List and Menzies on High-Level Causation

Jens Jäger

Abstract

I raise two objections against Christian List and Peter Menzies'
influential account of high-level causation. Improving upon some of
Stephen Yablo's earlier work, I develop an alternative theory which
evades both objections. The discussion calls into question List and
Menzies' main contention, namely that the exclusion principle, applied
to difference-making, is false.

1 Introduction

At first pass, it seems obvious that properties studied by the special sciences
can be causes. If you ask virologists what caused recent spikes in viral
infections, they might point out (among other things) that as the weather
gets colder people tend to stay indoors for longer, where viruses spread more
easily. If you ask economists about the causes of the 2007 financial crisis,
they might cite the prevalence of subprime mortgages in the run-up to the
crisis. Here, the cited causes are high-level properties, concerning (in the
first case) human behavioral tendencies and (in the second case) properties
of financial markets.

At second pass, however, the claim that special science properties are
causes is in tension with three common theses about the metaphysics of these
properties: (1) that special-science properties supervene on microphysical
properties;[1] (2) that special-science properties are nonetheless not identical to
(since multiply realized by) microphysical properties; and (3) a thesis about
the causal powers of the microphysical: that every physical property which is
caused at all is caused by a microphysical property.

The tension arises as a result of exclusion arguments, prominently defended in various forms by Jaegwon Kim (2000; 2005). The centerpiece of these arguments is the 'exclusion principle', which Kim holds to be 'virtually an analytic truth' (Kim, 2005, p. 51). One possible formulation of the principle, adapted from List and Menzies (2009), is as follows:

> EXCLUSION: 'For all distinct properties F and F^* such that $[F$ necessitates $F^*]$, F and F^* do not both cause a property G.' (List and Menzies, 2009, p. 490)[2]

To see the tension, suppose that a special-science property S is a cause of some physical property. By (3), there's a microphysical property, P, causing the physical property. By (1), S is metaphysically necessitated by some microphysical property; very plausibly, that property is P itself. Finally, by (2), P is distinct from S. So, both a microphysical property and a distinct special-science property it necessitates are causes of the same effect. This conflicts with EXCLUSION. So S isn't a cause of any physical property.

In two papers, Christian List and Peter Menzies (2009; 2010) (henceforth "LM") set out to defend the causal efficacy of special science properties. They propose to interpret "cause" in terms of a relation of difference-making, and claim that this interpretation can salvage the causal autonomy of special-science properties. Sometimes, high-level properties make a difference to the occurrence of an effect, without any low-level property making a difference. Hence condition (3), regarding the causal power of the microphysical, is false: Some physical effects don't have microphysical difference-makers. Further, according to LM, sometimes *both* the high-level property *and* the low-level property are difference-makers. EXCLUSION is false too.

The basic idea behind LM's account is not new: LM acknowledge that the concept of difference-making is closely related to the idea that causes must be *proportional* to their effects; an idea which they attribute to Stephen Yablo (1992). Roughly, causal proportionality says that causes have to contain *just the right amount of detail* given the effect. They should neither incorporate overly specific information that's causally irrelevant, nor omit detail that's causally significant.[3]

This paper identifies two problems for LM's account. The centerpiece of their account is a theory of the truth conditions of claims about difference-making. The truth conditions involve counterfactuals, for which LM supply a nested-sphere semantics similar to the one proposed by David Lewis (1973). The first problem is that LM's truth conditions are often insufficiently sensitive: They qualify too many properties as difference-making causes. The second problem arises from a deviation from Lewis's semantics LM implement. The deviation, I'll argue, introduces an additional unexplained free parameter into their semantics, diminishing the account's explanatory power.

I'll then develop a different implementation of the proportionality intuition, in the spirit of Yablo (1992). The implementation solves both problems. Further, it still identifies proportional causes which consist of special science properties, thereby recovering special sciences' causal autonomy. *Contra* LM, however, the account entails EXCLUSION. (The only thesis it denies is (3); that is, it confirms that some physical properties have special-science causes without having microphysical causes.)[4,5]

2   The Account

Recall Lewis's semantics for counterfactuals (Lewis, 1973): The semantic model consists of an assignment to each world of a system of centered, nested spheres. The system is supposed to convey information about the comparative similarity of possible worlds in the following sense: If some possible world i lies outside of a given sphere S centered around world w, all possible worlds within S are more similar to w than i is to w. Simplifying Lewis's account somewhat, require that any system (set) $\boldsymbol{S}_w$ of spheres centered around w satisfy the following constraints. Where P is any proposition, call any sphere which has non-empty overlap with P a "P-permitting sphere":[6]

- *Nestedness*: For all spheres $S, T \in \boldsymbol{S}_w$, either $S \subseteq T$ or $T \subseteq S$.

- *Strong Centering*: $\{w\} \in \boldsymbol{S}_w$, i.e. the smallest sphere containing w is the sphere that contains only w.

- *Exhaustiveness*: $\mathcal{W} \in \boldsymbol{S}_w$, i.e. the system of spheres contains a largest sphere consisting of all possible worlds.

- *Limit Assumption*: For any non-empty proposition P, there is a *smallest* P-permitting sphere around w. In symbols: For any non-empty P, define $\min_P(w) := \bigcap \{S \in \mathbf{S}_w \mid S \cap P \neq \varnothing\}$. Then $\min_P(w) \in \boldsymbol{S}_w$ and $\min_P(w) \cap P \neq \varnothing$ (that is, $\min_P(w)$ is the smallest P-permitting sphere around w).

The truth conditions for counterfactual conditionals are then as follows. For any proposition X, call any possible world satisfying X an X-world. Then,

> **(CF)** $P \, \square\!\!\rightarrow Q$—read "If it was the case that P, it would be the case that Q."—is true at w if and only if every P-world in the smallest P-permitting sphere around w is a Q-world. In symbols: $P \, \square\!\!\rightarrow Q$ is true at w iff $P \cap \min_P(w) \subseteq Q$. (Cf. Lewis, 1973, p. 20)

In the following, I'll call the intersection $P \cap \min_P(w)$ the "closest P-worlds to w". In more casual terms, a counterfactual is true at w if and only if its consequent is true in all worlds which satisfy the antecedent while remaining 'as similar as possible' to w.

LM provide necessary and sufficient conditions for difference-making in terms of two counterfactuals (List and Menzies (2009, p. 483), adapted for notation; see my fn. 7 below):

> **(DM):** x makes a difference to y in the actual world if and only if actually (i) $x \, \square\!\!\rightarrow y$ and (ii) $(\neg x) \, \square\!\!\rightarrow \neg y$.

Here x and y quantify over *events*, and I've adopted the following short-hand: Inside counterfactuals, "x" stands for "x occurs", and analogously for "y" (and "z", introduced later). E.g. "$\neg x$" abbreviates "$\neg(x$ occurs)"; etc. So "$(\neg x) \, \square\!\!\rightarrow \neg y$" is to be read "If x hadn't occurred, y wouldn't have occurred". The same holds for the other logical connectives; e.g. "$x \wedge y$" abbreviates "(x occurs) $\wedge$ (y occurs)". Where x is an event, I'll call any world where x occurs an x-world.[7]

The truth conditions for LM's counterfactuals (i) and (ii) are as above, with one crucial exception: LM replace Lewis's *strong centering* with

- *Weak Centering*: For all $S \in \boldsymbol{S}_w$, $w \in S$.

