

Dispositional Predicates in Context

by

Alexander Anthony

Class of 2010

A thesis submitted to the
faculty of Wesleyan University
in partial fulfillment of the requirements for the
Degree of Bachelor of Arts
with Departmental Honors in Philosophy

CONTENTS

Introduction	4
Part I: Recent Analysis of Dispositional Predicates	
1. The Simple Conditional Analysis	7
2. The closeness counterexamples	8
3. Relevance, exclusion and tolerance	10
4. The Lewis conditional analyses	16
5. Fara's habitual analysis	41
6. Manley and Wasserman's proportional account	47
Part II: Context-Sensitivity Arguments	
7. The context-sensitivity of dispositional predicates	53
8. Context shifting arguments	54
9. The conditional analyses	62
10. Fara's habitual analysis	69
11. Manley and Wasserman's proportional account	83
Part III: Paradigms and Interpretation	
12. Toward an <i>exclusionary</i> proportional account	87
13. The basic model of interpretation	89
14. Interpretational constraints	93
15. Fink-, mask-, and mimic-exclusion	98
16. Two final revisions	100
17. The final model	103
18. The model at work	105
19. Applications to lexical competence, vagueness, and conceptual unity	117
Works Cited	122

Abstract: In this paper I critically examine six analyses of dispositional predicates. In Part I, I argue that none of the analyses are extensionally adequate. In Part II, I consider the context-sensitivity of dispositional predicates and argue that it is not well explained by any of the analyses considered in Part I. In Part III, I propose a new analysis and argue that it is extensionally adequate and accommodates the context-sensitivity of dispositional predicates.

Dispositional predicates in context¹

ALEX ANTHONY

Introduction

Dispositional properties and *dispositional predicates* have attracted significant recent attention within metaphysics, philosophy of science, and philosophy of language. Discussion has centered on a few central topics: the dispositional/categorical distinction, dispositions and causal bases, dispositions and laws of nature, and the intrinsicness of dispositions.

Perhaps the most active area of recent work has been the semantics of dispositional predicates and disposition ascriptions. Ever since Rudolf Carnap's verificationism led him to produce reduction sentences for disposition ascriptions in 'Testability and Meaning,'² there has been significant interest in the semantics and logical form of disposition ascriptions. For much of the second half of the twentieth century there was a broad tacit consensus that the Simple Conditional Analysis of dispositional predicates (SCA) – for object x and dispositional predicate 'D', ' x is D' is true just in case, were x to be in a D-associated testing condition, it would give a D-associated response (e.g. ' x is fragile' is true iff were x stressed, x would break) – was extensionally adequate and provided a nice explanation of the meaning of disposition ascriptions. However, in the early 90's Mark Johnston and C.B. Martin published papers describing cases that were almost universally accepted as counterexamples to the Simple Conditional Analysis. In the wake of these counterexamples, there has been a wave of papers attempting to provide extensionally adequate analyses of the semantics of dispositional predicates.

Why have philosophers paid so much attention to the semantics of dispositional predicates? Why worry about the failure of the (SCA)?

¹ I'd like to acknowledge Markus Schrenk and Sanford Shieh for a great deal of help with this paper. Many of the ideas presented here grew out of a Hilary Term 2009 tutorial with Dr. Schrenk and were aired in a Fall 2009 thesis tutorial with Professor Shieh. Both provided enormously helpful comments and direction both in the tutorials and on drafts of the paper.

² Carnap 1936, pp. 440-441.

Michael Fara makes the case that understanding the semantics of dispositions is important for philosophers working on a wide range of topics:

The need for a correct account of dispositions is urgent in philosophy. For just about every concept that philosophers have been interested in, there are some who hold a so-called 'dispositional analysis' of that concept. Among the most well known are dispositional analyses of meaning, offered in response to Kripke's presentation of Wittgenstein's skeptical challenges [Saul Kripke, Wittgenstein on Rules and Private Language], and dispositional accounts of the mental in the form of versions of behaviourism and functionalism. But there are also dispositional analyses of the colors [Colin McGinn, The Subjective View]; of value [David Lewis, 'Dispositional theories of value']; of goodness [Michael Smith, The Moral Problem]; of 'stimulus meaning' [W.V. Quine, Word and Object]; of properties [Sydney Shoemaker, 'Causality and properties']; and of conditionals [D.H. Mellor, 'How to believe a conditional']. Advocates of these various analyses have offered arguments in support of them, and their opponents have responded with counterarguments. But as Mark Johnston [(1992), pp. 232-233] has emphasized with regard to the particular case of dispositional theories of the colors, care should be taken when adjudicating these arguments. For a dispositional theory may wrongly be taken to be mistaken while the real mistake lies in a misunderstanding of the dispositional vocabulary employed by that theory. We need an account of dispositions before we can confidently assess dispositional analyses of one or another philosophically problematic concept.³

