

# Wide Causation<sup>1</sup>

Stephen Yablo  
University of Michigan, Ann Arbor

appeared in *Philosophical Perspectives* 11 (1997), 251-281

## 1. Introduction

Are physical events subject to mental influence? Even to raise the question suggests what the answer had better be. Deny mental causation and you are denying that anyone ever does anything: answer a question or anything else.<sup>2</sup> Tongues may wag and arms may wave about, but there is no action unless these things occur at the bidding of appropriate mental states. Nor is action the only casualty if mental states are physically inert. Smirking, beaming, moping about, shivering in anticipation, raising a skeptical eyebrow, favoring a tender limb -- these are just an inkling of the human phenomena making no sense in a world where thoughts and feelings keep causally to themselves.

Of course, to say that mental states had better be physically influential does not begin to explain how such a thing is possible. And the fact is that bafflement about the how of mental causation has been growing, to the point that doubts are now creeping in about the whether. A good many philosophers seem ready to give in to these doubts and accede to some form of epiphenomenalism: here, the view that mental phenomena exert no causal influence over the course of physical events. A good many others "resist" epiphenomenalism by maneuvers so subtle that it is mainly on their own impassioned testimony that they are not counted into the first camp. Still other philosophers would junk mental phenomena altogether rather than see them causally enfeebled.

All of this adds up to what has been described as an outbreak of epiphobia. (Epiphobia =<sub>df</sub> the fear that one is turning into an epiphenomenalist.<sup>3</sup>) Even allowing for the strange logic of thought disorders, it has to be said that this one is asserting itself at rather a surprising historical moment. Epiphenomenalism was supposed to be somebody else's problem: somebody long dead, or at any rate hopelessly out of touch with recent materialist developments like multiple realization and supervenience. Why epiphobia now?

## 2. A story

Time was when epiphobics had a genuine threat to point to: the gaping divide Descartes had postulated between the mental realm, said to be lacking in kinematical qualities, and the thoroughly kinematical physical realm. Not even Descartes himself claimed to understand how causal relations were supposed to reach across this divide,<sup>4</sup> and his critics (notably Gassendi) found the notion positively incoherent:

you must explain to us how this "directing" of movement can occur without some effort -- and therefore motion -- on your part. How can there be effort directed against anything, or motion set up in it, unless there is mutual contact between what moves and what is moved?<sup>5</sup>

Even at the time, however, such worries were easily shrugged off by philosophers who, agreeing with Descartes that the mind was something apart, had their own ideas about its particular nature. (Gassendi is a case in point: "I will grant you [that you are really distinct from your body], but will not therefore grant that you are incorporeal..."<sup>6</sup>) Centuries of subsequent squabbling about the intelligibility of cross-category interaction never quite succeeded in breaking this stalemate. No argument from the gaping ontic divide between mind and body could get a grip simply because no one felt sure of what the divide's mental side looked like.

Then the brainstorm hit that the mind's precise characteristics might not matter; trouble for mental causation could be conjured out of physical assumptions alone.<sup>7</sup> Never mind whether mental causation is beyond understanding (that depends on the natures of the relata, hence in particular on the nature of mind), it is enough for the epiphenomenalist if it is beyond belief. And that mental causation is beyond belief can be maintained just on the strength of the physical realm's well attested autonomy and self-sufficiency. A strategy like this was employed by C. D. Broad in his "argument from energy"

I will to move my arm, and it moves. If the volition has anything to do with causing the movement we might expect energy to flow from my mind to my body. Thus the energy of my body ought to receive a measurable increase, not accounted for by the food that I eat and the oxygen that I breathe. But no such physically unaccountable increases of bodily energy are found,<sup>8</sup>

and his "argument from the structure of the nervous system:"

...the nervous processes involved in deliberate action do not differ in kind from those involved in reflex action; they differ only in degree of complexity....So it is unreasonable to suppose that the mind has any more to do with causing deliberate actions than it has to do with causing reflex actions.<sup>9</sup>

But it was Norman Malcolm in "The Conceivability of Mechanism" who first grasped the new genre's full potential. Assuming a physical theory rich enough to "provide sufficient causal explanations of behavior,"

the movements of the man on the ladder would be completely accounted for in terms of electrical, chemical, and mechanical processes in his body. This would surely imply that his desire or intention to retrieve his hat had nothing to do with his movement up the ladder.<sup>10</sup>

The most important stimulus to contemporary epiphobia is this argument of Malcolm's. Because it sees would-be mental causes as preempted by underlying physical states, we can call it the argument from below. Later it will be set out in more detail but the essence is simply this: with each physical effect causally guaranteed by its physical antecedents, what is there left for its mental antecedents to do?

Although the argument as stated targets mental causes, the underlying logic applies to all nonphysical states. If an effect is causally inevitable given preexisting physical conditions, then the effect's biological, geological, economic, etc. antecedents are just as much out of a job as its mental ones. Whether because of concern about the sweepingness of this result or for some other reason, attention has been shifting to a second and in some ways more discriminating argument, the argument from within.

The target this time is intentional mental states: states like belief and desire individuated in terms of truth or satisfaction conditions. If Putnam is right that truth conditions can vary between internally indiscernible agents (e.g., me and my doppelganger on Twin Earth), then intentional states are extrinsic, or not wholly a matter of what goes on within the thinker's skin.<sup>11</sup> Add to this that it is intrinsic states that determine causal powers --

you can change [extrinsic states], remove them, or imagine them to be different in various respects, without ever changing the causal powers of the object or person that is in this extrinsic condition --

and you see the problem:

how can extrinsic facts about  $\Delta$ , depending as they do on factors that are spatially and temporally remote from  $\Delta$ , help explain  $\Delta$ 's current behavior? Surely what explains, causally explains,  $\Delta$ 's raising her arm or pushing a button are intrinsic facts about  $\Delta$ .<sup>12</sup>

Any behavior that beliefs and desires might seem to generate must really be due to some intrinsic surrogate: syntactic states, perhaps, or narrowly contentful attitude-analogues, or even brain states.<sup>13</sup> Intentional causes

are thus displaced by factors internal to the agent, which gives the argument its name.

### 3. The connection

The key point for us is that mental causation faces two separate threats: "from below" and "from within." With so much effort gone into distinguishing these threats in recent years,<sup>14</sup> no one has noticed that they are connected, and in a way that makes them more formidable as a package than taken separately. Any decent response to BELOW (as I'll call it) will have to make use of the principle of

proportionality: causes must be proportional to their effects.

But, WITHIN can claim to be little more than an application of the very same principle! If that is right then we are damned if we do (accept proportionality) and damned if we don't; for either way, one of the two arguments goes through. How proportionality is supposed to play this double role is the topic of the next few sections, but in general terms the idea is this.

Start with the physical states that BELOW casts as preempters. Seen in the light of proportionality, these appear to be overloaded with unneeded microstructural detail. (Had my pain been implemented in a different microphysical way, the effect would in all likelihood still have occurred.) As for the mental states they are said to preempt, these are simply the result of stripping some of the unneeded detail away. But then to call the mental state an unneeded excrescence gets matters exactly backwards. You might as well say that since my screaming "wake up right now!!" in my cat's ear sufficed to wake him, my screaming in his ear as such made no contribution; it was only along for the ride.

All right so far. But now WITHIN chimes in that intentional "causes" are also overloaded with unneeded detail, not microstructural this time but extrinsic. Regardless of whether my desire had been for water or twater, as long as I stayed intrinsically the same the behavioral results would not have been any different. The question is, why should this excess extrinsic detail be any less offensive to proportionality than the unneeded microstructural detail of the last paragraph?

This is a question I hope eventually to answer. After various preliminaries and softening up exercises in the next few sections, the argument will unfold in three stages. First, WITHIN is not an application of the same principle used to defeat BELOW. Second, BELOW falls to a principle that is tolerant of intentional causation, albeit intentional causation of an interestingly unexpected sort. Third, WITHIN relies on an

enormously stronger principle that undermines just about any intuitive causal relation you care to mention. Details will be given in due course; for now proportionality is left at an intuitive level so as to give WITHIN the best possible run for its money.

#### 4. The argument from below<sup>15</sup>

According to closure, each physical outcome E is causally guaranteed by some prior physical C.<sup>16</sup> Dualism says that no mental C\* is identical to any physical C. Exclusion says that if E is causally guaranteed by C, then no C\* distinct from C is causally relevant to E. These three assumptions granted, no physical effect owes anything to its mental antecedents. How can it, with underlying physical states already ensuring that the effect is going to occur?<sup>17</sup>

All of the assumptions could be quarreled with, but closure and dualism can be considered the price of admission to the debate. Someone who denies dualism (e.g.) thinks that mental states are physical states and so is not interested in any supposed threat from below. It makes sense then to focus on exclusion, which has in any case an obvious problem. Look again at what is being claimed:

for every "form of nonidentity" R (every irreflexive relation) and every R-related pair C and C\*, if C is causally sufficient for an effect then C\* is causally irrelevant to it.

No doubt there are some irreflexive relations R whose relata do compete for causal influence as the principle says. But for many Rs this competition arises only sometimes, and for others it never arises. R = causation is a case in point; taken at its word, the exclusion principle predicts that E owes nothing to the causal intermediaries by which C brings E about! This shows that the exclusion principle is overdrawn. But is it overdrawn in a way that bears on the causal relevance of my pain? How plausible is it really that my pain serves as a causal intermediary between its physical basis C and my grimace?

Never mind that this would require my pain to be literally an effect of C, whereas pain intuitively stands in a closer than causal relation to its physical basis.<sup>18</sup> The relation pain bears to C is, as the word "basis" attests, often thought to be causal-like; it is considered a dependency relation of some sort.<sup>19</sup> And that ought to be just as good. The real difficulty is still to come. Much as we might like the idea of our thoughts and feeling functioning as intermediaries, how exactly are they supposed to be slotted in? If there were gaps in the physical event-chains linking brain states to behaviors, then (who knows?) mental states might perhaps find work plugging them. This would violate the exclusion principle but

only in the way that intermediaries do generally. To foist my pain on a process that is complete and self-sufficient without it, though, goes against what seems right in exclusion: a thing can do causal work only when causal work is there to be done.

