explain all of the metaphysical impossibilities associated with normative properties, if strong supervenience is true. You could, of course, say that it is in the nature of normative properties to be coinstantiated with particular non-normative properties, but that just looks like the reductivist's explanation of supervenience. So Wedgwood needs a different kind of explanation of specific supervenience facts, and that brings us to the second part of his account.

5.2 contingent explananda for modal explanans

Wedgwood's explanation of specific supervenience facts is simple, and takes advantage of giving up (4). His thought is that if it is in the nature of A properties to strongly supervene on B properties, and some particular A property, F, happens to be actually coinstantiated with some particular B-maximal property, G, then it will have to be impossible for anything to be G but not F – otherwise the A properties wouldn't supervene on the B properties after all. So the idea is that, having first explained strong supervenience by appeal to a claim about essence, to use strong supervenience, together with *contingent* facts, to explain specific supervenience facts. (Even though it is necessary that something is, if G, F, it is only contingent that anything is both F and G.)

It is easier to see how Wedgwood's explanation works, to think about it in terms of a picture of the space, not only of possibility, but of *hyper*possibility. Let the space of hyperpossibility include not only the possible worlds, but the possibly possible worlds, and the possibly possibly possible worlds, and so on – out to the limit. This space is represented by the model structure for a (B) model of modal logic. Rather than thinking about the space of possibility as constrained only by essence, Wedgwood thinks of the space of *hyper*possibility as constrained only by essence. Anything that is compatible with essence is hyperpossible – including any way of recombining A properties with B properties, because since the essence of the A properties only constrains them to strongly supervene, from Blackburn's point it follows that they might still supervene in any of a number of ways. Consequently, each of these ways of supervening is hyperpossible.

But facts about essence don't just constrain which *worlds* there are in the space of hyperpossibility; they also place primitive constraints on which worlds are possible relative to which others. For example, even though it is hyperpossible for F to be coinstantiated with the B-maximal property G, and hyperpossible for ~F to be coinstantiated with G, neither of these two worlds is possible relative to the other. At the closest, they are each possible relative to a third world where G is uninstantiated, as in figure 6. Consequently, what is possible in this picture depends on *where we are* in the space of hyperpossibility.

So this is what Wedgwood's appeal to contingent facts does – it isolates where we are in the space of hyperpossibility, in order to isolate which of the many hyperpossibilities are really, genuinely, possible.

The nonreductivist at whom the explanatory argument is directed accepts supervenience theses and consequently metaphysical impossibilities, without having any explanation for why those things are genuinely impossible. Insofar as they have no explanation, they appear to be brute. But the problem is that metaphysical impossibilities don't seem like the kind of thing that *can* be brute. In Wedgwood's picture, however, the metaphysical impossibilities posited by specific supervenience facts *do* have an explanation – they are explained by contingent facts. And then the explanation stops. But that seems like a much more acceptable place for the explanation to stop – after all, contingent facts *do* seem like the kind of thing that can simply be brute. Ultimately, things simply have to be one way or the other.

So far, so good – the second step of Wedgwood's answer to the explanatory argument exploits the assumption that (4) is false, by working with a model in which not everything that is hyperpossible is genuinely possible. Like the earlier ways of avoiding the constructability results from Kim and Jackson, therefore, it works by exploiting a relaxation in the assumptions built into the S5 modal logic. But that's not all it needs; as we'll see in the next section, Wedgwood's answer to the explanatory argument also shares with the attempts to avoid the constructability results, the feature that it trades on *which* versions of strong and global supervenience we assume. This is best illustrated by introducing the third part of Wedgwood's account.