In this paper, I critically examine five analyses of disposition terms proposed since (and in response to) the Johnston/Martin counterexamples. I find each of the analyses wanting, and propose a new analysis of my own.

The paper is organized into three parts. In the first part of this paper, I consider the extensional adequacy question – are any of analyses proposed in recent years extensionally adequate? I argue that none are.

³ Fara 2005, pp. 43-44.

In the second part of the paper, I turn to the context-sensitivity of dispositional predicates. In recent years, some philosophers have argued that their analyses provide a mechanism for explaining and predicting the context-sensitivity of disposition ascriptions. I consider explanations of the context-sensitivity of dispositional predicates available to proponents of each of the analyses from Part I and argue that while some of the analyses might provide a *partial* explanation for the context-sensitivity of dispositional predicates, none can accommodate all of the context-dependence displayed by dispositional predicates.

In the third part, I propose an analysis of dispositional predicates and argue that it is both extensionally adequate and provides a mechanism for the context-sensitivity of dispositional predicates. I then consider some applications of the analysis to semantic questions about dispositional terms concerning lexical competence, vagueness, and conceptual unity. At the center of the analysis is a theoretical odd couple – *paradigms* (category exemplars in the tradition of prototype-theory) and *interpretation* – but I argue that the extensional adequacy and explanatory power of the analysis nonetheless provide good reason to prefer it to the other analyses.

PART I: RECENT ANALYSES OF DISPOSITIONAL PREDICATES

1. The Simple Conditional Analysis

For a number of years, the Simple Conditional Analysis of dispositional predicates like 'fragile' and 'soluble' enjoyed widespread tacit acceptance:

For a dispositional predicate 'D', there is an associated testing condition T and associated response R such that for an object x , 'x is D' is true iff if x were in T, x would R.

Suppose, for instance, the dispositional predicate 'fragile' is associated with the testing condition *is stressed* and the response *breaks*; thus, on the Simple Conditional Analysis (SCA), 'x is fragile' is true iff if x were stressed, it would break. The (SCA) does a fine job approximating the truth conditions of many dispositional predicate ascriptions, but in the early nineties, papers by Mark Johnston and C.B. Martin introduced a battery of counterexamples⁴:

Finks: x is D, but were x to be in T, before R occurred, x would lose D; thus R would fail to occur.

Reverse-finks: x is not D, but were x to be in T, x would gain D; thus R would occur.

Masks: x is D, but were x to be in T, some extrinsic interference would prevent R from occurring.

Mimics: x is not D, but were x to be in T, some extrinsic factor would cause R to occur.

⁴ Martin 1994, pp. 2-4 and Johnston 1992, pp. 232-233.

Martin's original fink and reverse-fink cases are counterexamples to the analysis 'x is live' is true iff if x were touched by a conductor, electricity would flow from x to the conductor. Suppose x is a live wire, but is connected to an *electro-fink*, a device which detects when x is touched by a conductor and deadens x until it ceases to be touched by the conductor. *Ex hypothesi*, x is live, but were x to be touched by a conductor, electricity would *not* flow to the conductor, since the electro-fink would deaden the wire. The dispositional analysandum holds, but the conditional analysans does not. An electro-fink running on a reverse-cycle pumps electricity into a dead wire when it is touched by a conductor, 'livening' it. Were a reverse-finkish dead wire to be touched by a conductor, electricity would flow to the conductor, but the wire (*ex hypothesi*) is not live.

Johnston suggested that a certain glass's fragility might be *masked* by special internal packing which prevents it from breaking when stressed.⁵ Though the glass does not lose its fragility, were it to be stressed, it would not break. A very sturdy rock with a stress-activated mine attached would *mimic* fragility. Were the rock stressed, the mine would detonate and cause the rock to break, but the rock is not fragile.