This replacement allows the smallest sphere around w to contain additional worlds besides w.

The modification is required since otherwise condition (i) of (DM) wouldn't do its proper work. It would automatically be true for any actually occurring x and y: If x and y actually occur, then, by *strong centering*, y occurs in all closest x-worlds (namely, the actual world). (That is, strong centering implies "And-to-If": "$x \mathbin{\square\!\rightarrow} y$" is true if antecedent and consequent are both true.) LM want to avoid this. Replacing *strong centering* by *weak centering* achieves this, intuitively, by expanding the set of closest x-worlds, such that a consequent may fail to hold at some closest-antecedent worlds despite being true at the actual world.

Let's illustrate LM's account with an example (List and Menzies, 2010): Suppose you're waving for a taxi. Consider two candidate difference-makers for your action: a high-level mental state, your *intention* to wave for a taxi; and the state's microphysical realizer, the exact *neural state n*. Which one made a difference to your waving? Your mental state plausibly satisfies condition (ii): If you hadn't intended to wave for a taxi, you wouldn't have waved. By contrast, LM claim, the microphysical realizer violates condition (ii): If you hadn't been in the exact neural state n, you would have been in a slightly different neural state n' which *still* realizes your intention to wave for a taxi. So you'd still wave. If that's right, we have an instance of what LM call 'downwards exclusion': Some property is a cause, while one of its low-level realizers is not.

There's also the opposite phenomenon, 'upwards exclusion': Some property is a cause, while some *higher-level* property is not. On LM's account, condition (i) is typically responsible for upwards exclusion. To illustrate, consider a third candidate difference-maker: your intention to wave (*simpliciter*). (What I mean here is [your intention to wave] *simpliciter*, not your intention to [wave *simpliciter*]. When you intend to wave for a taxi, necessarily you

instantiate the former, but you may lack the latter.) Suppose condition (i) is satisfied by your intention to wave for a taxi: In all closest worlds in which you intend to wave for a taxi, you indeed wave for a taxi. Your intention to wave *simpliciter*, however, may violate condition (i): There may be some closest world in which you intend to wave, but do not intend to wave for a taxi. (Instead, you might intend to wave at a friend across the street, or at an airplane passing overhead.) (Note the crucial role of strong centering failure in this reasoning.) If so, your intention to wave *simpliciter* violates condition (i). For presumably when you intend to wave in some alternate way, you do not wave for a taxi. Hence your intention to wave *simpliciter* isn't a difference-maker according to (DM). So we have an instance of upwards exclusion.

Now for the problems of LM's account.

3   First Problem: Missing Downwards Exclusion

The first problem is that condition (ii) doesn't achieve downwards exclusion in all relevant cases. Consider the following variant of a well-known case from Yablo (1992):

> **(Sophie)** Sophie is a pigeon trained to peck at all and only red objects. Scientists place two objects, a crimson disk and a blue disk, into a box. They pull out one disk at random and place it in front of Sophie. They happen to pick the crimson disk, and Sophie pecks.

The disk's being crimson seems superfluous for Sophie's pecking in the same way in which one's neural state seems superfluous for one's waving for a taxi. An account of proportionality plausibly shouldn't draw the line between crimson in (Sophie) and my neural state in the taxi case. Yet this is what LM's account does. Let "RED" and "CRIMSON" abbreviate "the target's being red" and "the target's being crimson", respectively. Condition (ii) excludes CRIMSON iff, had the target not been crimson, Sophie *might still* have pecked. But this is implausible in the current setup. The only other

available target is blue. Surely, then, had the target not been crimson, it would have been blue. So Sophie wouldn't have pecked. So CRIMSON satisfies condition (ii). (And of course, since RED satisfies condition (i), and RED entails CRIMSON, CRIMSON also satisfies condition (i).)

More formally, let "PECK" abbreviate "Sophie's pecking". To include RED and exclude CRIMSON we need both

(ii-RED) $\neg\text{RED} \,\square\!\!\rightarrow\, \neg\text{PECK}$, and

($\neg$ii-CRI) $\neg(\neg\text{CRIMSON} \,\square\!\!\rightarrow\, \neg\text{PECK})$.

But in the current setup we have

(SETUP) $\neg\text{CRIMSON} \,\square\!\!\rightarrow\, \neg\text{RED}$.

Since CRIMSON entails RED, (ii-RED) is equivalent to

(ii-RED*) $(\neg\text{CRIMSON}) \wedge (\neg\text{RED}) \,\square\!\!\rightarrow\, \neg\text{PECK}$.

But LM's semantics (like Lewis's) validates the schema ("Cumulative Transitivity"):

$$A \,\square\!\!\rightarrow\, B,\ (A \wedge B) \,\square\!\!\rightarrow\, C \models A \,\square\!\!\rightarrow\, C.$$

Hence, (SETUP) and (ii-RED*) entail

(ii-CRI) $\neg\text{CRIMSON} \,\square\!\!\rightarrow\, \neg\text{PECK}$,

contradicting ($\neg$ii-CRI). So, provided that LM's account counts RED as a difference-maker, it counts CRIMSON as a difference-maker too. And that's the wrong result.

Intuitively, the general problem is that LM's condition (ii) excludes a low-level realizer only if the presence of the high-level difference-maker is "counterfactually robust" against changes in the low-level realizer (i.e., in the example: only if (SETUP) is false). But, as (Sophie) shows, counterfactual robustness doesn't always track difference-making; sometimes even a fragile difference-maker excludes its realizer.

Yablo (1992) makes a similar point. He first asks us to imagine that Socrates *guzzled* the deadly hemlock rather than drinking it slowly. Guzzling

7

the hemlock doesn't make a difference to Socrates's death—drinking it does. LM's counterfactual criterion concurs: Presumably, had he not guzzled the hemlock, Socrates would still have consumed the hemlock (drinking it in an orderly manner). But now consider the following variant, where death *does* counterfactually depend on guzzling:

> **(Sloppy Socrates)** 'Imagine that Socrates, always a sloppy eater, had difficulty drinking without guzzling, to such a degree that if the guzzling hadn't occurred, the drinking wouldn't have either. Then Socrates' death *was* contingent on his guzzling the hemlock ... Intuitively [however], it appears that not *all* of the guzzling was needed, because there occurred also a lesser event, the drinking, which would still have done the job even in the guzzling's absence.' (1992, p. 276)

Condition (ii) goes astray by picking up on additional features of the case, Socrates's habits, which are irrelevant to the guzzling's proportionality.

3.1 A Fix: Requiredness

Notwithstanding its problem, condition (ii) seems to capture *something* right. At least in the sort of simple systems we're dealing with (cf. fn. 4), an effect should be *contingent* on the difference-maker. This is well captured by condition (ii) (which, following Yablo, I'll call **Contingency**). But the previous discussion suggests that even antecedents which satisfy Contingency can contain excessive detail. So we need a better criterion to capture "excessive detail".