## 5. Dependence

A lot of people seem to think that the best way of getting mental states in on the causal act is to make them strongly enough dependent on physical states. (Supervenient dependence is particularly recommended in this regard.) But a dependent is ontologically posterior to what it depends on, and so all the dependency hypothesis achieves is to cast my pain as a lagging indicator of the fact that a process causally sufficient for the effect is already under way. T. H. Huxley saw this implication already in the last century and did not flinch from it:

all states of consciousness ... are immediately caused by molecular changes of the brain-substance ... our mental conditions are simply the symbols in consciousness of the changes which take place automatically in the organism...the feeling we call volition is not the cause of a voluntary act, but the symbol of that state of the brain which is the immediate cause of that act.<sup>20</sup>

Those of us who do flinch from Huxley's conclusion have our work cut out for us. If mental states do not depend on "molecular changes of the brain-substance," how are they connected to brain activity? What alternative picture of mental/physical relations is available?

Of course, one clear alternative is mapped out by the identity theory. There is no question on this theory of pain's depending on (brain state) C, for C is already a state of pain. Another thing there is no question of on this theory is C's beating pain to the causal punch. It is only in the matter of truth value that the theory disappoints. Identicals necessitate one another, but any brain state specific enough to necessitate pain (a condition we assume C to meet) is too specific to be necessitated by it in return. C is thus one of a number of brain states C<sub>i</sub> each necessitating pain asymmetrically.<sup>21</sup>

No surprises so far. The surprise is that an essentially similar picture, in which (certain) brain states are already states of pain, continues to be available even if the identity theory is rejected. An analogy shows how this can be. Just as pain is not identical to any of the brain states C<sub>i</sub> that necessitate it, red is not identical to any of the more precise shades R<sub>i</sub> (scarlet, crimson, etc.) that necessitate it. Yet there is no question of redness depending on scarlet, for to be scarlet is already to be red. Scarlet is, as we say, a way of being red, or, in an older terminology, a

determinate of redness. Why shouldn't the  $C_i$ s likewise be determinates of pain?<sup>22</sup>

At last we have hit on a relation that brain states plausibly bear to mental ones<sup>23</sup> and that makes nonsense of the causal competition idea. Imagine a pigeon Sophie trained to peck at red shapes. No one would call the triangle's redness irrelevant to her pecking on the grounds that the effect was already provided for by its specific shade of red.<sup>24</sup> Nor would anyone think that my screaming as such was irrelevant since my screaming "wake up!!" was sufficient. Examples like these confirm what seems obvious anyway: determinates do not preempt their determinables.<sup>25</sup> Understand pain as a determinable of the  $C_i$ s, and preemption should not be possible in this case either.

## 6. Determinables and causation

The argument from below rests everything on a certain principle: a sufficient cause drains whatever it bears R to of causal relevance. But the principle is not true when R = the determinate/determinable relation. Since this is a relation in which physical and mental states plausibly stand, my pain can (for all anyway that BELOW has to say about it) be relevant to effects for which my brain state suffices.

To stop here though leaves the impression of a power sharing arrangement between pain and brain state -- an arrangement, indeed, favoring the brain state, since it after all suffices for an effect to which the pain claims only some unspecified relevance.

One could try to counter this impression by enlarging on what has already been said, viz. that to be in pain is part of what it is to be in such and such a brain state. When one state is included in another, any influence that the first has on subsequent events is included in the influence had by the second. Brain state and pain thus share power in a more literal sense than the one intended: not by dividing it up between themselves, in the way that books share space on a shelf with other books, but by possessing it in common, in the way that an encyclopedia shares shelf space with the volumes making it up.

And yet built into this account of how the two states share power appears to be a concession that the brain state has more power. (Just as the encyclopedia occupies more space.) This greater power shows up quantitatively in the fact that my brain state bears the most powerful form of causation -- causal sufficiency -- to more effects than my pain. And it shows up qualitatively in the fact that each of these extra events (e.g., say,

my grimace) is more the effect of my brain state than of my pain. Because again, it is the brain state that stands to the grimace in the most powerful form of the causal relation there is.

Wrong, wrong, wrong -- in a way we need the principle of proportionality to help us sort out. There is no denying that my brain state has the quantitative advantage mentioned. But sufficing for more effects is one thing, greater license to claim them as your effects, another. And proportionality says that my brain state may well be in a worse position to cause some of these additional effects than is my pain.

How we confused ourselves was by thinking of sufficiency and relevance as unequally powerful forms of causation, when in truth they are not forms of causation at all. X can be relevant to Y despite omitting factors crucially important to Y's occurrence (my addressing the cat was relevant to its waking) and sufficient for Y despite incorporating any number of irrelevant extras (its waking was causally guaranteed by my shrieking in its pointier ear at a prime number of decibels a message with the semantic content that it should immediately wake up). But X does not cause Y unless it is proportional to it, in a sense that at least implies some degree of freedom from these excesses.<sup>26</sup>

If causation is subject to a proportionality constraint, what does that say about my brain state's claim to be more the cause of my grimace than its mental competitor? Arguably it is the brain state, weighed down with superfluous microphysical detail, that suffers in the comparison. After all, I would still have grimaced even if my pain had occurred in a different microphysical way. Whereas the issue of how I would have behaved had the brain state occurred in the pain's absence cannot even be raised, because the brain state includes the pain.

## 7. The argument from within

This is where WITHIN sees its opening. I desire water and extend my hand. But of course Twin Me, who desires not water but twater, would have done the same in my circumstances<sup>27</sup> -- as indeed would anyone intrinsically just like me, even a Swampelganger Me with no intentional states whatever. So intentional states, like brain states, are overloaded with unneeded detail. The only difference is that this time the unneeded detail is "without" rather than "below."

If beliefs, desires, and the like do not cause behavior, what does? The only remaining candidates would seem to be intrinsic states of some sort: syntactical in nature, or neural, or narrow analogues of the attitudes. But we know from the Twin Earth examples that states like these do not of themselves represent the world as being any particular way. (What they



can perhaps claim is association with a staggering array of different truth-conditions, which depending on the causal/historical context in which they are imagined to be embedded; see section 11. But context aside, the intrinsic counterpart of my belief that water is wet no more concerns H<sub>2</sub>O than XYZ, and no more these than a pattern of electrical signals emanating from the walls of some brain-ready vat.) And now we see the real threat posed by WITHIN: the part of our mental life with the strongest intuitive claim to influence behavior -- the part representing the circumstances which that behavior seeks to change, and the outcomes it seeks to bring about -- may have to take a back seat to states with limited or nonexistent representational powers.

## 8. A nomic analogue

Notice a way in which this reasoning stops curiously short. Fixated as we become on the causally excessive aspects of intentional states, and determined to find relief in intrinsic surrogates, it never occurs to us to ask whether the intrinsic surrogates might not be excessive in their own way.<sup>28</sup> I want to sneak up on this question by switching temporarily (until section 13) to a nomic version of the argument.

Imagine that we are asked to find the cause of someone's receiving a speeding ticket near a police radar unit; in a familiar jargon, we are asked to solve "X caused her to be ticketed" for X. Bearing in mind proportionality's call for an X that is enough for the effect without being too much, we quickly see that there are two opposite ways of bungling the task, illustrated by

- (1) her driving through the radar caused her to be ticketed, and
- (2) her speeding through the radar sober caused her to be ticketed.

respectively. Her driving through the radar was not enough, since she had to be driving over the speed limit, while her speeding through the radar sober was too much, since her sobriety had nothing to do with it. The true cause will be an event that avoids these failings.

Now let the task be to solve "Xs ceteris paribus conduct electricity" for X -- to find a nomic ancestor rather than a causal one. Again there seem to be two roughly opposite ways of going wrong. This time let our examples be

- (3) matter c.p. conducts electricity, and
- (4) pennies c.p. conduct electricity.

Because lots of matter doesn't conduct electricity, including some paradigmatic enough not to be scared off by the ceteris paribus clause, (3) has an underspecific antecedent, making it an overgeneralization. (4) has

the opposite problem; copper conducts electricity regardless, so (4) is an undergeneralization with an overspecific antecedent.

This suggests that laws too observe a kind of proportionality constraint. For it to be a law that As are c.p. Bs, A should be determinate enough to make (other things equal) for B, but that's all; there should be no piling on of nomically irrelevant detail. Otherwise we run the risk of breaking unitary generalizations up into a large number of pointlessly different variants: "pennies conduct electricity," "copper foil conducts electricity," "the bottoms of RevereWare pans conduct electricity," and so on.

Isn't nomically irrelevant detail just what we are getting, though, in intentional generalizations like "people who want water c.p. go ahead and drink"? Any behavior issuing from my intentional states issues equally from the very different intentional states of my otherworldly Twins. Set against the intrinsic properties they and I share, that it is water I want looks like precisely the sort of nomic irrelevancy that proportionality warns against. Ignoring this warning amounts to turning our back on a great mass of unitary causal generalizations, namely all those entailed by the fact that doppelgangers behave identically despite believing and desiring different things.

That was the promised nomic analogue of WITHIN. No one could object to the principle behind it; carving up unitary generalizations is a bad thing. But if it is bad when the generalizations are over Twins, then it is bad whatever they are over. At a minimum, then, the argument is too quick. Nothing can be concluded until we consider what other generalizations might be on the chopping block; and whether it was the Twin generalizations that put them there; and which should be sacrificed if we are forced to choose.

## 9. Missed generalizations

A fact that tends to get lost in all the excitement about our Twins is that we have no Twins. Neither here on Earth nor anywhere in darkest space can molecule for molecule duplicates of flesh and blood human beings be found. As an immediate consequence, the forgone generalizations of the last paragraph (which say in effect that anyone intrinsically just like SY is in his circumstances going to do just as he does) are generalizations over things all but one of which fail to exist. This may not make the generalizations any less true, but neither does it recommend them as crashingly important.<sup>29</sup> Still less does it recommend them as crashingly more important than the generalizations we forgo if we insist on intrinsic typing -- especially since these latter range over things a great many of which do exist.

I am about to drink some water, and something tells me I'm not the only one. Is there anything about the members of this group to set us apart from the general run of others? The tempting reply is that we are the ones who want some water.<sup>30</sup> Pretending for argument's sake that soda, coffee, whiskey, and the rest are yet to be developed, so that water is the one and only drinkable, we can put the relevant law like this:

(5) people who want water c.p. have a drink.