5.3 the problem of global supervenience

The third part of Wedgwood's account is needed to address the problem that if GS^{abs} is true, then there will be further necessary truths which still remain unexplained. The problem goes like this: let Q be a B-maximal world description, as in section 3.3. That is, Q is a proposition which describes the world in every B detail, so that necessarily, any two worlds in both of which Q is true are B-indistinguishable. And let P be any A proposition whatsoever. Since GS^{abs} says that worlds which are B-indistinguishable must be A-indistinguishable as well, it follows as a special case that if Q is true at two worlds, then P must be either true at both or false at both (assuming bivalence). That is, GS^{abs} entails that one of the following two theses will be true:

1 $\Box(Q \supset P)$
2 $\Box(Q \supset \sim P)$

Moreover, this would be a problem for Wedgwood because his explanation of specific supervenience facts does not generalize to explain either of these facts. Wedgwood's explanation of specific supervenience facts, in the second part of his account, can only explain necessities that follow from supervenience facts, together with contingent facts. But even holding fixed the truth of GS^{abs} – assuming that it might be true in virtue of the essence of normative properties – no contingent fact suffices to explain *which* of 1 and 2 is true, unless Q happens to be true at the actual world. Unfortunately, however, if Q&P is merely *possible*, that suffices, together with GS^{abs}, to mean that 1 has to be true. Similarly, if Q&~P is merely possible, that suffices, together with GS^{abs}, to mean that 2 has to be true. So to explain whichever of 1 or 2 is true, Wedgwood would need to explain either why Q&P is impossible, or why Q&~P is impossible. But nothing about supervenience or about contingent facts suffices to say which of these is impossible – just that one of them *has* to be impossible. This is the revenge of the problem raised by Blackburn's point – to be true, supervenience theses require that one or another of some stronger, more specific necessities also be true, but they don't specify which one needs to be true.

The third part of Wedgwood's account is therefore to reject GS^{abs}. Distinguishing GS^{abs} from GS^{WR}, he endorses only the latter, and rejects the former. In contrast to GS^{abs}, GS^{WR} does *not* entail in general that either 1 or 2 is true; the only choice of 'Q' for which GS^{WR} does entail that either 1 or 2 is true is where Q is true. This is the specific actuality entailment that we constructed in section 4.1. But this *particular* necessity *can* be explained by Wedgwood's method from the second part of his account: it is explained on the basis of the contingent fact that Q&P (or that Q&~P – whichever the case may be).

Given that Wedgwood endorses the entailment of global supervenience by strong supervenience, this means that he is committed to denying SS^{abs} as well. Wedgwood is unfortunately less than perspicuous on this point, and neglects to distinguish between the absolute and world-relative versions of strong supervenience, as he does for the two versions of global supervenience. The only version of strong supervenience that he even discusses is SS^{WR}. But as we saw in section 2.2, if SS^{WR} entails GS^{WR}, then SS^{abs} entails GS^{abs} as well.

So Wedgwood's response to the explanatory argument, just like the responses to the constructability arguments considered in section 4.3, requires not only weakening our modal logic (the response to the constructability argument required giving up (E) and Wedgwood's account requires giving up (4)), but also requires the assumption that the absolute formulations of both strong and global supervenience are false. So if he wants to endorse supervenience at all (as he does), he needs to endorse only the world-relative versions of each. In the final part of the paper, we'll now evaluate whether it is plausible to reject the absolute formulations of strong supervenience, even in a relaxed modal logic. We'll be arguing that this is still a highly counterintuitive way to go – *even if* (4) is false.

6.1 absolute supervenience theses are independently intuitively compelling

As we have just seen, the viability of Wedgwood's account requires a specific package of views on strong and global supervenience. It requires him to affirm the relative versions of strong and global supervenience while denying their absolute versions. In this section we will argue that denying the absolute version of strong supervenience is very implausible. We can test for these intuitions by formulating the different versions (and their negations) in natural language and inquiring which of the formulations 'sound' intuitively true and also, on a comparative scale, whether some versions seem more plausible than their counterparts.

We will be arguing not only that it is unintuitive to deny the absolute version of SS *per se*, but that it is especially odd to deny it while at the same time affirming its world-relative counterpart. This can be made particularly vivid by considering some of the cases in which supervenience is often invoked, such as the supervenience of the moral on the non-moral or natural. In fact, it is hard to see why any reason[12] to deny the absolute version would not also give us (similar) reasons to deny the relative versions. What is it that sets the two theses apart? We believe that similar points will go for both strong and global supervenience, but here we will confine our remarks to strong supervenience.