Yablo (1992) offers some guidance. Let x be some event, and consider all events necessitated by x. Now suppose that any of those high-level events occurs *without x*. Then, following Yablo, say that x's occurrence is *required* for an effect iff, in all those cases, the effect would have been absent. Intuitively, x's being required means that the effect wouldn't occur even if a very similar event was instantiated ("very similar", in the sense that the event's occurrence necessitates many of the same high-level events as x does). Formally, define:[8]

> **Yablo-Requiredness:** x is Yablo-required for y iff, for all z whose occurrence is necessitated by x's occurrence, $(\neg x) \land z \; \Box\!\!\rightarrow \neg y$.

For example, had the target not been crimson but still red, the pigeon would still have pecked. Hence CRIMSON is *not* (Yablo-)required for PECK.[9] The idea is then to add the following condition to LM's account (DM):

(iii) x is *Yablo-required* for y.

Is this satisfactory? It seems to do a good job with (Sophie), and handles related cases like (Sloppy Socrates) analogously. It seems to constitute a significant improvement over LM's demand for mere Contingency.

But there's trouble: The criterion mishandles disjunctive events. To illustrate, let WORM be the target's having a worm attached to it, and consider the occurrence of (RED ∨ WORM). This occurrence is necessitated by RED's occurrence. But it is not the case that ($\neg$RED ∧ (RED ∨ WORM)) $\Box\!\!\rightarrow$ $\neg$PECK, for the antecedent logically entails WORM's occurrence, and Sophie (let's suppose) always pecks at a worm. So, according to the above criterion, RED is *not* (Yablo-)required for PECK; it is "screened off" by (RED ∨ WORM).[10] It follows from (iii) that RED doesn't make a difference to PECK. (McGrath (1998) raises the same issue.) Indeed, any event that's "robustly sufficient" for PECK spoils the soup: If its occurrence suffices for PECK even in RED's absence, then disjoining it with RED yields an event that screens off RED. (For another example, consider (RED ∨ SPASM), where SPASM is the target's being such that Sophie has a muscle spasm, making her peck at the target.)[11,12]

We can fix this bug by stipulating that being screened off by *badly disjunctive* events does not detract from requiredness:

Requiredness: x is required for y iff, for all *not badly disjunctive* z whose occurrence is necessitated by x's occurrence, $((\neg x) \wedge z)$ $\Box\!\!\rightarrow \neg y$.

The practice of excluding badly disjunctive events is familiar from counterfactual accounts of causation. For instance, Lewis (1986b, p. 267) wants to avoid that, in a situation in which Fred's talking causes Ted's laughing, Fred's talking-or-walking causes Ted's laughing as well. He achieves this by barring bad disjunctions of events from qualifying as events altogether. We may

implement that ban too. If we do, the restriction to "not-badly-disjunctive" z's follows for free, and Requiredness and Yablo-Requiredness are equivalent. For explicitness, I'll continue to use "events" in the more permissive sense.

(For the purpose of this essay, I'm intending to stay largely neutral on how "bad disjunction" is best defined. But for concreteness, here is one proposal.[13] Langton and Lewis (1998) and Lewis (2001) define a *property's* bad disjunctiveness via a notion of comparative naturalness for properties. Analogously, we may define an *event's* bad disjunctiveness via a notion of comparative naturalness for events.[14] Then say that an event x is badly disjunctive iff, for some of x's disjunctive expansions, all of the disjuncts are much more natural than x itself.[15])

McGrath (1998) considers, but rejects, a similar-sounding response on Yablo's behalf:

> '[Yablo] could rule out disjunctive properties as determinables of their disjuncts. The problem is that this may well be the best justification for counting pain as a determinable [of] its underlying brain states'. (171)

However, McGrath's proposal does not distinguish between *good* and *bad* disjunctions. Perhaps non-reductive physicalists, like Yablo and LM, should agree that the property of being in pain is a disjunction of properties of the form *being in x*, where x is a neural state. But they shouldn't agree that the disjunction is *bad* (e.g. that it is much less natural than all of its disjuncts). In fact, mental properties seem like paradigm examples of good disjunctions. (The property of being in pain, the property of having a red sensation, and so on, don't seem less natural than some of the gazillion detailed neural states which possibly realize them—arguably, they even seem *more* natural than *any* of them.) So McGrath's critique doesn't threaten our account.

With the notion of Requiredness in place, we should supplement (DM) instead with:

(iii*) x is *required* for y.

Condition (iii*) is satisfied by RED, though also by RED ∨ WORM and RED ∨ SPASM. (We'll add an additional criterion to (DM) in section 4.)

3.2 Dropping Contingency

Requiredness turns out to entail Contingency, thus making an explicit demand for it redundant. Consider the class of necessarily occurring events—events which occur in any world whatsoever. Some of these events are intuitively highly non-disjunctive. For example, take the event which, on Lewis's (1986) formalism, is the set of all possible worlds. (The event may be essentially describable as "the world's existing", or "the world's being identical to something".) It is intuitively highly non-disjunctive: Disjunctiveness measures how "gerrymandered" an event is; and the all-encompassing event is intuitively highly non-gerrymandered. (Indeed, literal gerrymandering is impossible in US states with a single congressional district, such as Wyoming or Alaska.) (The Langton-Lewis account agrees: The set of all possible worlds is the necessary proposition, and so it is definable by any simple tautology, which is highly natural—so natural that plausibly at least some disjunct is always less natural than it.)

Let now x be an event required for y, and let $\top$ be the all-encompassing event. x's occurrence necessitates $\top$'s occurrence, and so Requiredness implies $((\neg x) \wedge \top \;\square\!\!\rightarrow \neg y)$. Since $\top$ necessarily occurs, this implies $(\neg x) \;\square\!\!\rightarrow \neg y$. So a demand for Requiredness includes a demand for Contingency.[16]

On to the second problem for LM.

4 Second Problem: Weak Centering

4.1 Difference-Making Is Highly Sensitive to the Size of the Innermost Sphere

My intending to wave for a taxi is proportional to my waving for a taxi, though my intending to wave is not. This is upwards exclusion, and LM's framework enables it (mostly[17]) via condition (i), combined with strong centering failure. To recall, on LM's semantics, my intending to wave *simpliciter*—abbreviated "INTEND-S"—is supposed to violate condition (i): $\neg(\text{INTEND-S} \;\square\!\!\rightarrow \text{TAXI})$.

The trouble with LM's semantics is that what events violate (i) depends very sensitively on the size of the innermost sphere. (Here and throughout, by

"innermost sphere" I mean "innermost *non-empty* sphere".) Consider figure 1. It depicts three possible systems of spheres, centered around the actual world w, where I wave for a taxi. (INTEND is my intending to wave for a taxi. In a slight abuse of notation, I also use INTEND to denote the *proposition* that I intend to wave for a taxi. Analogously for INTEND-S, etc.) In each figure, the inner circle/ellipse indicates the innermost sphere around w, and the two outer circles indicate the second-closest and third-closest spheres to w, respectively. The shaded oval region T around w is the set of all relevantly close worlds where I wave for a taxi. ("Relevantly close", since there may be other worlds where I wave for a taxi, but they lie far enough away to be safely ignored.) T is a proper subset of INTEND, since close to the actual world I end up waving for a taxi only if I also *intend* to wave for a taxi. (E.g., involuntary tics are far removed.) INTEND, meanwhile, is a proper subset of INTEND-S: Necessarily, I intend to wave for a taxi only if I intend to wave *simpliciter*.