But now notice something important about the world's water-wanters. Once we get beyond their shared extrinsic property of being in a state with waterish satisfaction conditions, they are an exceedingly miscellaneous bunch. Unprincipled disjunctions aside, any intrinsic feature they possess in common is likely to be shared as well by a good many non-water-wanters.<sup>31</sup> If we insist on intrinsic typing nevertheless, the unitary generalization (5) breaks up into a jillion variations on the theme of

(6) people intrinsically just like SY c.p. have a drink.

And why should decoupling me from my Twins, who after all don't exist, be thought worse than decoupling me from my drinking buddies, who after all do?

Stop right there, you say -- "generalizations entailed by the fact that doppelgangers behave identically" wasn't supposed to mean generalizations limited to doppelgangers, but rather generalizations subsuming doppelgangers; the plea in other words was not on behalf of (6) but something more like

(7) people in intrinsic state E c.p. have a drink,

where F is some limited shareable aspect of SY's total intrinsic state. So, contrary to the last paragraph, the generalizations we forgo by typing intentionally, and incur by typing intrinsically, have plenty of real world instances. Add to this that (7) improves on (5) in extending to these instances' counterfactual doppelgangers, and the verdict is clear.

Some such line of response is the intrinsicalist's best bet. But it overlooks one thing: only (6) can be described as a generalization entailed by the fact that doppelgangers behave identically in the same circumstances.<sup>32</sup> What we have in (7) is the form of a generalization that the intrinsicalist hopes for. Generalizations like this may well exist, but only if it proves possible to pare my total intrinsic state down to a part that, while specific enough to make c.p. for drinking, is not so specific as to be peculiar to myself. And the fact that doppelgangers behave identically cannot itself decide this issue -- at least, not by any argument that we have yet seen.

## 10. Bracketing

To think of (7) as a casualty of intentional typing is premature; all that it represents so far is a lost opportunity. And yet, it is possible to wonder how can there fail to be interesting (7)-type generalizations. Aren't these guaranteed, more or less, by the existence of (5)-type intentional generalizations, together with the fact that how people behave in a given situation depends only on what they are like intrinsically? If it is true, for instance, that

(5) people who want water c.p. have a drink,

then given the irrelevance to this behavior of their purely extrinsic features, it should also be true that

(8) people who [want water] c.p. have a drink --

where [wanting water] is wanting water with its purely extrinsic aspects bracketed away.<sup>33</sup> This amounts in fact to a recipe for intrinsicalizing intentional generalizations like (5). Simply substitute for each offending attitude the corresponding battitude, that is, its image under the operation of bracketing.<sup>34</sup>

Sounds promising, but why stop there? If the recipe works at all, it gives a way of intrinsicalizing nonintentional generalizations as well: substitute for each offending G the corresponding [G]. I have heard, for example, that people from large families are by and large gregarious. But gregariousness in a given context depends on intrinsic features alone; a gregarious person's intrinsic duplicates are not going to be taciturn and

withdrawn. Apparently then there has got to be an intrinsic property of [being from a large family] -- the intrinsic "core" of being from a large family -- that also makes c.p. for gregariousness. Similarly, poor people must share an intrinsic property of [poverty] that accounts for their feelings of not having enough food in their stomachs. And now the fallacy must be plain. The most that follows from the irrelevance of the purely extrinsic is that each water-wanter has some intrinsic feature or other that leads c.p. to drinking. The further conclusion that they share an intrinsic feature that leads to drinking is just wishful thinking.

## 11. Battitudes

How wishful it is can be seen by looking at the two main theories of the battitudes. One gives us states that are shared but not sufficiently specific, the other, states that are specific but not shared.

The simpler of the two theories says that you share my [belief that p] iff some possible doppelganger of yours believes that p.<sup>35</sup> (Similarly for desire and the other attitudes.) Twin Me on Twin Earth [wants water], for instance, since he has a doppelganger, myself, who wants water. Doppelgangers of other terrestrial water-wanters [want water] too, not only on Twin Earth but on all planets, be they actual or hypothetical.

But it is not just doppelgangers elsewhere of terrestrial water-wanters who [want water]. Doppelgangers here of extraterrestrial water-wanters [want water] too. And now it becomes hard to think who does not [want water] on this theory. For let Dino be a person wanting essentially any old thing.<sup>36</sup> And let Twin Dino be Dino's doppelganger in a world where the-thing-that-manifests-itself-in-the-way-that-the-object-of-Dino's-desire-actually-manifests-itself is water. Twin Dino wants water in that world, so Dino [wants water] in this one. Battitudes as explained by the first theory are thus wildly underspecific, turning "people [wanting water] c.p. have a drink" into a gross overgeneralization along the lines of "matter conducts electricity."

Why does the theory deliver such coarse-grained results? Reformulate it like so and the reasons jump out: you and I share a battitude iff there are worldly contexts, not necessarily identical, in which your doppelganger judges the same proposition as mine. Each of the highlighted phrases makes for a separate kind of trouble. "Not necessarily identical" leaves the door open to tailoring the two contexts so as to offset bona fide battitudinal differences. Perhaps Dino is a martini fiend would sooner chew tinfoil than take in a drop of any other liquid, but the fact that he (or rather, his doppelganger) would want water in a world where it was water that lay behind martini-appearances suffices to make him a [water-wanter] like me.

All right; we need to drop the "not necessarily identical" and require that the same proposition be judged in the same context. That we are free to choose this context at will (remember the quantifier "there are") means that a problem remains. Attitudes that are capable of latching onto different propositions are absolutely distinct.<sup>37</sup> But a capability is not the sort of thing that every context can be relied on to register. Sargon's [longing to visit the Morning Star] was quite a different attitude from his [longing to visit the Evening Star], even if he lived out his days in a setting where their distinctness did not manifest itself in different propositional outcomes desired.

Where are we? Not only should attitudinalizers be compared in the same context, that context should be allowed to vary arbitrarily. Both of these modifications together give us the second main theory of the attitudes.<sup>38</sup> For someone to share my [belief that p], their doppelganger in w should believe (not the proposition p that I in fact believe, as on the first theory but) the proposition p(w) that my doppelganger in w believes.<sup>39</sup> And this should hold not for a single world w (as on the first theory) but all of them.<sup>40</sup> Put another way, attitudes are individuated by the functions they induce from worldly contexts to the singular propositions that get judged in those contexts.

Not just anyone is going to share my [belief that p] on the new approach. They are going to have to grasp or conceive p at least somewhat as I do, lest the difference induce a different proposition believed in some faraway world. But how much similarity of conception are we talking about here? All it takes for a thinker not to share my [belief that p], remember, is that there be something in their take on reality, no matter how little connected to p, that in some world w, however distant or contrived, swings the propositional content of their belief away from that of my doppelganger in w. And it is natural to wonder whether there is any difference in [attitude] that could not be exploited to achieve this result in a suitably wacky world.

Here is why. What my [belief that p] is about in a world w depends on what it is in covariational thrall to there. But on anybody's account, the covariational channels through which content flows are shaped and sustained by various sorts of external props: paradigms, measuring devices, experts, and the like. No doubt there are worlds in which all available props converge on the same external referent; all the instruments agree as it were. But there will also be worlds in which switching the prop puts the thinker en rapport with a different referent. Any change in [attitude] with even the potential to shift my allegiances as between props thus engenders an actual change in the function from contexts to attitudes that constitutes my [belief that p]. And it is hard to think how a change in [attitude] could lack this potential -- how I could "change my mind" without in any circumstances

whatever tipping the balance in favor of deference to a different class of paradigms, measuring devices, experts, or what have you. Variation in any [attitude] therefore entails variation in all of them.<sup>41</sup>

This problem for the second theory of the battitudes can be called subjective meltdown. Because what we are seeing is that to share my [belief that p], you must share my total subjective outlook -- or, what comes to the same, my [belief that p] is my total subjective outlook.<sup>42</sup> If there is anything to subjective meltdown at all, the second theory essentially just inverts the difficulties we found with the first; it delivers overspecific battitudes, turning "[water-wanters] c.p. drink" into an undergeneralization along the lines of "pennies conduct electricity."

## 12. Battitudes as overcommittal anyway

Now for the "real" reason not to take it for granted that proportionality backs the battitudes over the attitudes. This is a reason that continues to apply even if subjective meltdown is somehow avoided -- even if battitudes are the separately identifiable cognitively revealing items their proponents have wanted them to be.

Imagine that to each of my attitudes A corresponds a distinct subjective state [A]<sub>SY</sub> that sums up what I within the privacy of my own head to be in A.<sup>43</sup> So, [desire for water]<sub>SY</sub> is what I do internally to desire water (as opposed to what I do to desire fried green tomatoes or to believe that okra is slimy). The state thus picked out might be a hankering after the odorless, tasteless, transparent, river-filling stuff that etc. etc. But it might equally well be some sort of syntactical and/or neural state.

The point in either case is the same. The appeal to these states can exempt the intrinsicalist from charges of fracturing the intentional generalization (5) only on a certain condition: all or almost all water-wanters must be in the state of [wanting water]<sub>S</sub> for some value of S. And this is just not plausible. How a person's water-desire is neurally implemented, the precise mentalese orthography involved, the fine detail of the water's internal mode of presentation, all of these may be expected to vary enormously without much effect (ceteris paribus) on the desirer's probability of drinking.

## 13. Back to WITHIN

A case can thus be made that WITHIN's nomic analogue is guilty of double dealing. After much handwringing about intentional states' overspecificity relative to this or that intrinsic surrogate, that the

surrogate states are similarly overspecific relative to their intentional originals is completely overlooked. The question is how much of this transfers over to WITHIN itself, which you'll recall goes as follows: According to the proportionality principle, causes should not be overloaded with unneeded detail -- detail in whose absence the effect would still have occurred. But unneeded detail is exactly what we are getting when my desire for water is nominated as the cause of my hand going out to the cup. Had it been twater I wanted rather than water, then, holding my [desire] fixed, my hand would still have gone out.