Earlier in this paper we gave formulations of the two versions involving quantification over possible worlds. Doing so made it easier to see the logical relationships among supervenience theses. But now we will translate these formulations of supervenience into equivalent natural language sentences that involve modal adverbs like *necessarily* and modal complementizers like *it could have been (the case) that.*[13] Since these expressions are more natural ways of talking than expressions like 'in every possible world', the idea is that framing supervenience using them will make it easier to think about the intuitive plausibility of supervenience theses and their negations.

Another standard way of formulating SS^{WR} and SS^{Abs} which can be found in the literature[14] better lends itself to natural-language formulations. It looks like this:

---

[12] Any reason *other* than the reason that your theory requires you to hold one version while denying the other

[13] We choose this more complicated way of expressing alethic modality because we think that 'it is possible that p' mostly expresses epistemic possibility.

[14] See, for example, Kim [1984, 65], where Kim formulates world-relative strong supervenience in this way. He does not call it 'world-relative', of course, because he does not distinguish it from absolute strong supervenience.

(i) $\Box\forall x\forall F\in A[(F(x)\rightarrow\exists G\in B^*(G(x)\&\Box\forall y(G(y)\rightarrow F(y)))]$

(ii) $\forall F\in A\forall G\in B^*[\Diamond\exists x(F(x)\&G(x))\rightarrow\Box\forall y(G(y)\rightarrow F(y))]$[15]

For the proof of the equivalence of SS^{Abs} and (ii), see Appendix A.1. The relationship between SS^{WR} and (i) is slightly more complicated, but it turns out that SS^{WR} holds at all worlds in a model just in case (i) holds at all worlds in that model; we prove this in Appendix A.2. Since Wedgwood's explanation of supervenience in terms of essence requires it to hold at all worlds in the model (that is, in all hyperpossible worlds), we take it that SS^{WR} and (i) are interchangeable for his purposes.

The advantage of working with (i) and (ii), rather than only with SS^{Abs} and SS^{WR}, is that they make it easier for us to look at something closer to natural-language versions of supervenience, starting with (i) (world-relative strong supervenience):

(I) Necessarily, if anything has one of the A properties, then there is a B-maximal property which it has, and which necessarily implies having that A property.

It might seem harder to translate (ii) into English. In fact, most formulations of absolute strong supervenience that can be found in the literature seem to be in a language that allows quantification over possible worlds – similar to our earlier formulation (SS^{abs}).[16]

Here is how one might try to render (ii) in something approximating ordinary English:

(II) For any A property and any B-maximal property, if they could have been co-instantiated, then necessarily, anything that has that B-maximal property has that A property.

Despite doing our best to formulate both SS^{WR} and SS^{Abs} in comprehensible, natural-language formulations, we believe that (II), which states the absolute version of strong supervenience, is both much easier to understand[17], and deeply independently intuitively compelling in its own right, in cases for which supervenience claims are often made, such as the supervenience of the moral on the non-moral.

---

[15] Note that in (ii) the property-quantifiers have to take scope over the whole conditional - which means that they don't occur within the scope of any modal operator. While (i) requires the assumption that all A-and B-maximal properties exist in all possible worlds (regardless of whether they are instantiated in those worlds), (ii) requires only the (weaker) assumption that all A- and B-maximal properties exist at the *actual* world. This creates a complication we'll need to gloss over, here.

[16] For example, see Paull and Sider [1992, 834] and McLaughlin [1997, 210].

[17] The difficulty in being sure that we understand (I) properly when we consider it in ordinary language, lies in being sure that the second 'necessarily' is clearly understood as having scope inside the first. Any reading on which this is not clear may simply be confusing world-relative strong supervenience with absolute strong supervenience.