Figures 1a-1c vary only in the size of the innermost sphere. But those small variations make a big difference, given (DM).

- Figure 1a: Both INTEND and INTEND-S are difference-making causes of T.[18]

- Figure 1b: With the innermost sphere now extending outside of the T-worlds, INTEND-S is no longer a difference-maker. It violates condition (i), since $\neg$(INTEND-S $\Box\!\!\rightarrow$ T). (In some relevantly similar situations in which I intend to wave, I don't wave for a taxi.) INTEND remains a difference-maker.

- Figure 1c: The innermost sphere is now large enough to contain even some *INTEND-worlds* that are $\neg$T-worlds. As a result, INTEND too violates condition (i): $\neg$(INTEND $\Box\!\!\rightarrow$ T). (In some relevantly similar situations in which I intend to wave for a taxi, I don't wave for a taxi—perhaps I experience a sudden weakness, or am distracted by a friend.) Neither INTEND nor INTEND-S are difference-makers.
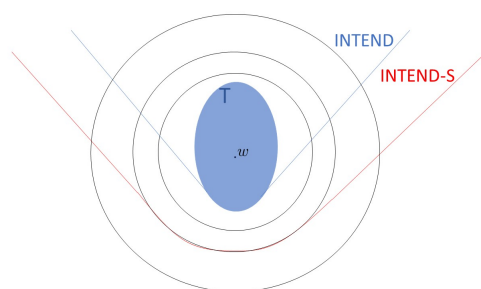
This illustrates the issue: Upwards exclusion occurs when high-level events

INTEND

INTEND-S

T

.w

(a)

INTEND

INTEND-S

T

.w

(b)

INTEND

INTEND-S

T

.w

(c)

Figure 1

13
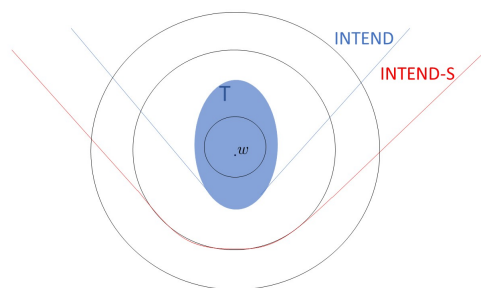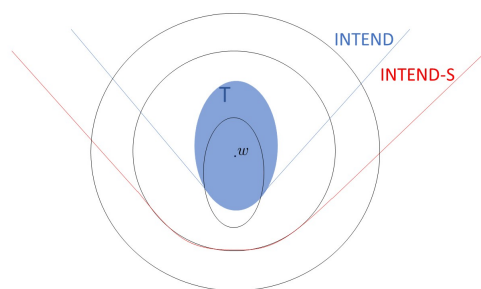
violate condition (i),[19] but whether events violate condition (i) sensitively depends on the size of the innermost sphere. More concretely, we can prove that, where x necessitates z, x satisfies (i) while z violates (i) *only if the innermost sphere contains some $\neg x$-worlds.*[20] Figures 1a and 1b witness this. Figure 1c, meanwhile, is a case where an event on *an even lower level* than INTEND excludes INTEND and INTEND-S.

4.2 The Size of the Innermost Sphere Is Unexplained

The sensitivity to the size of the innermost sphere is a problem, given that, as I'll now argue, the parameter remains wholly unexplained.

To start, it is well known that similarity metrics for counterfactual semantics don't simply match our naive pretheoretical judgments about "differences between worlds". As Lewis (1979, p. 466) points out, the truth of certain plausible counterfactuals requires that 'the similarity relation ... disagrees with ... explicit judgments of what is "very different"' (List and Menzies (2009) recapitulate this point).[21] So we can't just rely on our pretheoretic judgments about similarity to fix the size of the innermost sphere.

One might think we could instead rely on our pretheoretic judgments about *counterfactuals*. Take any truth which holds throughout the entire innermost sphere—some tautology $\top$, say. The innermost sphere is the intersection of all P such that $\top \: \square\!\!\rightarrow P$ (for P holds everywhere throughout the innermost sphere iff $\top \: \square\!\!\rightarrow P$). These conditionals may thus be used to map out the innermost sphere. A pre-theoretic reading of counterfactuals with strong centering failure would therefore allow us to approximate the size of the innermost sphere in a given context.

But it's doubtful that there is such a pre-theoretic reading. Suppose Anna says to Beth: "I bet you that, if I flicked the switch, the lights would turn on. If I'm right, you'll pay me \$1; if I'm wrong, I'll pay you \$1." Beth accepts, and Anna flicks the switch. The light turns on. Does Beth have to pay up? Of course she does. Anna flicked the switch and the light turned on. This is enough to ensure that Anna wins her bet. Suppose Beth responded: "Ah, you see: The cable is brittle, and the light bulb isn't screwed in all the way, and

the fuse is unstable. It wouldn't have taken much for the light *not* to turn on. Indeed, in some very nearby worlds you flick the switch and the light stays *off*. So, please, I want my money back!" This complaint is obviously pointless. No detail of the setup can change the fact that Anna won her bet. (If such details *were* to matter, we'd see many more bettors hurry to prove the modal frailty of the relevant material conditional, after losing a bet on a counterfactual.) This suggests that there is no pre-theoretic reading of the counterfactual with merely weakly centered models.[22]

LM fail to provide any further linguistic data to underwrite a reading which implements strong centering failure. They recapitulate Lewis's (1979) point, noted above, that not every difference between worlds should count toward dissimilarity. But this doesn't even provide indirect evidence for strong centering failure (let alone provide linguistic data). For, as Lewis himself notes, the point is perfectly compatible with strong centering: '[T]here may not be any worlds that differ from ours only in the respects that don't count, even if there are some respects that don't count' (466).[23]

LM go on to suggest some general theoretical reasons for strong centering failure. They claim that strong centering 'introduces an unjustified asymmetry into our reasoning about counterfactuals' (486):

> 'Consider the inference pattern *strengthening the antecedent*, which goes from the premise P $\Box\!\!\to$ Q to the conclusion (P & R) $\Box\!\!\to$ Q. ... When the antecedent P is false, this inference is generally invalid. ... However, under Lewis's strong centering requirement, this inference pattern is valid when the antecedents of the counterfactuals are true. But the inference seems equally bad when the antecedent is true as when it is false.' (486)

The non-standard use of "valid" in this quote is distracting. By itself, *strenghtening the antecedent* is invalid (in the proper sense of the term) on either account of centering. Hence there's no asymmetry here.

LM's worry, perhaps, is rather this. On either centering assumption, the following inference is invalid:

1. ¬P

2. ¬R

3. P □→ Q

Thus, (P & R) □→ Q

Further, on weak centering, *but not on strong centering*, the following inference is invalid:

1.* P

2.* R

3.* P □→ Q

Thus, (P & R) □→ Q

But it's unclear where the problematic asymmetry is. The two inference patterns above aren't symmetrical: 1.* negates 1., 2.* negates 2., but 3.* and 3 are identical. It's hard to see why the inference patterns should stand and fall together.