Now, it might well be asked why (in the absence of information about how I came by my altered desire) this counterfactual should strike us as correct rather than merely baffling. But our problem is much more basic. Assume that my [desire] does screen off my desire in the way described. This can't itself put my [desire] in the driver's seat, for my desire might well return the favor. Even if it is true, in other words, that

(9) had my desire been different, then provided my [desire] had been the same, my hand would still have gone out,

it might also be the case that

(10) had my [desire] been different, then provided my desire had been the same, my hand would still have gone out.<sup>44</sup>

And since (9) and (10) are absolutely symmetrical, any causal advantage the one might seem to confer on my [desire] is nullified by the other.

## 14. Symmetry

At least, my [desire]'s advantage is nullified if (10) is true. The intrinsicalist will say that it is not. Don't we have a million Frege-inspired examples to show that tiny differences in the way a proposition is presented can have enormous behavioral ramifications? Whereas if different propositions are presented in the same way (as in the Twin Earth examples), the same behavior results. The clear lesson of these examples is that behavior is driven less by what one believes/desires -- by the propositional content of one's attitude -- than by how that content is grasped. And if so then the very last thing we would expect is that switching the [desire] behind my desire, as in (10), will leave my behavior in place.

Sorry, but the clear lesson of the Frege and Twin Earth examples is only this. If we distinguish "what I believe/desire" from "how I believe/desire it" as factors in my extending my hand, then adjusting the



how-factor alone can affect my behavior while adjusting the what-factor alone cannot. And this is compatible with (10), as an example brings out.

Whenever Isaac spots his bubbe in a photograph, he grins in recognition. Distinguish two factors in the grin on his face right now: what the photograph depicts (its subject or subjects, in this case my mother), and how it depicts (how intrinsically speaking the colors are arrayed). These two factors interact in something very like the way under discussion. Adjusting the how-factor alone can affect Isaac's behavior -- had the photograph been much fuzzier Isaac would have been baffled by it -- while adjusting the what-factor alone cannot -- leave the colors alone, and regardless of subject, Isaac grins. Shouldn't we then conclude that Isaac's behavior is controlled more by the picture's intrinsic color properties than by its extrinsic, representational, ones? And if it is controlled more by the color properties, then the very last thing we would expect is that a differently colored picture of his bubbe would still have lead Isaac to grin.<sup>45</sup>

And yet, this is precisely what we would expect. Isaac is a boy capable of tracking his bubbe through a huge variety of photographic images, and the image at issue here is not anything far out but the one his bubbe would have given rise to if the actual image were for some reason ruled out. Why Isaac should suddenly lose sight of his bubbe in the alternative-image world nearest to this one is hard to understand. Harping on the fact that a change in intrinsic color properties is necessary and, if suitably dramatic, sufficient for a change in Isaac's reaction only drives the problem home; why should there be a dramatic change in color properties in the nearest alternative-image world to actuality?

Here is what the Frege and Twin Earth cases may indeed show: if you want to stop me from extending my hand, mucking with my [water-desire] alone can do it, whereas mucking with my water-desire alone cannot. But this is fully compatible with saying that many or most ways of mucking with my [desire] leave my behavior in place, provided that I keep on wanting water. And it is supremely compatible with the notion that I would still have extended my hand if I had wanted water in the way involving the least possible departure from actuality.<sup>46</sup>

So what, in other words, if a desire for water conceived as the-stuff-I-once-saw-through-an-electron-microscope, or as whatever-she's-drinking, would not have set my hand in motion? The screening off issue concerns not these modes of presentation but the one(s) I would have enjoyed had I not conceived of water in the way that I in fact do. (Given the richness and multifacetedness of my actual conception of water, it would seem bizarre for there to be no closer alternative to my actual conception of water than one robbing my desire of its motive power.) So what if a sufficiently perverse mentalese encoding would have cut my desire off

from its behavioral effects, as long as the closest alternative encoding(s) are not perverse?

## 15. Proportionality

Generalizing madly, let us assert the following: any intrinsic state rich and complex enough to count as what-I-do-internally-to-judge-that-p is bound to exceed in some respects the causal requirements of any particular bit of behavior. If that is right, then the intrinsic causes that WITHIN favors are as open to charges of disproportionality as the extrinsic, intentional, causes that it rejects. Either the charges stand up in both cases -- in which case nothing causes behavior -- or they stand up in neither case. I say that they stand up in neither case. But then there must something wrong with WITHIN's understanding of proportionality.

What could it be? Proportionality has been kept at an intuitive level until now, mainly in order not to rain prematurely on WITHIN's claim to be relying on the same principle used in the response to BELOW. Suppose we look at that response again, this time with an eye to what it has in mind by proportionality:

my brain state cannot expose my pain as causally irrelevant to my grimace, because it is a determinate of my pain; my pain, however, can knock my brain state out of contention for the role of cause, by screening it off and so exposing it as not required for, and hence out of proportion with, my grimace.

Working backwards, my brain state is not proportional to my grimace because it is not required; and it is not required because my pain -- one of its determinables, note -- screens it off. Here are the definitions right way around:<sup>47</sup>

(11) C<sub>1</sub> screens C<sub>2</sub> off from E iff, had C<sub>1</sub> occurred without C<sub>2</sub>, E would still have occurred.

(12) C is required for E iff none of its determinables screens it off, and C is enough for E iff it screens off all of its determinates.<sup>48</sup>

(13) C is proportional to E iff it is both required by and enough for E.<sup>49</sup>

To complete the response we should explain how my pain, having knocked my brain state out of contention for the role of cause, might come to occupy that role itself. If it were to screen off its other determinates (other than the brain state, that is), then by (12), my pain would be enough for the grimace. If it escaped a similar fate at the hands of its determinables, then by (12) again, it would be required. Both results together would by (13) make my pain proportional to the grimace and to that extent its cause.

## 16. Thick and thin

Here is what proportionality means in the response to BELOW: you are proportional iff you screen off your determinates, and you avoid being screened off by your determinables. The question is whether WITHIN can get by on the same interpretation. Does the fact that attitude  $\Delta$  is screened off from a behavioral effect by battitude  $[\Delta]$  knock  $\Delta$  out of proportion with that effect in the sense of proportionality just laid down?

That depends on how we resolve an unremarked ambiguity in talk of attitudes like  $\Delta$ . That  $\Delta$  is extrinsic is agreed (remember Putnam and Twin Earth). But an extrinsic state need not be extrinsic through and through; it can have intrinsic parts or aspects. This is obvious in the case of rigged-up examples like being spherical and P, where  $P$  is a property as extrinsic as you like. But there are plenty of ordinary examples as well. Being a horse (stamp, crater, ...) involves a horsy history together with a horsy intrinsic character. Even that paradigm of extrinsicness, the property of being five miles from a burning barn, is not altogether free of intrinsic content. To be five miles from anything you need spatial boundaries, and it seems an intrinsic property of a thing that, along some dimensions at least, it finally peters out.

Now, from the Twin Earth examples we know that  $\Delta$  is extrinsic in respect of its truth-conditions or singular propositional content.<sup>50</sup> But this does not prevent it from being intrinsic in other respects. One possibility is that  $\Delta$  includes the thinker's internal contribution to the fact that such and such is the truth-conditional content she judges; that is,  $\Delta$  might be a determinate of  $[\Delta]$  = its image under the bracketing operation. Attitudes like this, which have their corresponding battitudes as determinables, will be called thick. Another possibility is that  $\Delta$  is (relatively) noncommittal about the thinker's internal contribution; it is not a determinate of  $[\Delta]$ . Attitudes like this will be called thin.<sup>51</sup>

How does the thick/thin distinction affect  $[\Delta]$ 's ability to knock  $\Delta$  out of proportion with behavioral effects? Simple -- thick  $\Delta$  has  $[\Delta]$  as a determinable, and (11)-(13) say that  $\Delta$  had better not be screened off by any of its determinables if it wants to come out proportional with  $\underline{E}$ .  $\Delta$  is not proportional to  $\underline{E}$ , then, if it is screened off by  $[\Delta]$ . (Compare: my screaming "wake up!!" in my cat's ear is not proportional to its waking up if it is screened off by my screaming in the cat's ear as such.) Whether screening off in fact occurs depends on the details of the case -- on whether  $\underline{E}$  would still have occurred had the thinker judged a different proposition by way of the same battitude. But if the factors responsible for the switch in proposition are far enough removed from the causal scene, then  $\underline{E}$  is probably not going to be affected.

About thick attitudes, then, WITHIN has a point; they really are in danger of being knocked out of proportion with typical behavioral effects by their intrinsic counterparts. But if you have been following me this far, you will see that thin attitudes are in no comparable danger. This is because thin A has no intrinsic determinables worth speaking of -- certainly not [A], for A does not determine [A]<sup>52</sup> -- and it takes a determinable of A to expose it as not required for the effect. Thin attitudes have nothing to fear from WITHIN.

## 17. Superproportionality

If a determinable of C screens it off, then C is not required for E. But, why the restriction to determinables? What is so special about them that only they have the power to break C's causal connection with E? This is crucial because extending equivalent veto power to non-determinables would bring thin A under the same proportionality pressure as thick.<sup>53</sup> And if thin A loses its advantage over thick, then not much remains of our defense of wide causation.

So again, what is so special about C's determinables? And while we're at it, what is so special about its determinates that C need only screen them off to be proportional to E?

Nothing, you might say. The fact that C is screened off at all shows that, suitable other things holding fixed, the effect would still have occurred without it. And even a single state not screened off by C shows that C cannot itself supply all of the effect's causal needs. What is the point of a proportionality condition if not to show "causes" like this the door? Never mind the ineffectual (12) and (13); let's have

(14) C is superrequired for E iff nothing screens it off, and superenough for E iff there is nothing it fails to screen off<sup>54</sup>  
and

(15) C is superproportional to E iff C is superrequired and superenough for E.

Bertrand Russell seems to have been in the grip of some such idea in "On the Notion of Cause." Because here is what he argued, or provided the materials for arguing:

C cannot cause a strictly later event E except via some causal intermediary D. But then C is not superenough for E, since it would not have been followed by E but for D's assistance.<sup>55</sup> (Nor is it superrequired, since given D it makes no difference to E whether C occurs or not.) So there can be no temporal gap between cause and effect. Can we at least say that C begins earlier than E, ending no earlier than the time at which E begins? No, for the parts of C occurring prior to E would have to be written out as not superrequired.<sup>56</sup> The only true causation is simultaneous causation.<sup>57</sup>

So much for the old canard about the future being causally beholden to the past. By the time Russell is done the universe has disintegrated into a loose succession of moments, each sponsoring feverish causal activity on a rigidly intramural basis.