What it says, in the moral case, is that for any moral property and any maximal non-moral property – that is, any complete way that something could be in every non-moral respect – if it is possible for something to have had both properties, then necessarily, anything which is that same way in every non-moral respect would also have to have that moral property. For example, if it is possible for an action to be wrong under some maximally determinate set of circumstances, then necessarily any action in the very same circumstances would be wrong. This claim is highly intuitive in its own right.

6.2 the negations of absolute supervenience theses are independently deeply implausible

Moreover, we also believe that the intuitive compellingness of (II) for cases like the moral/non-moral case is even clearer, when we turn to look at its negation. Negating (II) and simplifying the resulting formulas by pushing the negation through (until all the negation-signs of the formula are contained in some atomic subformula) gives us:

> Neg-(II) There is a B-maximal property which, for some A property F, could have been coinstantiated with F and could also have been instantiated without F.

We think that Neg-(II) is both particularly easy to understand and *very* implausible, for ordinary cases in which supervenience theses are maintained, such as that of the moral – in particular, of wrongness – on the non-moral. Observe:

> **Wrong** There is some complete characterization of an action in absolutely every non-normative respect, such that it is both possible for an action to satisfy that characterization and be wrong, and possible for an action to satisfy that characterization and not be wrong, without any other non-normative difference between the two situations.

This is extremely unintuitive. We think that such a scenario is very odd – it seems to undermine the core intuition in which our commitment to the supervenience of the normative on the natural is grounded.

The moral so far, is this: even though there is such an important logical difference between world-relative and absolute strong supervenience theses, the absolute version is extremely compelling in cases like this one. But of course, what we've observed in the earlier part of the paper is that the payoffs for Wedgwood's resistance to the S5 modal logic require him to deny the absolute version of strong supervenience for the moral on the natural. So he is committed to the unintuitive result illustrated by Wrong.

6.3 two possible responses?

Presumably, one natural response for Wedgwood, is to claim that our reasoning rests on a subtle confusion. Perhaps he might argue that when we are finding SS^{Abs} intuitively compelling, that is because we are really confusing it with SS^{WR}, which is the real truth in the neighborhood. This confusion, if we were really making it, could explain our intuitive judgment. But why would anyone confuse SS^{Abs} and SS^{WR}? One reason for such a possible confusion may be that we intuitively take the necessity-operators to satisfy (E) and both versions of SS are of course equivalent – as we have shown – if (E) holds. Call this the confusion hypothesis.[18]

We agree that this is a good strategy for Wedgwood to try, but we think that it is particularly implausible, and for two reasons. First, we believe that Neg-(II), which is equivalent to the negation of SS^{Abs}, is one of the easiest-to-understand ways available of formulating any supervenience thesis or its negation. It is very implausible to think that when we consider such an easy-to-understand sentence, what is really happening, is that we are mixing it up with the negation of SS^{WR}, and not least because the negation of SS^{WR} is rather hard to understand, in its own right. Here is our best pass at how to express the negation of (I) in a natural way:

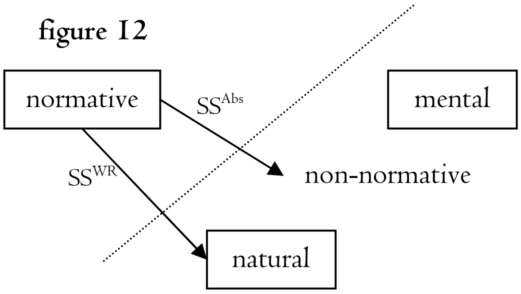
> Neg-(I) There could have been something with an A property, but which would have possibly not had that A property, even if it had had the same B-maximal property.

It is extremely implausible that our grasp of such an easy-to-understand sentence as Neg-(II) is mediated by confusing it with the difficult-to-grasp content of Neg-(I), as the Confusion Hypothesis would require. Moreover, our own plausibility-judgments persist when we stipulate that (E) fails – in fact, we don't even have to *stipulate* that (E) fails because we both *do* reject (E) independently of the intuitions discussed in this section and so the fact that our judgement is robust with respect to the rejection of (E) is evidence that the Confusion Hypothesis is false.