LM continue:

> 'This point is especially significant in the case of future-tense counterfactuals whose antecedents are not known to be true or false. It is crucial to the use of such counterfactuals for prediction and decision-making that inferences be valid or invalid regardless of the truth of their antecedents.' (ibid.)

"Valid" and "invalid" are again used non-standardly (an antecedent's truth value obviously doesn't affect an inference's validity, normally understood). LM presumably mean to reiterate that the previous two inference patterns should be either simultaneously valid or simultaneously invalid. As a reason they note that otherwise our confidence in (P & R) □→ Q would generally depend on our confidence in P and R. And this, they claim, is undesirable for the purpose of 'prediction and decision-making'. But why think that? In a situation where one believes that Q is true, for example, it seems perfectly reasonable to increase one's confidence in (P & R) □→ Q upon gaining evidence that P and R are true.

So it's not clear that there are any theoretical reasons to embrace strong centering failure, nor that strong centering failure explains any (pretheoretical) linguistic data. Both issues make LM's theory unappealing as a semantic theory of counterfactuals. Moreover, they make it unappealing as a theory of (difference-making) causation. For we are left with a parameter, the size of the innermost sphere, which sensitively influences the account's predictions yet remains wholly unexplained—it is determined neither by linguistic data nor by independent theoretical considerations. This doesn't make for a good theory of difference-making causation.

Further, it also limits the interest of LM's main theoretical result, Exclusion failure. Whether in any given case an event excludes another depends crucially on the size of the innermost sphere. So we are left without a good sense of when (and when not) exclusion occurs. Indeed, for all we know, it always does occur: Suppose x and z, such that x properly necessitates z, satisfy conditions (i) and (ii) for some effect y. It turns out that we can always modify the innermost sphere such that exactly one of x and z satisfies (i) and (ii).[24] So, for all we know, the innermost sphere conspires to save Exclusion, even on LM's account.

4.3   A Fix: Enoughness

We should look for a theory that's compatible with strong centering. Yablo (1992) is again of some help. One of Yablo's criteria does the work of condition (i) even given strong centering.

Plausibly, x's being proportional to an effect requires that x's realizers aren't *required* for the effect. A realizer is required, intuitively, only if the effect depends sensitively on the realizer's occurrence. A required realizer thus indicates that x contains too little information (if x had occurred, but with the a slightly different realizer, the effect would not have occurred). Define then[25]

> **Enoughness:** x is *enough* for y iff no event whose occurrence necessitates x's occurrence is *required* (in our sense) for y.

Now replace (i) in (DM) with:

(i*) x is *enough* for y.

(i*) excludes overly unspecific events. COLORED is not enough for PECK, since RED is required. Similarly, INTEND-S is not enough for TAXI, since INTEND is required.

Moreover, though, even RED isn't enough; for example, RED∧*Sophie's being able to move* is also required. (Sophie wouldn't peck if she were immobilized.) And even the latter is excluded, since an *even finer* conjunction is required (the target also has to be in pecking distance, it must be bright enough to see, etc). In general, on our account, difference-makers will be rather detailed conjunctions of events, capturing *all* aspects of the situation required for the effect.

This is a demand for causal detail, but one compatible with causal proportionality. Proportionality is, in the first instance, a theory of *vertical* causal selection, of locating causes at the "correct level of nature". It is therefore acceptable if *horizontal* selection—the project of distinguishing salient causes from mere "background conditions", primarily within a *single* level of nature—is treated separately.[26]

Since Mill (1874, Ch. 5) it is commonplace to distinguish between "whole" and "contributing" (or "partial") causes.[27] The new account may be viewed as delivering proportional *whole* causes, as opposed to mere proportional *contributing* causes. (Contributing causes may be recovered from the whole cause, of course, by considering its conjunctive expansions (ways of writing the event as a conjunction of events).) This strikes me as a reasonable output for a theory of proportionality, and a marked improvement upon LM's account.

5 EXCLUSION Saved

Let's review. My first criticism against LM has been that condition (ii) sometimes picks up on factors irrelevant to proportionality. I've suggested to replace it with a version of Yablo's (1992) criterion, that causes be "required", with a proviso ruling out badly disjunctive properties. My second criticism

18

has targeted LM's modification to Lewis's counterfactual semantics. The worry has been that their modification makes causation sensitive to the size of the innermost sphere, an unexplained parameter with questionable theoretical and pre-theoretical credentials. An alternative approach, compatible with orthodox strongly centered semantics, adopts Yablo's demand that proportional causes be "enough". (At the same time, though, not everything from Yablo (1992) should be imitated.[28])

The modified Yablo-style proposal can be summarized thus. We start with two formal features of events (I've added the terms "higher-level" / "lower-level" for easier comprehension):

- **Requiredness:** x is *required* for y iff, for all not-badly-disjunctive *higher-level events* z (events whose occurrence is necessitated by x's occurrence), $((\neg x) \wedge z) \; \Box\!\!\rightarrow \neg y$.

- **Enoughness:** x is *enough* for y iff no *lower-level event* (event whose occurrence necessitates x's occurrence) is *required* for y.

An event x is then *a difference-making* (or "*proportional*") *cause* of an event y iff x is both required and enough for y. (Of LM's original proposal only condition (ii) remains, as an implication of x's Requiredness.)

Having developed our alternative account, let's harvest its fruits. We have at least two noteworthy upshots. First, and most plainly, the new account shows that superior alternatives to LM's account exist. Our account is superior both in being more extensionally adequate, and in being theoretically and explanatorily more satisfactory (due to a strongly centered semantics).

Second, and more interestingly, the new account denies LM's main theoretical result, the falsity of Exclusion. LM's account permits cases in which both an event and its realizer cause the same effect. (This happens whenever the causal influence is 'realization-sensitive': whenever, that is, an event x is a difference-maker of effect y, but where, had x occurred without some actual realizer z, y wouldn't have occurred (List and Menzies, 2009, p. 493). On LM's theory (DM), z is a difference-maker of y in this case.[29]) Thus LM's theory permits failures of Exclusion.

By contrast, our Yablo-style account *entails* Exclusion: Suppose, for contradiction, that z is distinct from but necessitates x, and both x and z make a difference to y. Then, by definition of difference-making, x is enough for y, and so y requires no event distinct from but necessitating x. In particular, y does not require z. But then, by definition of difference-making, z does not make a difference to y, in contradiction with our assumption. It follows that no two distinct events x and z, where the occurrence of one necessitates the occurrence of the other, are difference-making causes of the same effect.[30]

Now, I do not claim that my account is the final word on the matter. (In particular, it would be desirable to extend the account to handle cases of preemption and overdetermination.) But so much at least is clear: The truth of Exclusion, when restricted to difference-making causation, is an open possibility. And settling the question requires careful attention, not only to the theoretical virtues of specific theories of causation, but also to the semantic accounts underwriting them.[31]

Department of Philosophy
New York University

Notes

[1]The relevant notion of supervenience here is *strong individual supervenience*: That is, necessarily, for any special-science property S, everything that has S has some microphysical property P such that, necessarily, everything that has P has S. (The relevant modality is *metaphysical* necessity.)