Now, Russell intended his argument as a reductio of the whole notion of cause. But it works better as a reductio of (13)'s overheated conception of proportionality. The real lesson of Russell's argument is that to insist that causes screen off subsequent events, while not being screened off by them in return, imposes an absurd degree of intimacy on causal relations. This perhaps explains why no one has ever tried to deduce epiphenomenalism from the fact that mental states are screened off by the causal chains they extend towards behavior. If this sort of screening off were truly disqualifying, epiphenomenalism would be the least of our problems; essentially everything would be robbed of its intuitive causal powers.

No one imagines it makes beliefs and desires epiphenomenal to be screened off by events subsequent to themselves. But many do seem to think it makes them epiphenomenal that they are screened off by associated [beliefs] and [desires]. This is interesting because it seems to me that to count this sort of screening off disqualifying also imposes a disastrous degree of intimacy on causal relations.<sup>58</sup> The difference is that now the intimacy is of a modal nature rather than a temporal one. Instead of being forced to exist at the same times, C and E are forced to occur at the same or similar worlds.

## 18. Dedicated pseudocauses

Why a modal intimacy this time? Because it is primarily in modal respects that attitudes differ from their corresponding battitudes. As far as this world is concerned, my desire for water and my [desire for water] are just alike. They occur at the same time and, Putnam's slogan that "meanings ain't in the head" notwithstanding, in the same place. (He might as well have said that pennies ain't in the pocket, since events within the pocket do not suffice to make them pennies.) To the extent that content is categorical, they can even be said to have the same content or contents. All of their categorical properties are shared, or near enough not to matter. Where my desire and my [desire] differ is in which of these properties they have essentially, or, what comes to the same, in their counterfactual careers. The desire persists into worlds where it is water that I want, even water grasped in a different intrinsic way; the [desire] persists into worlds where it is thusly that I want, even if the thing thusly wanted is not water.

Using the term coincident for items that are categorically alike but hypothetically different, we can put the claim like this. Applied to events occurring at different times, superproportionality imposed an undue degree of

temporal intimacy; applied to coincident events (events occurring at the same time but in different worlds) it imposes an undue degree of modal intimacy.<sup>59</sup> Epiphenomenalism is the least of our problems either way, because too much intimacy of either sort makes an absolute hash of the causal order.

A few sections back we saw how coincident would-be causes can screen one another off; in the terms of (14), each exposes the other as not superrequired for the effect. That was just the tip of the iceberg, however. Another scenario involves three candidate causes, all coincident, with the first screening off the second, the second screening off the third, and the third screening off the first. (Had the miller girl guessed the little man's name without guessing "Rumpelstiltskin" -- his name was "Ralph" -- or guessed "Rumpelstiltskin" without guessing his deepest secret -- he had a still deeper secret -- or guessed his deepest secret without guessing his name -- "Rumpelstiltskin" was not his name but that of his invisible friend -- he would still have stamped himself into the ground.<sup>60</sup>) Again, none of the candidate causes is superproportional with the effect. How often does this sort of situation arise?

Here are some crude statistics to suggest what the superproportionalist is up against. If  $C_1, \dots, C_n$  are coincident events each up for the role of causing  $E$ , then  $C_i$  causes  $E$ , according to superproportionality, only if

for all  $C_j$ ,  $E$  would still have occurred had  $C_i$  occurred in  $C_j$ 's absence, and  
for all  $C_j$ ,  $E$  would not have occurred had  $C_j$  occurred in  $C_i$ 's absence.

Call the scenario where none of the  $C_i$ s passes this test -- where each has its candidacy destroyed by some other -- collective self-destruction. What we are after is an estimate of its probability. As a basis for calculation let's say that between the hypothesis that  $E$  would have occurred had  $C_j$  occurred without  $C_i$ , and the hypothesis that it wouldn't have occurred, there is nothing to choose; one candidate cause is a priori as likely to screen another off as not to do so. (This is debatable but never mind; any other estimate only increases the chances of collective self-destruction.) Then the probability of  $C_i$ 's escaping elimination at the hands of  $C_j$  is  $1/4$  - - for there is half a chance of its being screened off by  $C_j$  and half a chance of its failing to screen  $C_j$  off. Assuming that these probabilities are relevantly independent,<sup>61</sup> we can reason as follows:

the chance of  $C_i$  escaping elimination by  $C_j = 1/4$ , so  
the chance of  $C_i$  escaping elimination altogether =  $(1/4)^{n-1}$ , so  
the chance of  $C_i$  being eliminated =  $1-(1/4)^{n-1}$ , so  
the chance of each  $C_i$  being eliminated =  $(1-(1/4)^{n-1})^n$ .

This is not a negligible figure, even for small values of  $n$ . With two candidate causes, self-destruction is 56% likely; with three it is 82% likely; with four it is 94% likely; and with five it is 98% likely. With six candidate causes there is only one chance in a hundred that some  $C_i$  will stave off elimination.<sup>62</sup>

It is true that the "right" candidate cause could beat the odds. But think what "right" has to mean here. A  $C_i$  which occurred in the very same worlds as  $E$  would not be in any danger. But any departure from this ideal is potentially a departure from superproportionality. For  $E$  to occur without benefit of  $C_i$  in even a single world  $w$  opens  $C_i$  up to charges of not being superrequired for  $E$ . (What it would take to make the charges stick is a  $C_j$  such that  $w$  = the closest world to actuality in which  $C_j$  occurs in  $C_i$ 's absence.) Likewise a single world in which  $C_i$  occurs without  $E$  opens  $C_i$  up to charges of not being superenough for  $E$ . (Here we would need a  $C_j$  such that  $w$  = the closest world in which  $C_j$  occurs in  $C_i$ 's absence.) Superproportionality comes perilously close to the demand that causes be unconditionally necessary and sufficient for their effects -- as close as the pool of candidate causes permits.<sup>63</sup>

Pressurizing causes to exist in the same worlds as their effects is a bad idea. That  $E$  is not likely to have an antecedent quite this modally attuned to it is only part of the problem. Even if such an antecedent were found, call it  $C_e$ , we would be hard put to regard it as  $E$ 's cause. After all, this would be an event with existence-conditions roughly as follows:  $E$ 's causal needs are somehow or other met. Surely it is not  $E$ 's causal needs being met that does the causing, it's the whatever-it-is that in fact meets them. (Shades of dormativity.)

Imagine though that we stifle our doubts and accept  $C_e$  as cause; then our troubles are just begun. An event so closely identified with  $E$  is in a poor position to cause other effects, especially if causation requires the high degree of modal attunement now being contemplated. (Do not suppose that it will cause these other effects via  $E$ . Ordinary events like  $E$  have long since fallen out of superproportion with their supposed progeny.) And how  $C_e$  is supposed to be provided with a superproportional cause of its own is anybody's guess.<sup>64</sup> Any comfort that superproportionality might seem to lend epiphenomenalism is thus a sideshow compared to its real project. The world we have now is a richly connected cosmos, run through with multiply branching causal chains. Given the right sort of ammunition (the right pool of candidate causes) superproportionality would lay waste to this arrangement, leaving behind a great disorderly mass of effects each tracing back to an unmoved mover dedicated precisely to it.<sup>65</sup>

## 19. BELOW vs. WITHIN

About one thing WITHIN is right: intentional causes incorporate unneeded detail. But all intuitive causes, intentional or not, are like this. Do we really want to deny that the miller girl's guessing "Rumpelstiltskin" caused the little man to stamp himself into the ground, on the basis that so long as she had guessed his name (whatever it happened to be) the result would have been the same? If so then we are well on the way to a world of dedicated pseudocauses consisting in whatever it takes for a given effect to eventuate.

Now this, coming on the heels of our objection to brain states as incorporating unneeded microphysical detail, may seem to raise double dealing to new heights. Intentional causes can do it, but neural ones cannot -- is that it? But there is an objective difference here. C incorporates unneeded detail iff it incorporates detail that (as I keep on saying) the effect could have done without, suitable other things holding fixed. Focus with me for a moment on these "suitable other things," the ones in whose continued presence the effect would still have occurred. Are they included in C, or do they lie outside it?

If the suitable other things are included in C, then they remain when the unneeded detail is stripped away. Stripping that detail away therefore yields a determinable of C that would still have been succeeded by the effect even in C's absence. This is how it is with my brain state and my pain.

Now suppose that the suitable other things are not included in C, as when C is a thin attitude and what gets held fixed is [C]. Then the result of stripping the unneeded detail away (unneeded extrinsic detail in the case of interest) is too impoverished to do meaningful causal work.<sup>66</sup> This sort of unneeded detail will have to be tolerated, because there is nothing to cover for it in its absence.

The point not to lose sight of is that there is no hope of evading the difficulty by attempting to compensate the cause somehow for its extrinsic losses. This can only push the bulge elsewhere, because apart from the whatever-it-takes pseudocauses rejected above, all causes, even purely intrinsic ones, contain an element of the unneeded. Tradeoffs are unavoidable; we buy relief from one sort of unneeded detail by taking on detail of another sort. When the tradeoffs balance out, we can attribute the effect to a relatively extrinsic cause or a relatively intrinsic one as we choose.<sup>67</sup> When a modicum of extrinsic detail buys up an abundance of intrinsic, we have wide causation pure and simple.

## 20. Innocence



If we could but recover our pre-Fregean intentional innocence, it would seem incredible that the desire leading me to reach just now for water had much more to its content than this: I get water.<sup>68</sup> What normally and primarily drives behavior is outwardly directed attitudes, not how those attitudes happen to be encoded in people's heads.<sup>69</sup>

And a good thing too. Because think what life would be like if the same truth-conditional contents, variously grasped, induced a comparable variety of behaviors. Frustrating, that's what. The more behaviors a fixed set of attitudes issues in, the harder it becomes for these behaviors to converge on desired results.

How is it that people are so good at getting what they want?<sup>70</sup> Three generalizations go a long way towards accounting for this. First, people have a tendency to do the subjectively reasonable thing, as defined by their [desires] and [beliefs]. (Decision theory is not a complete descriptive failure.) Second, the subjectively reasonable thing is quite often the objectively reasonable thing, in a sense defined by the agent's desires and beliefs. (Lois Lane snubbing Clark Kent, whom she de re adores, is the exception that proves the rule; it piques our interest because it doesn't usually happen.) Third, the objectively reasonable action is quite often the objectively right action, in a sense given by the agent's desires and the world. (Facts relevant to the success of our behavior are generally known to us.)