Alternatively, instead of appealing to the Confusion Hypothesis, Wedgwood could fall back on the important distinction between the thesis that the moral supervenes on the natural and the thesis that the normative supervenes on the *non-normative*. As he has reminded us in conversation, Wedgwood is not committed to denying the absolute strong supervenience of the normative on the *non-normative* – because he is inclined to hold that the class of non-normative properties includes mental properties. His picture therefore looks like this:

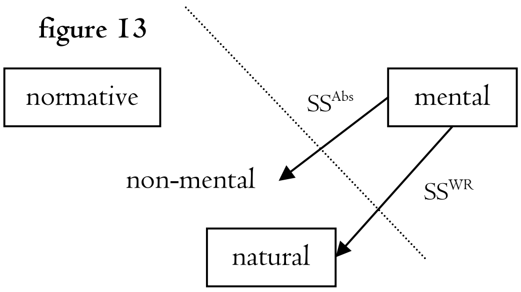
[18] Thanks to an anonymous referee for reminding us how the assumption of (E) may be thought to mediate such confusion.

figure 12



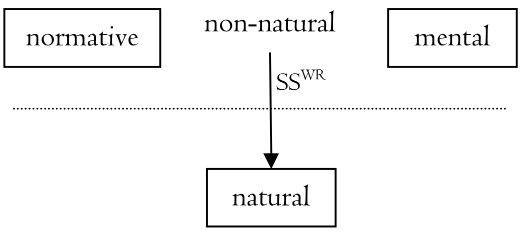
As illustrated in the figure, Wedgwood accepts that the set of normative properties absolutely strongly supervenes on the set of non-normative properties – provided that this set is understood to include intentional mental properties as well as strictly natural properties. He holds an analogous relationship to hold for the supervenience of the mental properties:

figure 13



Despite the acceptance of these absolute strong supervenience theses, however, this picture is still compatible with Wedgwood's nonreductive ambitions, because he is a nonreductivist about both the normative and the mental:

figure 14



So instead of hypothesizing that we are confusing SS^{Abs} with SS^{WR}, Wedgwood might claim that we are confusing the absolute strong supervenience of the normative on the non-normative (which is both plausible and true) with the absolute strong supervenience of the normative on the natural (which he claims is false, and only gains plausibility by confusion with the former).

It would be interesting enough to find that this sophisticated maneuver is required in order make good on taking advantage of relaxing S5 in order to accommodate supervenience without reduction. If this is right, then Wedgwood's views really will stand or fall as quite a large package, and it will be difficult to get the benefits of his idea of relaxing S5 without going whole hog on many of his other controversial ideas. But even so, though this respone complicates matters, we find it ultimately unpersuasive. It is true that as we have formulated **Wrong**, above, it is formulated in terms of the supervenience of wrongness on the non-normative. But we could just as well formulate it in terms of the natural, which we continue to find just as unintuitive:

> **Wrong*** There is some complete characterization of an action in absolutely every natural respect, such that it is both possible for an action to satisfy that characterization and be wrong, and possible for an action to satisfy that characterization and not be wrong, without any other natural difference between the two situations.

Against this, one might claim that **Wrong*** is only unintuitive if property-dualism is false. To see the force of this objection suppose that phenomenal properties are non-natural properties. Now, consider a scenario in which Mary viciously hits John. If whether this action is wrong may depend on how much pain it causes, and if pain is a phenomenal property, not a natural property, then Wrong* could be true.[19]

Two comments are in order: First, Wedgwood seems to be committed to denying absolute strong supervenience *even if phenomenal properties are included in the supervenience base* because not only does he think that the normative cannot be reduced to the natural, he also thinks that the normative cannot be reduced to the natural *cum* phenomenal. The normative is not reducible to any set of properties unless that set includes *intentional* properties and so – given what we have argued in section 5.3. – it seems to us that Wedgwood is committed to SS^{Abs} being false for any supervenience-base that does not include intentional properties.[20]

Second, even though we wish to remain somewhat cautious about this point, we are both very sympathetic to a kind of naturalism which entails that phenomenal properties are natural properties, so we happen to think that the kind of property-dualism that the objection presupposes is most probably false (in fact, necessarily false). Consequently, we think it is anything but costless to show that the uninituitiveness of **Wrong*** can be explained away or resisted.