[2]By 'F necessitates F^*' I mean that, necessarily, every F is an F^*. List and Menzies instead write 'F^* supervenes on F'. This is loose talk, as supervenience, properly understood, relates not individual properties but *sets* of properties. To avoid confusion, I use strict talk throughout my presentation, distinguishing supervenience and metaphysical necessitation where necessary.

[3]For continuity's sake, I accept LM's (and Yablo's) assumption that there is a preferred meaning of the word "cause" expressing the difference-making relation. But LM's and Yablo's projects remain of interest even if one denies this, e.g. if one thinks that not every instance of causation is an instance of difference-making. What's important is

that difference-making occupies a distinctive explanatory role. This guarantees that it's independently interesting to model our usage of the concept, and to study its behavior vis-a-vis exclusion. Further, for the special sciences to be, in many cases, the exclusive supplier of *proportional causes* would still amount to a noteworthy sort of autonomy, and would secure them a privileged explanatory role. I ask readers who are loath to equate causation with difference-making to replace all occurrences of "cause" with "difference-making" or "proportional causation". (For a recent view distinguishing causation from difference-making, see Lee (forthcoming).)

[4] There are alternatives to counterfactuals-based approaches to causal proportionality, for example Michael Strevens's (2004) 'production' account. In this essay, I put those aside. Further, I'll follow Yablo and LM (2009, 478, Fn. 8) in focusing on simple causal systems only. In particular, I'll ignore causal systems involving preemption or overdetermination. Troubles with such systems are familiar and serious, but not unique to the current proposal.

[5] Initially LM also discuss an alternative reading of the exclusion principle, which replaces one occurrence of "cause" with "causally sufficient":

> EXCLUSION*: 'If a property F is *causally sufficient* for some effect G, then no distinct property F^* [metaphysically necessitated by] F can be a cause of the effect G.' (ibid.: 475; my emphasis)

Following Woodward (2008), List and Menzies (2009, p. 490) read 'causal sufficiency' as something like nomological entailment, and the remaining 'cause' as 'difference-maker'. On this reading, I agree that EXCLUSION* is false: Microphysical properties can nomologically necessitate special-science properties without being difference-makers. Throughout the essay, I'm concerned with EXCLUSION and not with EXCLUSION*.

[6] Cf. Lewis (1973, Ch. 2) and List and Menzies (2010, p. 6). Compared to Lewis, our account simplifies by adopting Exhaustiveness and the Limit Assumption. This is merely for expository convenience; nothing essential hangs on them.

[7] LM do not explicitly state (DM) in terms of "events". Instead, they ask when '[t]he presence of [some *property*] F' makes a difference to 'the presence of [some other *property*] G'. I stick to "event"-talk throughout this paper since that's the more established ideology. Nothing I'll say essentially hangs on this terminological departure from LM. For it's reasonable to suppose that, for any candidate difference-making property F, the presence of F is an event. (This is arguably false for generic properties, cf. Lewis's (1986) worry about disjunctive cases such as "Jeff's talking-or-walking". But what disqualifies them from denoting events—their disjunctiveness—plausibly also disqualifies them from denoting difference-makers.) Conversely, my criticism of LM goes through even if we focus only on the subclass of events expressible, for some F, by ⌜the presence of F⌝.

[8] Yablo formulates this condition in his own determinate/determinable framework; but it is straightforward to convert it into an event-theoretic framework.

[9] McGrath (1998) proposes a slight variation of Yablo's condition (*pace* McGrath, who claims to reproduce Yablo's account):

> **McGrath-requiredness:** x is McGrath-required for y iff, for all z realized by x, $\neg((\neg x) \wedge z \mathbin{\Box\!\!\rightarrow} y)$.

Restricted to non-vacuous counterfactuals, McGrath-requiredness is weaker than Yablo-requiredness. They can come apart in rare circumstances, namely when, for some z realized by x: it might be that y obtains when z is present without x, but it might also be that y doesn't obtain in that case. In this case, x may still count as McGrath-required, but wouldn't be Yablo-required. Yablo-requiredness seems slightly more faithful to the intuitive meaning of the word "required". The fact that Sophie *might* still have pecked had the target been non-red intuitively ensures that redness isn't required for pecking. (That she also *might not* have pecked doesn't seem like a good reason for thinking otherwise.) But being closer to the intuitive meaning of "requiredness" doesn't ensure being the superior condition for difference-making causation. In this respect, I don't see much daylight between the two conditions. Did RED make a difference to PECK, since, had the target not been red Sophie might not have pecked? Or does it not make a difference since she might still have pecked? I don't think I have a preference. Indeed, provided that facts about difference-making causation don't run metaphysically deep, I'm inclined to think that there is no fact of the matter here. For concreteness, I'll stick with Yablo-requiredness in the following (everything I say will hold for McGrath-requiredness too).

[10] B "screens off" A relative to effect C iff, had B occurred without A, C might still have occurred; *viz.* iff $\neg(((\neg A) \wedge B) \mathbin{\Box\!\!\rightarrow} \neg C)$.

[11] LM's account on its own doesn't suffer from this problem, but it still runs into trouble. It qualifies both RED *and* RED $\vee$ WORM as difference-makers. But if difference-making is causation, as LM claim, we arguably shouldn't admit disjunctive difference-makers like RED $\vee$ WORM, on pain of abundant overdetermination (cf. Lewis (1986b, p. 267)).

A later account of Yablo's (2003), based on naturalness, shares this trouble. Yablo defines: 'A property P of x is egregiously weak (relative to effect y) iff some more natural stronger property of x is better proportioned to y than P is.' Further: A property $Q+$ is better proportioned to y than a weaker property Q- is 'iff y would not have occurred, had x possessed Q- but not $Q+$' (324). On this definition, RED is no better proportioned to PECK than (RED $\vee$ WORM) is. Hence, (RED $\vee$ WORM) is not egregiously weak. That's enough to ensure that it qualifies as a proportional cause of PECK on Yablo's (2003) account.

[12] This issue is known in the literature as an instance of the "disjunction problem" for proportionality. Franklin-Hall (2016) explores the disjunction problem for accounts implementing Woodward-style interventionism. Franklin-Hall's discussion is independent of ours, since our counterfactual accounts are not wedded to interventionism. Shapiro and Sober (2012) also discuss the disjunction problem. Their discussion is independent of ours as well, since they only target accounts which employ condition (ii) as the *sole* criterion for difference-making.

[13] A recent alternative proposal is Rubenstein (unpublished), who identifies badly dis-

junctive events as those which are not *cohesive,* a notion which he links to our scientific practice of curve-fitting.

[14]If one is skeptical of extending naturalness from properties to events, here's an alternative: Restrict yourself to simple events for whose occurrence a proposition of the form ⌜a is F⌝ (for example, "Sophie pecks") states necessary and sufficient conditions. Where ⌜a is F⌝ expresses such a proposition for event e, call the property expressed by F a *characteristic property* of e. Then define: e is badly disjunctive iff any of its characteristic properties is badly disjunctive. Then follow Langton and Lewis's definition of bad disjunctiveness for properties.