Imagine that we were the sort of creature that was liable to be driven hither and thither by variation in [desire] and [belief], even with all relevant desires and beliefs held fixed. Then the second of the three generalizations would be undermined. According to it, the subjectively reasonable action (the one that typically gets performed, remember) tends to be the objectively reasonable action. But how is it possible for these actions to remain the same when the one is changing with each shift in [attitude] and the other is staying put? Sensitivity to pure variation in [attitude] hurts our chances of doing the objectively reasonable thing, and hence of doing the objectively right thing, and hence of obtaining satisfaction.<sup>71</sup>

## 21. Conclusion

Nourished from earliest days on a one-sided diet of Frege examples, and impressed by the vast causal difference a slight shift in subjective conception can make, philosophers have assumed that the richer intentional states are in subjective detail, the better adapted they are to the causation of behavior.

But (as one might have guessed from the fact that it took a Frege to think them up) Frege examples are special. What ordinarily happens is that the agent could have grasped her proposition in any number of ways at no cost to the ensuing behavior. This and related oversights lead the standard view to reverse the true state of affairs. The richer an intentional state is in subjective detail, the more proportionality argues for rejecting it in favor of its subjective core.<sup>72</sup> Better equipped for causal duty are subjectively impoverished attitudes. These are safe from WITHIN, and, stressing as they do the external situation grasped over subjective nuances, more commensurate with typical behavioral effects. Normally I reach for water because I want water, never mind the phenomenology.<sup>73</sup>

## BIBLIOGRAPHY

- Anderson, C.A. and Owens, J. 1990. Propositional Attitudes: The role of Content in Logic, Language, and Mind (Stanford: CSLI)
- Bedau, M. 1986. "Cartesian Interaction," in Midwest Studies in Philosophy X, 483-502
- Block, N. 1991. "What Narrow Content is Not," in Loewer & Rey 1991
- Block, N. 1980. Readings in the Philosophy of Psychology, vol. 1 (Cambridge: Harvard University Press)
- Broad, C. D. 1925. Mind and its Place in Nature (London: Routledge & Kegan Paul)
- Brown, C. 1993. "Belief States and Narrow Content," Mind & Language 8, 343-365
- Cottingham, R. Stoothoff, and D. Murdoch. 1984: The Philosophical Writings of Descartes, vol. 2 (Cambridge: CUP)
- Davidson, D. 1984. Inquiries into Truth & Interpretation (Oxford: OUP)
- Dretske, F. 1988. Explaining Behavior: Reasons in a World of Causes (Cambridge, MA: MIT)
- Dretske, F. 1993. "The Nature of Thought," Philosophical Studies 70, 185-199

- Fodor, J. 1980a. "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology," Behavioral and Brain Sciences 3, 63-98; reprinted in Fodor 1981
- Fodor, J. 1980b. "Methodological Solipsism: Reply to Commentators," Behavioral and Brain Sciences 3, 99-109
- Fodor, J. 1981. Representations (Cambridge, MA: MIT)
- Fodor, J. 1982. "Cognitive Science and the Twin-Earth Problem," Notre Dame Journal of Formal Logic 23, 98-119
- Fodor, J. 1987. Psychosemantics (Cambridge, MA: MIT)
- Fodor, J. 1986. "Individualism and Supervenience," Proceedings of the Aristotelian Society, supp. vol. 60, 235-262
- Fodor, J. 1989. "Making Mind Matter More," Phil. Topics 17, 59-79
- Fodor, J. 1991a. "Replies to Critics" in Loewer & Rey 1991
- Fodor, J. 1991b. A Theory of Content and Other Essays (Cambridge, MA: MIT)
- Fodor, J. 1991c. "A Modal Argument for Narrow Content," Journal of Philosophy 88, 5-26
- Fodor, J. 1994. The Elm and The Expert: Mentalese and Its Semantics (Cambridge, MA: MIT)
- Heil, J. & Mele, A. 1993. Mental Causation (Oxford: OUP)
- Horgan, T. 1991. "Actions, Reasons, and the Explanatory Role of Content," in McLaughlin 1991
- Huxley, T.H. 1911. Method and Results (New York: Appleton)
- Jackson, F. & Pettit, P. 1988. "Functionalism and Broad Content," Mind 97, 381-400
- Kim, J. 1989. "Mechanism, Purpose, and Explanatory Exclusion," Philosophical Perspectives 3, 77-108; reprinted in Kim 1993
- Kim, J. 1991. "Dretske on How Reasons Explain Behavior," in McLaughlin 1991; reprinted in Kim 1993
- Kim, J. 1993. Supervenience and Mind (Cambridge, UK: CUP)

- Loar, B. 1985. "Social Content and Psychological Content," in Grimm and Merrill, Contents of Thought (Tucson: University of Arizona Press)
- Loewer, B. and G. Rey, 1991. Meaning in Mind: Fodor and his Critics (Oxford: Blackwell)
- Long, A. A., & Sedley, D. N. 1987. The Hellenistic Philosophers, vol. 1 (Cambridge, UK: CUP)
- Malcolm, N. 1968. "The Conceivability of Mechanism," Philosophical Review 77, 45-72; reprinted in Watson 1982
- McLaughlin, B. 1991. Dretske and His Critics (Oxford: Blackwell)
- Millikan, R. 1993. White Queen Psychology and Other Essays for Alice (Cambridge, MA: MIT)
- Owens, J. 1987. "In Defense of a Different Doppelganger," Philosophical Review 96, 521-554
- Pettit, P. & McDowell, J. 1986. Subject, Thought, and Context (Oxford: OUP)
- Putnam, H. 1975. "The meaning of 'meaning'," in Putnam, Mind, Language, and Reality (Cambridge: CUP)
- Rudder Baker, L. 1995. Explaining Attitudes (Cambridge, MA: CUP)
- Russell, B. 1917. Mysticism and Logic (London: Allen & Unwin)
- Schiffer, S. 1989. Remnants of Meaning (Cambridge, MA: MIT)
- Searle, J. 1983. Intentionality (Cambridge, UK: CUP)
- Searle, J. 1995. "The Mystery of Consciousness," New York Review of Books 42, no.17, 60-66
- Stalnaker, R. 1989. "On What's In the Head," Philosophical Perspectives 3, 287-316
- Stalnaker, R. 1990. "Narrow Content," in Anderson & Owens 1990
- Stalnaker, R. 1991. "How to Do Semantics for the Language of Thought," in Loewer & Rey 1991

- Stich, S. 1978. "Autonomous Psychology and the Belief-Desire Thesis," The Monist 61, 573-591
- Stich, S. 1980. "Paying the Price for Methodological Solipsism," Behavioral and Brain Sciences 3, 97-8
- Stich, S. 1983. From Folk Psychology to Cognitive Science: The Case Against Belief (Cambridge, MA: MIT)
- Stich, S. 1991. "Narrow Content Meets Fat Syntax," in Loewer & Rey 1991
- Walker, V. 1990. "In Defense of a Different Taxonomy: Reply to Owens," Philosophical Review 99, 425-431
- Walton, K. 1984. "Transparent Pictures: On the Nature of Photographic Realism," Critical Inquiry 11, 246-277
- Watson, G. 1982. Free Will (Oxford: OUP)
- White, S. 1991. The Unity of the Self (Cambridge, MA: MIT)
- Woodfield, A. 1982. Thought and Object (Oxford: OUP)
- Yablo, S. 1987. "Identity, Essence, and Indiscernibility," Journal of Philosophy 84, 293-314
- Yablo, S. 1992a. "Mental Causation," Philosophical Review 101, 245-280
- Yablo, S. 1992b. "Cause and Essence," Synthese 93, 403-449

---

<sup>1</sup> This paper follows up on "Mental Causation" (Yablo 1992a) ; it takes back the suggestion there that "wide content" events may be inefficacious since "their "narrow" counterparts seem ordinarily to be more commensurate...with their supposed effects." I owe thanks to Louise Antony, Mark Crimmins, Frank Jackson, Joe Levine, Philip Pettit, and Sydney Shoemaker for written comments, Georges Rey for an eye opening conversation, and especially Krista Lawlor for discussions in 1993 by which time she had hit independently on the

---

same basic idea. Too late I discovered Jerry Fodor's recantation in The Elm and the Expert (1994). Section 20 however owes a lot to his pp. 40 ff. Research was done at the National Humanities Center with support from a University of Michigan Humanities Fellowship.

<sup>2</sup> I use "mental causation" for mental causation of physical effects.

<sup>3</sup> Fodor 1989.

<sup>4</sup> See Bedau 1986.

<sup>5</sup> CSM II, 237.

<sup>6</sup> CSM II, 237. Cf. Lucretius: this "reasoning proves the nature of the mind and spirit to be corporeal. For when it is seen to hurl the limbs forward, to snatch the body out of sleep, to alter the face, and to govern and steer the entire man -- and we see that none of these is possible without touch, nor touch without body -- you must surely admit that the mind and spirit are constituted with a corporeal nature" (Long & Sedley 1987, 67).

<sup>7</sup> Along, of course, with the assumption that mental phenomena aren't physical. This recalls another crucial stimulus to contemporary epiphobia, the Putnam/Fodor multiple realization argument. See their papers in Block 1980.

<sup>8</sup> Broad 1925, 104. I should stress that he is not impressed by either argument.

<sup>9</sup> Broad 1925, 100.

<sup>10</sup> P. 133 of Watson 1982

<sup>11</sup> Putnam 1975. Despite our perfect intrinsic similarity, my doppelganger on Twin Earth wants twater, the colorless drinkable stuff in his environs, while it is water that I desire.

---

<sup>12</sup> Dretske 1993, 187, with inessential deletions.

<sup>13</sup> See various papers in Woodfield 1982, especially McGinn's; Loar 1985; Stich 1978, 1980, 1983; Fodor 1980, 1987, 1991a, 1991c; Dretske 1988 and 1993; and various papers in McLaughlin 1991, especially those by Kim and Horgan. Here is Kim's version of the argument: "semantical properties [are] relational, or extrinsic, whereas we expect causative properties involved in behavior production to be nonrelational, intrinsic properties of the organism. If inner states are implicated in behavior causation, it seems that all the causal work is done by their "syntactic" properties, leaving their semantic properties causally idle....How can extrinsic, relational properties be causally efficacious in behavior production?" (1991, 55).