2. The closeness counterexamples

If we get more precise about the logic of the subjunctive conditional by adopting the semantics of David Lewis (1973), we can bring out two additional problems with the (SCA).⁶ Lewis suggests that evaluating counterfactual conditionals requires a set of sets of possible worlds, which he calls a system of *spheres*. Spheres are sets of possible worlds centered on a world *i* to form a nested hierarchy arranged by the 'closeness' of the worlds in each sphere to *i*:⁷

⁵ Alexander Bird's *antidotes* are a subset of masks, viz. masks which are 'administered' after the disposition test.

⁶ These problems arise just the same in natural language, but it is much easier to describe them with the semantics for counterfactuals in place.

⁷ I'll leave the notion of comparative similarity at the intuitive level here and drop the scare quotes when using terms like 'close' and 'nearby' to describe comparative similarity relations. For a classic explication of similarity – Lewis's 'four weights' account – see Lewis 1979, p. 472.

Any particular sphere around a world i is to contain just those worlds that resemble i to at least a certain degree. This degree is different for different spheres around i . The smaller the sphere, the more similar to i must a world be to fall within it. To say the same thing in purely comparative terms: whenever one world lies within some sphere around i and another world lies outside that sphere, the first world is more closely similar to i than the second.⁸

Call a world where it is the case that ϕ a ϕ -world. On Lewis's semantics, $\lceil (\phi \Box \rightarrow \psi) \rceil$ is non-vacuously true at w iff there is a sphere centered on w in which there are ϕ -worlds, and every ϕ -world in that sphere is a ψ -world. In the language of comparative similarity: $\lceil (\phi \Box \rightarrow \psi) \rceil$ is non-vacuously true at w iff the most similar ϕ -worlds to w are all ψ -worlds.

The next counterexample arises because there can be cases in which even though an object is not disposed, all of the closest *test condition*-worlds just so happen to be *response*-worlds by a sort of fluke of comparative similarity. Here's an example, adapted from David Manley and Ryan Wasserman.⁹ A sturdy brick is precariously placed to hold open a twentieth story apartment window. Consider the (SCA) with the Lewis conditional ' $\Box \rightarrow$ ' playing the role of the subjunctive conditional: '*fragile*(x) iff *stressed*(x) $\Box \rightarrow$ *breaks*(x)'. All of the *stressed*(brick)-worlds closest to the actual world happen to be *breaks*(brick)-worlds, because all of these closest *stressed*(brick)-worlds are ones in which the brick slips out of the window and endures a quite severe stress hitting the pavement after falling twenty stories. Despite the fact that the brick wouldn't break in response to most ordinary, less severe stress, there is a *stressed*(brick)-permitting sphere at which every *stressed*(brick)-world is a *breaks*(brick)-world. Thus, the analysis holds – were the brick stressed it would break – but *ex hypothesi* the brick is not fragile.

The final counterexample arises because the system of spheres is *strongly centered*: on Lewis's semantics, no world is as close to w as itself, so $\lceil (\phi \& \psi) \rceil \not\models \lceil (\phi \Box \rightarrow \psi) \rceil$ (since if w is a ϕ -world and a ψ -world, each

⁸ Lewis 1973, p. 14.

⁹ Manley and Wasserman 2008, p. 70.

of the closest φ -worlds (just w) is a ψ -world). Suppose the precariously placed brick *actually* falls and breaks (rather than merely *possibly* breaking in nearby test-worlds). Since the inner sphere of *stressed*(x)-worlds consists of the actual world alone, and it is a *breaks*(x)-world, were the brick stressed it would break. But *ex hypothesi* the brick is not fragile.

3. Relevance, exclusion, and tolerance

The counterexamples above show that these analyses of ‘fragile’ and ‘live’ are extensionally inadequate. Does the extensional inadequacy rest on the particular dispositional predicates and associated conditions, or can counterexamples like these be produced for other similar analyses?

Counterexamples like those above can plausibly be provided for any analysis of an English language dispositional predicate in terms of vague, rough-and-ready associated test and response conditions. For instance, consider

‘ x is soluble’ is true iff if x were placed in a solvent, x would dissolve.

‘ x is poisonous’ is true iff if x were ingested by y , x would kill y .

‘ x is buoyant’ is true iff if x were placed in water, x would float.