---

[19] Thanks to Tristram McPherson for pushing us on the relative implausibility of **Wrong** and **Wrong*** along these lines.

[20] For this reasoning to go through we have to assume that intentional properties are not reducible to or constructible out of phenomenal properties.

In conclusion, despite the allure of the idea that relaxed assumptions about modal logic can resuscitate nonreductive theories and keep at bay a variety of important arguments based on supervenience, the moral is: not so fast. All of these potential benefits also turn on the idea that absolute versions of supervenience theses are false and only their world-relative versions are true. But if what we've argued here is on the right track, the absolute versions of plausible supervenience theses are intuitively compelling in their own right.[21]

Appendix A.I proof that ii. and SS^{Abs} are equivalent

We begin by noting that SS^{Abs} can be broken up into the following two parts (i.e., that it is equivalent to their conjunction):

SS^{Abs}_1 $\forall w_{:R(@,w)}\forall v_{:R(@,v)}\forall x_{\in D(w)}\forall y_{\in D(v)}[\forall G_{\in B}(Gx\equiv Gy)\supset\forall F_{\in A}(Fx\supset Fy)]$

SS^{Abs}_2 $\forall w_{:R(@,w)}\forall v_{:R(@,v)}\forall x_{\in D(w)}\forall y_{\in D(v)}[\forall G_{\in B}(Gx\equiv Gy)\supset\forall F_{\in A}(Fy\supset Fx)]$

Observe that SS^{Abs}_1 and SS^{Abs}_2 are notational variants of one another (swapping x for y and w for v, and switching the order of the universal quanfiers and the biconditional). Hence, since SS^{Abs} is equivalent to their conjunction, it is equivalent to each. So this reduces our problem to that of showing that ii is equivalent to SS^{Abs}_1.

ii. $\forall F_{\in A}\forall G_{\in B^*}[\Diamond\exists x(Fx\&Gx)\supset\Box\forall y(Gy\supset Fy)]$

⇔ (by possible-world equivalences for '□' and '◊')

1 $\forall F_{\in A}\forall G_{\in B^*}[\exists w_{:R(@,w)}\exists x_{\in D(w)}(Fx\&Gx)\supset\forall v_{:R(@,v)}\forall y_{\in D(v)}(Gy\supset Fy)]$

⇔ (by quantifier movement rule from predicate logic)

2 $\forall w_{:R(@,w)}\forall v_{:R(@,v)}\forall F_{\in A}\forall G_{\in B^*}[\exists x_{\in D(w)}(Fx\&Gx)\supset\forall y_{\in D(v)}(Gy\supset Fy)]$

⇔ (by second application of quantifier movement rule from predicate logic)

3 $\forall w_{:R(@,w)}\forall v_{:R(@,v)}\forall x_{\in D(w)}\forall y_{\in D(v)}\forall F_{\in A}\forall G_{\in B^*}[(Fx\&Gx)\supset(Gy\supset Fy)]$

⇔ (by exportation and importation)

[21] Special thanks to Billy Dunaway, Ralph Wedgwood, Karen Bennett, Tristram McPherson, and a blind referee for *Philosophical Studies*.