<sup>14</sup> See in particular Kim 1991, Horgan 1991.

<sup>15</sup> There is more on BELOW in Yablo 1992a; there it is called the exclusion argument.

<sup>16</sup> Or at least sufficient for E's objective probability.

<sup>17</sup> E in this paper is always a token event. But C, C\* and so on can be either tokens or types. Words like "state," "phenomenon," "antecedent" and "event" are meant to share in this ambiguity and as far as grammar permits they will be used indifferently either way. (Although see note 47.) I appreciate that some people are scandalized by type/token laxity and I apologize to each and every one of them. The alternative was to run essentially the same argument twice over. See Yablo 1992a,b for a more careful treatment.

---

<sup>18</sup> John Searle says that mental states are "caused by and realized in" physical states of the brain (1983, chapter 10). At times he even seems to suggest that they are caused by and identical to brain states:

if brain processes cause consciousness, then it seems to many people that there must be two different things, brain processes as causes, and conscious states as effects, and this seems to imply dualism. This...mistake derives from a flawed conception of causation (Searle 1995, 60).

Passages like this notwithstanding, Searle agrees that there are "two different things": "the sheer qualitative feel of pain is a very different feature of the brain from the pattern of neuron firings that cause the pain" (ibid, 63). His view is thus type dualism; mental types are caused by, realized in, and distinct from, physical ones. Searle sometimes presents (this version of) type dualism as a solution to the mental causation problem; for many people it is where the problem starts.

<sup>19</sup> Part of the reason that supervenience theories of mind met with such a euphoric response was supervenience's claim to

belong to that class of relations, including causation, that ... represent ways in which objects, properties, facts, events, and the like enter into dependency relationships with one another (Kim 1993, 54).



---

Hence the disillusionment when it sunk in that the standard covariational definitions of supervenience failed to capture any such dependency, and the subsequent insistence that "any physicalist who believes in the reality of the mental must accept pervasive psychophysical property covariance...plus the claim that a dependency relation underlies this covariance" (Kim 1993, 169).

<sup>20</sup> "Animal Automatism" in Huxley 1911, 244; the essay dates from 1874.

<sup>21</sup> I have run this as an argument against type identity but it is effective against token identity as well; see Yablo 1992a. Kripke in Naming and Necessity takes a similar position.

<sup>22</sup> Admittedly, the pain/C<sub>i</sub>: red/scarlet analogy isn't perfect. This doesn't concern me, unless the disanalogies are such as to make pain more causally competitive with C<sub>i</sub> than colors are with their shades. As far as I can see, all that "Y is a determinate of X" needs to mean in this paper is that Y necessitates X (not because it has a metaphysically infallible way of bringing X about but) because X is immanent in or included in Y. This is all it takes to kill the appearance of causal competition. To illustrate with a deliberately farfetched example, suppose that physical states turned out to be conjunctions with mental states as conjuncts. Conjunctions are not in any traditional sense determinates of their conjuncts, but so what? They do determine them in the sense just explained, and that is enough; P&Q can no more preempt P than scarlet can preempt redness.

---

<sup>23</sup> The determinate/determinable story is meant to apply to tokens as well as types; it is not just pain as such but the particular pain I am experiencing right now that can be had in a number of physical ways. Pain stands to its physical determinates in the relation that red bears to scarlet; my particular pain stands to its physical determinates in the relation that something's turning red bears to its turning scarlet. See Yablo 1992a, section 6, for more on token determinates and determinables.

<sup>24</sup> Not least because, for all that has been said so far, Sophie is shade-blind and can't tell crimson from any other sort of redness.

<sup>25</sup> I am not saying that redness inherits causal relevance from scarlet; I am just denying that scarlet can deprive redness of causal relevance.

<sup>26</sup> Details are given in section 15.

<sup>27</sup> There is a considerable tradition of attempting to answer WITHIN by denying this sameness; my Twin, unlike me, would have been reaching for twater. (See the first few papers in Pettit & McDowell 1986, and for criticism Fodor 1991c.) I agree that there is something my Twin does that is different from what I do, and vice versa. But I would hate to pin the case against WITHIN on this, for there is something else we do, viz. simply reaching out, that is the same in both cases. I want to argue that WITHIN is wrong even about the behaviors that my Twin and I have in common.

<sup>28</sup> Yablo 1992a is strangely complacent about this; see the "first remark" in that paper's note 51.

---

<sup>29</sup> If the campaign to purge geology of extrinsic notions has never taken off, the reason is that intrinsically-as-though-sedimentary rocks tend to be, well, sedimentary. No one cares about the counterfactual generality thus gained.

<sup>30</sup> Of course, it matters too that this desire is not outweighed by other desires, that water is available, that its whereabouts are known, and so on; I'll take all that for granted here.

<sup>31</sup> "Oh? Who's to say they don't all have the same sentence of mentalese in their desire box?" I have two responses. First, Fodor has promoted mentalese as providing a non-Fregean explanation of cognitive significance phenomena. This application falls apart if the relation between singular propositional contents and mentalese encodings is not one-many. (Do not say that the relation is one-many just when cognitive significance phenomena force it to be. This suggestion pulls in two directions at once, because a given attitude will engender lots of behaviors only some of which care how exactly the attitude is encoded.) Second, the argument was supposed to show that narrow taxonomy was better, because less generalization-killing, than broad. Now it looks as though narrow taxonomy might be less generalization-killing than broad; it is if a kind of narrow taxonomy can be made out that kills fewer generalizations. Who could argue with that?

<sup>32</sup> Actually, not even (6) is entailed by this fact since I have intrinsic duplicates in a huge (unlimited, in fact) range of external circumstances. Twin Me may be in circumstances like mine but he is very much the exception.

---

<sup>33</sup> This argument is already a bit of a stretch, for a reason hinted at in the last note: water-wanters may well find themselves in circumstances more favorable to drinking behavior than the common run of [water-wanters]. But let me not distract attention from the more serious worry raised in the text.

<sup>34</sup> "Bracketing" makes it sound as though battitudes were stripped down attitudes. This is true in the case of "thick" attitudes (section 16). But I want to leave the door open to "thin" attitudes too subjectively impoverished for bracketing so understood to yield anything worthwhile.

<sup>35</sup> That is, some possible doppelganger of yours has a belief with the singular propositional content that  $p$ . See Walker 1990 and Stich 1991. I am indebted to Brown 1993.

<sup>36</sup> This is sloppy but not I think in a way that matters.

<sup>37</sup> They are certainly distinct in the worlds where they exercise this capability; and if there, then everywhere, for duplicates are battitudinally indiscernible.

<sup>38</sup> See the first two chapters of White 1991 (one of which dates back to 1982) and Fodor 1987.

<sup>39</sup> This papers over a real problem, namely, how to decide which of the propositions believed by my doppelganger in  $w$  gets to counts as  $p(w)$  --  $p(w)$  being the proposition your doppelganger had better also believe in  $w$  if she wants to share my [belief that  $p$ ].

<sup>40</sup> Or as many as makes sense; one doesn't have doppelgangers in every world. I'm going to ignore this problem.

<sup>41</sup> Here is Fodor:

---

what I use to manipulate the correlation between my elm thoughts and elms is not an instrument but a botanist. To do that sort of thing, I must be able to pursue policies with respect to another person's mind as well as my own. And also with respect to the causal relations between our minds. I am relying on its being reliable that elms will cause the botanist to have elm thoughts; which in turn will cause him to utter elm reports; which in turn will cause me to believe that it is an elm I have to do with. Setting things up so that all this is reliable requires that I be very clever, that I know a lot (for example, I have to know which experts I can trust) and that I be prepared to pay what a botanist's services cost. But it is likely to be worth the trouble (1994, 36).

Fodor says that I have to know which experts I can trust. Had my elm thoughts been under the control of different experts, they might have been correlated in the content-determining way with a different kind of tree. But since who I trust about elm is a function of collateral [attitudes], so is elm's meaning in my neurolect.

<sup>42</sup> Compare Block 1991's arguments that narrow content is holistic and Fodor's response in the same volume. Fodor thinks that Block has confused two questions, namely "what fixes the propositional content of mentalese token "X" in a given context?" (answer: nomic relations with the outside world, regardless of relations with collateral mental items) and "what in my mental life helps to sustain "X" in the relevant nomic relations?" (answer: relations with collateral mental items among other things):

---

the N-relation ["the nomological relation N such that your "X" tokens refers to Xs...iff they bear N to Xs"] is...robust; many theories ... might succeed in sustaining the N-relation between 'dogs' and dogs in this world, 'dogs' and twin-dogs in Twin-world, 'dogs' and things-just-like-our-dogs-except-for-the-ears in Cousin-world... and so forth. So, many different belief systems might implement the narrow content DOG. Or, if this is not right, Block needs an argument to show that it isn't. And, so far, I don't see that he's got one (266).

The argument I would propose on Block's behalf is that while different belief systems will indeed implement the same N-relations in some worlds, the "and so forth" is unwarranted. Because the "and so forth" says in effect that differences in surrounding theory are necessarily (across all possible worlds ) incapable of bearing on what "dog" is N-related to. And it is not clear how both of the following can be true together: first, surrounding theory helps to sustain "dog" in its N-relations, but second, tweaking surrounding theory not only does not but cannot affect those N-relations. (Again, a change of N-relations in any world entails a change of narrow content here.)

<sup>43</sup> "Subjective" in the sense of "intrinsic to the subject." This is to allow for syntactical and/or neural battitudes. Note that we might want to relativize to other parameters as well; the same person may judge the same proposition in different intrinsic ways at different times, or even at the same time via different mental representations.

---

<sup>44</sup> Compare Ruth Millikan: "Jerry Fodor has been considerably exercised (as he likes to say) by the (undoubted) fact that, knowing only that it is true of the girl next door that John wants to meet her, we cannot predict that John will exhibit next-door-directed behavior. For John may believe that this girl whom he wishes to meet languishes in Latvia...But a very straightforward (though extremely fallible) surmise still follows immediately from the fact that John desires to meet Jane...and from this fact alone. Namely, eventually John will meet Jane (say, after he gets back from Latvia)" (1993, 69).