[15]A "disjunctive expansion" of x is any way to divide up x into jointly exhaustive (but not necessarily mutually exclusive) disjuncts. (This presupposes that events have the requisite logical structure, an assumption which follows e.g. on Lewis's (1986) account, where events are sets of spacetime regions.) Lewis considers different variations of the given proposal, including substituting "much more" with "more", and "all" with "some". Each move would enlarge the set of bad disjunctions.

[16]We also get that the right-hand side of Requiredness is never vacuous: No event necessitates only badly disjunctive events, since every event necessitates the all-encompassing event.

[17]There are also circumstances in which upwards exclusion is achieved by condition (ii). This happens if the effect y occurs somewhere inside the set of closest $\neg z$-worlds but *outside* the set of closest $\neg x$-worlds. I'll ignore this case for the most part, only returning to it briefly in Fn. 19.

[18]*Proof:* The set of closest INTEND-worlds is the innermost sphere, which is a subset of T. Hence INTEND $\square\!\!\rightarrow$ T. The same holds for INTEND-S, and so INTEND-S $\square\!\!\rightarrow$ T. The closest ¬INTEND-worlds are all and only the ¬INTEND worlds in the second-closest sphere, and so are a subset of ¬T. Similarly, the set of all closest ¬INTEND-S-worlds is a subset of ¬T. Hence also INTEND $\square\!\!\rightarrow$ T and ¬INTEND $\square\!\!\rightarrow$ ¬T. □

[19] In addition, upwards exclusion may occur due to condition (ii) (cf. Fn. 17). Let x be a low-level event and z be a high-level event necessitated by x. It's logically possible that, even though the effect is absent in all closest $\neg x$-worlds, it's nonetheless present in some closest $\neg z$-worlds. In this case, even though x satisfies condition (ii), z violates it. In the examples I am concerned with, this case doesn't occur.

[20]*Proof:* Suppose z violates (i) and x satisfies (i). Since both x and z actually occur, and x necessitates z, the set of closest x-worlds is a subset of the set of closest z-worlds. Now suppose, for contradiction, that the set of closest x-worlds is *identical* to the set of closest z-worlds. Then, since the effect occurs at all closest x-worlds (x satisfies condition (i)), it also occurs at all closest z-worlds, in contradiction with z's violating condition (i). □

[21]The counterfactuals in question are any truths of the form "If A, the world would be very different; but if A and B, the world would not be very different." Here, "very different" is supposed to track our pretheoretic similarity judgments. Since the closest $A \wedge B$-world

lies no closer to the actual world than the closest A-world, the counterfactual is true only if the formal similarity relation disagrees with pretheoretic similarity. (Lewis (1979, p. 466) attributes this point independently to Pavel Tichý and Richard J. Hall.)

[22]Cf. Lewis (1973, p. 27), who provides similar evidence for strong centering, in the form of an imagined conversation between two participants who consider a counterfactual conditional while disagreeing about the truth value of its antecedent.

Interestingly, Lewis (1973, p. 3) thinks that we can't 'directly' test strong centering, since ordinary counterfactuals 'carry some sort of presupposition that the antecedent is false'. The presupposition would make it hard to investigate whether our pre-theoretic reading validates "And-to-If", since it would be hard to disentangle our intuitions about the falsity of a counterfactual from our sense of defect resulting from presupposition failure. But Lewis's views about presupposition are doubtful. To take a famous example from Anderson (1951): 'In the investigation of Jones' death, a doctor might say, "If Jones had taken arsenic, he would have shown just exactly those symptoms which he does in fact show".' In this case, the counterfactual doesn't seem to carry any presupposition that the antecedent is false. These Anderson-type cases may be viewed as additional evidence in favor of strong centering. (Thanks to Cian Dorr here.)

[23]LM (2009, note 25) also claim that their account is *compatible* with Lewis (1979). But that's false. Lewis's similarity metric entails strong centering up to qualitative duplication: 'It is of the first importance to avoid big, widespread, diverse violations of law. It is of the second importance to maximize the spatio-temporal region throughout which perfect match of particular fact prevails.' (1979, p. 472) The actual world matches itself perfectly in these respects, unlike any world qualitatively different from it.

[24]*Proof:* Suppose that

(i) $x \mathbin{\Box\!\!\rightarrow} y$

(ii) $(\neg x) \mathbin{\Box\!\!\rightarrow} \neg y$

(i*) $z \mathbin{\Box\!\!\rightarrow} y$

(ii*) $(\neg z) \mathbin{\Box\!\!\rightarrow} \neg y$

Since $x \subset z$, there is a $((\neg x)\&z)$-world. Now simply modify the innermost sphere to include that world. If the world is a $\neg y$-world, then (i)* is false, while (i), (ii), and (ii*) remain true. If it is a y-world, then (ii) is false, while (i), (i*), and (ii*) remain true. □

[25]Since this condition involves requiredness, it also has a McGrath analogue (substituting "required" by "McGrath-required", cf. note 9). Again, there isn't much daylight between the two conditions (assuming that the quantifier in "McGrath-requiredness" is restricted to not-badly-disjunctive properties). They come apart only in cases where a high-level cause "barely" fails to screen off its low-level realizer, in the sense that, in the absence of the low-level realizer, the effect *might* occur but it also *might not.* In this case, the high-level cause isn't McGrath-enough, but it may still be (Yablo-)enough. McGrath-enoughness is *stronger* than (Yablo-)enoughness.

I'm borrowing the "horizontal"/"vertical" terminology from Franklin-Hall (2015). Incidentally, Franklin-Hall defends an account which achieves horizontal and vertical selection simultaneously. Her defense rejects the prevailing view, endorsed in Mill (1874), Lewis (1986a), Hall (2004), and Schaffer (2012), that our practice of horizontal selection is 'capricious' (Mill's words) and highly sensitive to conversational context.

[27]This mirrors the now likewise commonplace distinction between "full" and "partial" ground (Fine, 2012).

[28]We've already seen that his demand for Contingency is redundant. In addition, Yablo proposes a demand for "Adequacy". The idea here is to check not only for the truth of the consequent given the truth of the antecedent, but also for the truth of the consequent given the truth of the antecedent, *given the falsity of the antecedent.* This, intuitively, is another measure for how reliably the effect arises from the cause, and serves to exclude excessively high-level causes. To illustrate: In (Sophie), had the target not been colored—e.g. had it been uniformly white—then, *if it had been colored*, the target might plausibly have been blue. For, we may suppose, had the target been white, the scientists would still have had red and blue targets available. So, the scientists might have picked a blue target. So Sophie might not have pecked. Formally, Yablo defines Adequacy as follows:

Adequacy: x is *adequate* for y iff $(\neg x) \;\square\!\!\rightarrow (x \;\square\!\!\rightarrow y)$.