4 $\forall w_{:R(@,w)}\forall v_{:R(@,v)}\forall x_{\in D(w)}\forall y_{\in D(v)}\forall F_{\in A}\forall G_{\in B^*}[(Gx\&Gy)\supset(Fx\supset Fy)]$

⇔ (reversing quantifier movement rule from predicate logic)

5 $\forall w_{:R(@,w)}\forall v_{:R(@,v)}\forall x_{\in D(w)}\forall y_{\in D(v)}[\exists G_{\in B^*}(Gx\&Gy)\supset\forall F_{\in A}(Fx\supset Fy)]$

⇔ (by definition of B* (the set of B-maximal properties))

6 $\forall w_{:R(@,w)}\forall v_{:R(@,v)}\forall x_{\in D(w)}\forall y_{\in D(v)}[\forall G_{\in B}(Gx\equiv Gy)\supset\forall F_{\in A}(Fx\supset Fy)]$ QED

It is a trivial corollary of the equivalence of ii and SS^{Abs} that for any B model M, $M\vDash SS^{Abs}$ just in case $M\vDash$ii.

Appendix A.2 proof that for all B models M, $M\vDash SS^{WR}$ just in case $M\vDash$i.

As with SS^{Abs}, we begin by noting that SS^{WR} can be broken up into the following two parts (i.e., that it is equivalent to their conjunction):

SS^{WR}_1 $\forall w_{:R(@,w)}\forall v_{:R(w,v)}\forall x_{\in D(w)}\forall y_{\in D(v)}[\forall G_{\in B}(Gx\equiv Gy)\supset\forall F_{\in A}(Fx\supset Fy)]$
SS^{WR}_2 $\forall w_{:R(@,w)}\forall v_{:R(w,v)}\forall x_{\in D(w)}\forall y_{\in D(v)}[\forall G_{\in B}(Gx\equiv Gy)\supset\forall F_{\in A}(Fy\supset Fx)]$

Unlike SS^{Abs}, however, SS^{WR}_1 and SS^{WR}_2 are not equivalent. (This is why we show only that in B models, $M\vDash$i just in case $M\vDash SS^{WR}$.) So we'll proceed by first showing that for any world @, SS^{WR}_1 holds at @ just in case i. does. So if $M\vDash SS^{WR}$, then $M\vDash$i, giving us one direction, and if $M\vDash$i, then $M\vDash SS^{WR}_1$, giving us one half of the other.

SS^{WR}_1 $\forall w_{:R(@,w)}\forall v_{:R(w,v)}\forall x_{\in D(w)}\forall y_{\in D(v)}[\forall G_{\in B}(Gx\equiv Gy)\supset\forall F_{\in A}(Fx\supset Fy)]$

⇔ (by the definition of B* (the set of B-maximal properties))

1 $\forall w_{:R(@,w)}\forall v_{:R(w,v)}\forall x_{\in D(w)}\forall y_{\in D(v)}[\exists G_{\in B^*}(Gx\&Gy)\supset\forall F_{\in A}(Fx\supset Fy)]$

⇔ (by quantifier movement rule from predicate logic)

2 $\forall w_{:R(@,w)}\forall v_{:R(w,v)}\forall x_{\in D(w)}\forall y_{\in D(v)}\forall F_{\in A}\forall G_{\in B^*}[(Gx\&Gy)\supset(Fx\supset Fy)]$

⇔ (by exportation and importation)

3 $\forall w_{:R(@,w)}\forall v_{:R(w,v)}\forall x_{\in D(w)}\forall y_{\in D(v)}\forall F_{\in A}\forall G_{\in B^*}[Fx\supset(Gx\supset(Gy\supset Fy))]$

⇔ (reversing quantifier movement rule from predicate logic)

4 $\forall w_{:R(@,w)}\forall x_{\in D(w)}\forall F_{\in A}[Fx\supset\forall G_{\in B^*}\forall v_{:R(w,v)}\forall y_{\in D(v)}(Gx\supset(Gy\supset Fy))]$

⇔ (second application of quantifier movement rule from predicate logic)

5 $\forall w_{:R(@,w)}\forall x_{\in D(w)}\forall F_{\in A}[Fx\supset\forall G_{\in B^*}(Gx\supset\forall v_{:R(w,v)}\forall y_{\in D(v)}(Gy\supset Fy))]$