<sup>45</sup> Cf. Walton's claim that to see a person's photograph is, or can be, to see the person (Walton 1984). His concern is whether Isaac sees bubble; mine is why he sees, or seems to see, her.

<sup>46</sup> To say it a little more clearly: From (i) no change in behavior without a change in attitude, that is, a purely attitudinal change won't do it, and (ii) a purely attitudinal change will do it, it does not follow that (iii) attitudes cannot screen attitudes off. (ii) is irrelevant to (iii) since whether my desire screens my [desire] off turns on the results of holding my desire fixed while varying my [desire] -- not the other way around as in (ii). And between (i) and (iii) there is a palpable gap; what (i) says can happen through pure variation in [desire], (iii) says would happen were there pure variation in [desire].

<sup>47</sup> I have framed these definitions, and most of the subsequent discussion, with token causation in mind. For the application to types, think of C, with or without subscripts, as qualifying an

---

implicitly given token cause X, and substitute "X has C" for "C occurs." E remains a token effect.

<sup>48</sup> I used to say that C was enough for E iff no determinate of C was required for E (Yablo 1992a,b). This was a weaker reading of enoughness since determinates of C had only to be screened off by some determinable or other, not necessarily by C itself. I now prefer the definition in the text.

<sup>49</sup> Yablo 1992a and 1992b put two further conditions on proportionality which can be ignored here. E is contingent on C iff it would not have occurred if C had not occurred; and C is adequate for E iff had C not occurred, E would have occurred if it had.

<sup>50</sup> These are different but not in a way that matters here.

<sup>51</sup> "Attitude" in this paper has generally meant "thin attitude" and it will continue to do so. Note that on the "thick" reading it would be (trivially) false to say that my desire for water is more widely shared than my [desire for water].

<sup>52</sup> Nor does [A] determine it; the two are just incomparable, like the property of being a photo of Isaac's bubbe and the property of being a photo with such and such intrinsic color features.

<sup>53</sup> Demonstrably so, since [S] screens off both if either. Proof: Because thin S's differences from thick lie entirely within [S], [S] occurs without the one in the same worlds as it occurs without the other. But then E inhabits the nearest world containing [S] without thick S iff it inhabits the nearest world containing [S] without thin S.



---

<sup>54</sup> "Nothing" here means "no state or event actually in existence."

The mere fact that there could have been a state or event that, had it existed, would have screened C off does not prevent C from being superrequired.

<sup>55</sup> "[T]here must be some finite lapse of time...between cause and effect. This, however, at once raises insuperable difficulties.

However short we make the interval ... something may happen during this interval which prevents the expected result. In order to be sure of the expected result, we must know that there is nothing in the environment to interfere with it. But this means that the supposed cause is not, by itself, adequate to insure the effect" (Russell 1917, 136-7). One could easily read the passage as comparing C not to later events but a larger contemporary: C taken together with its simultaneous environment.

<sup>56</sup> "[I]f the cause is a process involving change within itself, we shall require...causal relations between its earlier and later parts; moreover it would seem that only the later parts can be relevant to the effect...Thus we shall be led to diminish the duration of the cause without limit, and however much we may diminish it, there will still remain an earlier part, which might be altered without altering the effect, so that the true cause...will not have been reached" (Russell 1917, 135).

<sup>57</sup> That is, cause and effect must begin at the same time. Similar reasoning suggests that they must end at the same time as well.

---

<sup>58</sup> The "proportionality" principle laid down in the last paragraph of Yablo 1987 amounted to (15) restricted to coincidents; I hereby withdraw it.

<sup>59</sup> Except of course when one of the coincidents is a determinate of the other; then we are back to simple proportionality. The relation between (token) determination and coincidence is this.  $\underline{D}$  is a determinate of  $\underline{C}$  iff (i)  $\underline{C}$  inhabits every world that  $\underline{D}$  does, and (ii) wherever both exist, they are coincident. Details can be found in Yablo 1987, 1992a,b.

<sup>60</sup> To avoid any appearance of scope confusion, the claim is this. Where  $\underline{C}_1$  = her guessing the little man's name,  $\underline{C}_2$  = her guessing "Rumpelstiltskin," and  $\underline{C}_3$  = her guessing his deepest secret, had  $\underline{C}_1$  occurred without  $\underline{C}_2$ , or  $\underline{C}_2$  without  $\underline{C}_3$ , or  $\underline{C}_3$  without  $\underline{C}_1$ , the effect  $\underline{E}$  would still have occurred. That  $\underline{C}_1$  can occur without  $\underline{C}_2$  (etc.) shows that we have not one event here but three, albeit three coincident events (see Yablo 1987 and 1992b).

<sup>61</sup> Assuming, that is, that (i) elimination at the hands of one candidate cause is independent of elimination at the hands of another, and that (ii) one candidate cause's being eliminated is independent of another's being eliminated. (ii) is not strictly true since the hypothesis that  $\underline{C}_i$  is eliminated raises the chances that it was eliminated by  $\underline{C}_j$ , which lowers the chances that  $\underline{C}_i$  eliminates  $\underline{C}_j$  in return. (If  $\underline{C}_j$  eliminated  $\underline{C}_i$  by screening it off, then  $\underline{C}_i$  cannot eliminate  $\underline{C}_j$  by failing to be screened off by it, and vice versa.) The formula in the text is close enough to the truth not to matter.

---

<sup>62</sup> If the power of elimination is reserved to  $\underline{C}_i$ 's determinates and determinables, chances of self-destruction are zero until the number of candidate causes hits four. And self-destruction will always be rare, because of the following fact. Using  $<$  for the is-a-determinable-of relation, and letting a zigzag be a sequence of  $\underline{C}_i$ s such that  $\underline{C}_1 < \underline{C}_2 > \underline{C}_3 < \underline{C}_4 > \dots$ , a set of candidate causes self-destructs only if each of its members is connectable by a zigzag to a circular zigzag of cardinality four or more.

<sup>63</sup> Ordinary proportionality raises similar problems (Yablo 1992b, section 11), but not on anything like the same scale. Technically this is because the chances of finding a  $\underline{C}_j$  screening  $\underline{C}_i$  off (a  $\underline{C}_j$  that  $\underline{C}_i$  fails to screen off) are greatly reduced if we require  $\underline{C}_j$  to exist in all (only) the worlds that  $\underline{C}_i$  exists in. Intuitively it is because a determinate of  $\underline{C}_i$  that screens it off (a determinable of  $\underline{C}_i$  that it fails to screen off) is prima facie a better candidate than  $\underline{C}_i$  for the role of cause. Superproportionality allows  $\underline{C}_i$  to be killed off by its causal inferiors; proportionality keeps  $\underline{C}_i$  alive until something better turns up.

<sup>64</sup> This would be an event with the existence-conditions that something has happened given which something has happened given which  $\underline{E}$ 's causal needs are met.

<sup>65</sup> Compare Yablo 1992b, section 11. There may be room for a screening-off-type condition stronger than proportionality but weaker than superproportionality. Several people have suggested the following:  $\underline{C}$  causes  $\underline{E}$  only if  $\underline{C}$  is not screened off asymmetrically, that is, by anything that it does not screen off in

---

return. This is helpful when the candidate causes number exactly two but when  $\underline{n} = 3$  or more it becomes possible for each  $\underline{C}_i$  to asymmetrically screen off one of its fellows while being asymmetrically screened off by another. The Rumpelstiltskin example is (or can be made to be) a case in point.

<sup>66</sup> Assuming that stripping it away yields an entity at all -- there are questions here about the limits of determinability.

<sup>67</sup> Note that as long as neither determines the other,  $\underline{C}_i$  and  $\underline{C}_j$  can both be proportional to a given effect. (Which is good; see the discussion in Yablo 1992b, section 12, of "world-driven" and "effect-driven" causes.) Superproportionality by contrast leaves at most one candidate cause standing. (Suppose for contradiction that  $\underline{C}_i$  and  $\underline{C}_j$  are both superproportional with  $\underline{E}$ . Either  $\underline{C}_i$  screens  $\underline{C}_j$  off from  $\underline{E}$ , or it doesn't. In the first case,  $\underline{C}_j$  is not superrequired; in the second,  $\underline{C}_i$  is not superenough.)

<sup>68</sup> This is a play on the last paragraph of Davidson's "On Saying That": "Since Frege, philosophers have become hardened to the idea that content-sentences in talk about propositional attitudes may strangely refer to such entities as intensions, propositions, sentences, utterances, and inscriptions. ... If we could recover our pre-Fregean semantic innocence, I think it would seem to us plainly incredible that the words 'The Earth moves,' uttered after the words 'Galileo said that,' mean anything different, or refer to anything else, than is their wont when they come in other environments" (1984, 108).

---

<sup>69</sup> If I said that this followed from the fact that attitudes screen battitudes off, I would be guilty of the very thing I've been warning against: drawing an asymmetrical conclusion from symmetrical premises. (Battitudes screen attitudes off as well.) I can't appeal to proportionality considerations either, for there is nothing in the definition of proportionality -- as opposed to superproportionality - - to prevent two candidate causes' both being proportional to an effect, provided that they are incomparable with respect to determination. (See the second last note, and Yablo 1992a, section 12.) Then why give the nod to the attitudes? Partly for shock value; partly because of skepticism about the battitudes; partly because of the rationality argument to follow; and partly because of a hard to defend intuitive feeling that that is the way the tradeoffs play out -- on the whole and for the most part, you can buy more intrinsic detail with a fixed amount of extrinsic (truth-conditional) detail than the other way around.

<sup>70</sup> I assume that "getting what you want" at least involves the referential content of your desire coming true.

<sup>71</sup> "But sensitivity to pure variation in [attitude] can work to our advantage. Julie's [desire to be with Jeckyll] and [desire not to be with Hyde] combine to put her in Jeckyll's company in his high-functioning periods while keeping her out of his clutches when he goes into monster mode. Hasn't she gotten what she wanted?" Apparently so. This kind of case needs further discussion. (Thanks here to Mark Crimmins.)

<sup>72</sup> Using "subjective," as above, for "intrinsic to the subject."

---

<sup>73</sup> And the orthography, and the neurology.