But as with Contingency, a demand for Adequacy seems redundant. Its work is already done by Enoughness. In (Sophie), COLORED is already excluded on the grounds that it is not enough; similarly with INTEND-S in the taxi case. Yablo himself introduces Adequacy with the following example:

> **(Boiler)** 'Safety valves are designed to open quickly under extreme pressure, thus easing the burden on the equipment upstream. This particular valve has begun to operate as advertised when a freak molecular misalignment stiffens the mechanism; this decelerates the opening to just past the point of endurance and the boiler explodes. Assuming that the explosion does not result from the valve's opening *per se*, I ask why not. Because the contingency condition is violated? But we can arrange it so that the explosion was contingent on the opening, say, by stipulating that if the opening had not occurred, rather than the boiler's exploding the connecting pipe would have burst. Adequacy does better: given the unlikelihood of the molecular mishap, had the opening failed to occur, it might easily have been quicker if it had. Speaking then of how things would have been if not for the opening, it cannot be said that, were it to have occurred, it would still have brought the explosion in its wake.' (1992, pp. 275–6)

A demand for Adequacy excludes the valve's opening (*simpliciter*): Had the valve not opened, then, if it had, it might not have opened in the freak-molecular-mishap-way it actually did. It might instead have opened sufficiently quickly to prevent the boiler's

explosion. However, Enoughness already excludes the valve's opening (*simpliciter*). This is because something more specific is required, namely the valve's opening *slowly*. (Had the valve opened, but not slowly, then the boiler would have been fine.) So Yablo's example fails to motivate a demand for Adequacy.

[29]*Proof:* Suppose x is a difference-making cause of y, i.e.,

(i*) $x \ \Box\!\!\rightarrow \ y$,

(ii*) $(\neg x) \ \Box\!\!\rightarrow \ \neg y$.

Since z necessitates x, (i*) entails, by (DM),

(j*) $z \ \Box\!\!\rightarrow \ y$.

Further, we suppose that the causal relation between x and y is realization-sensitive, i.e.,

(RS) $((\neg z) \wedge x) \ \Box\!\!\rightarrow \ \neg y$.

(RS) together with (ii*) and the fact that z necessitates x implies

(jj*) $(\neg z) \ \Box\!\!\rightarrow \ \neg y$.

(To see this last inference: Suppose, for contradiction, that some closest $\neg z$-world α is a y-world. By (RS), all closest $(\neg z) \wedge x$-worlds are $\neg y$-worlds. So α must be a closest $(\neg z) \wedge (\neg x)$-world. By (ii*) and the fact that z necessitates x, all closest $(\neg z) \wedge (\neg x)$-worlds are $\neg y$-worlds. Hence α is also a $\neg y$-world. Contradiction.) But (j*) and (jj*) are just the conditions for z's being a difference-maker for y. □

[30]Exclusion is maintained even for realization-sensitive setups. To repeat, the causal relation between cause x and effect y is realization-sensitive iff there is a lower-level z (a z whose occurrence necessitates x's occurrence) such that y wouldn't have occurred had x occurred without z. In this case, z *may or may not* be required for y. (For z to be required for y, it must be that for *every* higher-level x, y wouldn't occur had x occurred without z.) If z is required, then x isn't enough, and hence not a difference-maker. If z isn't required, then z itself is not a difference-maker (and x may still be enough).

References

Anderson, Alan Ross (Dec. 1951). "A Note on Subjunctive and Counterfactual Conditionals". In: *Analysis* 12.2, pp. 35–38. ISSN: 0003-2638. DOI: 10.1093/analys/12.2.35.

Fine, Kit (2012). "Guide to Ground". In: *Metaphysical Grounding*. Ed. by Fabrice Correia and Benjamin Schnieder. Cambridge University Press, pp. 37–80.

Franklin-Hall, L. R. (2015). "Explaining Causal Selection with Explanatory Causal Economy: Biology and Beyond". In: *Explanation in Biology: An Enquiry into the Diversity of Explanatory Patterns in the Life Sciences*. Ed. by P.-A. Braillard and C. Malaterre. Springer, pp. 413–438.

— (2016). "High-Level Explanation and the Interventionist's 'Variables Problem'". In: *British Journal for the Philosophy of Science* 67.2, pp. 553–577.

Hall, Ned (2004). "Two Concepts of Causation". In: *Causation and Counterfactuals*. Ed. by John Collins, Ned Hall, and Laurie Paul. MIT Press, pp. 225–276.

Kim, Jaegwon (Apr. 2000). *Kim, J: Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. English. Reprint edition. Cambridge, Mass.: Gb. ISBN: 978-0-262-61153-4.

— (2005). *Physicalism, or Something Near Enough*. Princeton University Press. ISBN: 978-0-691-13385-0.

Langton, Rae and David Lewis (1998). "Defining 'Intrinsic'". In: *Philosophy and Phenomenological Research* 58.2, pp. 333–345. ISSN: 0031-8205. DOI: 10.2307/2653512.

Lee, Samuel (forthcoming). "Building Low Level Causation out of High Level Causation". In: *Synthese*, pp. 1–29.

Lewis, David (1973). *Counterfactuals*. Blackwell.

— (1979). "Counterfactual Dependence and Time's Arrow". In: *Noûs* 13.4, pp. 455–476.

— (1986a). "Causal Explanation". In: *Philosophical Papers Vol. Ii*. Ed. by David Lewis. Oxford University Press, pp. 214–240.

— (1986b). "Events". In: *Philosophical Papers Vol. II*. Ed. by David Lewis. Oxford University Press, pp. 241–269.

— (2001). "Redefining 'Intrinsic'". In: *Philosophy and Phenomenological Research* 63.2, pp. 381–398. ISSN: 0031-8205. DOI: 10.2307/3071071.

List, Christian and Peter Menzies (Sept. 2009). "Nonreductive Physicalism and the Limits of the Exclusion Principle". In: *The Journal of Philosophy* 106.9, pp. 475–502.

— (2010). "The Causal Autonomy of the Special Sciences". In: *Emergence in Mind*. Ed. by Cynthia Mcdonald and Graham Mcdonald. Oxford University Press.

McGrath, Matthew (1998). "Proportionality and Mental Causation: A Fit?" In: *Philosophical Perspectives* 12, pp. 167–176. ISSN: 1520-8583.

Mill, John Stuart (1874). *A System of Logic*. Longman.

Rubenstein, Ezra (unpublished). "Cohesive Proportionality". en.

Schaffer, Jonathan (2012). "Causal Contextualisms". In: *Contrastivism in Philosophy: New Perspectives*. Ed. by Martijn Blaauw. Routledge.

Shapiro, Larry and Elliott Sober (2012). "Against Proportionality". In: *Analysis* 72.1, pp. 89–93. ISSN: 0003-2638.

Strevens, Michael (2004). "The Causal and Unification Approaches to Explanation Unified—Causally". In: *Noûs* 38.1, pp. 154–176. DOI: `10.1111/j.1468-0068.2004.00466.x`.

Woodward, James (2008). "Mental Causation and Neural Mechanisms". In: *Being Reduced: New Essays on Reduction, Explanation, and Causation*. Ed. by Jakob Hohwy and Jesper Kallestrup. Oxford University Press.

Yablo, Stephen (1992). "Mental Causation". In: *The Philosophical Review* 101.2, pp. 245–280. ISSN: 00318108, 15581470. DOI: `10.2307/2185535`.

— (2003). "Causal Relevance". In: *Philosophical Issues* 13.1, pp. 316–28. DOI: `10.1111/1533-6077.00016`.