⇔ (by possible-world equivalence for '□')

6 $\Box\forall x\forall F_{\in A}[Fx\supset\forall G_{\in B^*}(Gx\supset\Box\forall y(Gy\supset Fy))]$

⇔ (by assumption that $\forall x\exists!G_{\in B^*}(Gx)$))

i $\Box\forall x\forall F_{\in A}[Fx\supset\exists G_{\in B^*}(Gx\&\Box\forall y(Gy\supset Fy))]$

What remains, then, is to show that if M⊨i, then M⊨$SS^{WR}{}_2$. We'll show this by showing that if M⊨$SS^{WR}{}_1$, then M⊨$SS^{WR}{}_2$.

So let @ be any world in M, w be any world such that R(@,w), and v be any world such that R(w,v). To establish M⊨$SS^{WR}{}_2$, we seek to show that

$$\forall x_{\in D(w)}\forall y_{\in D(v)}[\forall G_{\in B}(Gx\equiv Gy)\supset\forall F_{\in A}(Fy\supset Fx)].$$

Since M⊨$SS^{WR}{}_1$, we know that $SS^{WR}{}_1$ holds at v – i.e., that (using a and b as bound variables over worlds)

$$\forall a_{:R(v,a)}\forall b_{:R(w,b)}\forall x_{\in D(a)}\forall y_{\in D(b)}[\forall G_{\in B}(Gx\equiv Gy)\supset\forall F_{\in A}(Fx\supset Fy)].$$

Moreover, since we are assuming that M is a B model, we know that R(v,v) and R(v,w). Hence, (substituting v for a and w for b and applying universal instantiation):

$$\forall x_{\in D(v)}\forall y_{\in D(w)}[\forall G_{\in B}(Gx\equiv Gy)\supset\forall F_{\in A}(Fx\supset Fy)].$$

But this is a notational variant on what we want to show: switching x and y:

$$\forall y_{\in D(v)}\forall x_{\in D(w)}[\forall G_{\in B}(Gy\equiv Gx)\supset\forall F_{\in A}(Fy\supset Fx)]\qquad \text{QED.}$$

references

Blackburn, Simon [1973]. 'Moral Realism.' Reprinted in Blackburn [1993], 111-129.

______ [1984]. *Spreading the Word.* Oxford: Oxford University Press.

______ [1985]. 'Supervenience Revisited.' Reprinted in Blackburn [1993], 130-148.

______ [1993]. *Essays in Quasi-Realism.* Oxford: Oxford University Press.

Fine, Kit [1994]. 'Essence and Modality.' *Philosophical Perspectives* 8: 1-16.

Hellman, Geoffrey [1985]. 'Determination and Logical Truth.' *Journal of Philosophy* 82: 607-616.

Hughes, G.E., and M.J. Cresswell [1996]. *A New Introduction to Modal Logic.* New York: Routledge.

Jackson, Frank [1998]. *From Metaphysics to Ethics.* Oxford: Oxford University Press.

Kim, Jaegwon [1984]. 'Concepts of Supervenience.' Reprinted in Kim [1993], 53-78.

______ [1987]. "Strong' and 'Global' Supervenience Revisited.' Reprinted in Kim [1993], 79-91.

______ [1993]. *Supervenience and Mind.* Cambridge: Cambridge University Press.

McLaughlin, Brian [1997]. 'Supervenience, Vagueness, and Determination.' *Philosophical Perspectives* 11: Mind, Causation, and World.

Paull, Cranston, and Ted Sider [1992]. 'In Defense of Global Supervenience.' *Philosophy and Phenomenological Research* 32: 830-845.

Schroeder, Mark [2005]. 'Realism and Reduction: The Quest for Robustness.' *The Philosophers' Imprint* 5(1): www.philosophersimprint.org/005001/.

Wedgwood, Ralph [2000]. 'The Price of Non-Reductive Physicalism.' *Nous* 34(3): 400-421.

______ [2007]. *The Nature of Normativity.* Oxford: Oxford University Press.