In each of these cases, subjecting an object with the disposition to the test condition does not causally suffice for the occurrence of the response. If I place some sugar in a beaker of solvent and promptly break the beaker with a hammer, the sugar will not dissolve in the solvent. If I ingest a poison but also its antidote, the poison will not kill me.

Here are some boilerplate cases that are counterexamples to almost any rough-and-ready associated conditions Simple Conditional Analysis. First boilerplate example, generalized from a familiar story from David Lewis¹⁰: x has disposition D and is subjected to the test condition at time t_1 , but at time t_2 very soon after t_1 , a meddling sorcerer removes the property of x which is its causal basis for D . The sorcerer is quick enough that object

¹⁰ Lewis 1997, pp. 147-148.

x doesn't have enough time to manifest the disposition before the removal of the causal basis.

The second boilerplate example is just as before, except instead of a meddling sorcerer, it is a nearby nuclear device that enters the scene at t_2 , detonating and destroying everything in the vicinity. x isn't even around after t_2 , so it doesn't give response R .

The analyses under consideration take the general form $\Box(D(x) \leftrightarrow T(x) \Box \rightarrow R(x))$ for disposition 'D', test condition 'T', and response 'R' — an object has the disposition iff if it were in the test condition, it would give the response. Both of the boilerplate examples are cases in which $D(x)$ and $T(x)$ but not $R(x)$. It appears that such counterexamples can generally be provided for analyses along these lines.

Most (all?) philosophers agree that the truth value of dispositional predicate ascriptions depend upon various actual and (non-actual) possible events.¹¹ For this reason, the logical machinery philosophers use to analyze dispositional predicates — e.g. the subjunctive conditional¹², habituals¹³, and quantification over possible events¹⁴ — usually also 'depend upon' various actual and possible events.

In what follows, I'm going to lean on this intuitive notion of a proposition's truth value 'depending upon' some possibilities.¹⁵ Let's call all of the possibilities that proposition p depends upon the possibilities *relevant* to p . For example, take the explicitly modal statement 'It is nomologically possible that grass is green,' evaluated at the actual world, @. The possibilities that the proposition expressed by that statement 'depends upon' for its truth — the ones that we 'look to' if we're interested in assessing the truth or falsity of the proposition — are intuitively all and only the nomologically possible worlds (accessible to @). Another

¹¹ This formulation is deliberately quite vague; how exactly it 'depends upon' possible events and especially *which* possible events it depends upon are not problems on which philosophers have reached a consensus.

¹² E.g. Lewis 1997, Mumford 1998, Malzkorn 2000, Gundersen 2002, Choi 2008

¹³ Fara 2001, 2005.

¹⁴ Manley and Wasserman 2007, 2008.

¹⁵ These *possibilities* or *possible events* will be more finely individuated than possible worlds, since disposition ascriptions may depend upon multiple events in the same world.

example: on Lewis's counterfactual semantics, the (non-vacuous) truth or falsity of $\lceil (\varphi \Box \rightarrow \psi) \rceil$ depends upon *the inner sphere of φ -worlds*, i.e. the φ -worlds most comparatively similar to the actual world.

In the above examples, the possibilities which are relevant can be determined just by looking at the semantics of the modal operator and counterfactual conditional, respectively. Lewis's semantics tell us to look to the inner sphere of φ -worlds (in non-vacuous cases) to evaluate $\lceil (\varphi \Box \rightarrow \psi) \rceil$; possible worlds semantics for alethic modals evaluate in a model $\langle W, R, v \rangle$ with set of worlds W , binary accessibility relation between worlds R , and valuation v which assigns truth values to atomic sentences in every $w \in W$. The worlds *relevant* to the proposition expressed by $\lceil \Diamond \varphi \rceil$ evaluated with respect to w are all and only those worlds u such that $R(w, u)$.

Though we don't have a formal account of which possibilities are relevant to disposition attributions, we do have an intuitive sense of the relevant possibilities: e.g. the possible events relevant to whether or not a certain object x is (for instance) soluble are (roughly) those in which x is placed in a solvent.

Two qualifications (almost?) everyone agrees upon: only *nomologically possible* worlds are relevant (whether or not my laptop is fragile doesn't depend upon how it would respond to stress in possibilities with zany laws), and only possibilities in which x is not intrinsically different than it actually is (we don't care about possible events in which, e.g. a non-fragile vase was, contrary to fact, made into a very fragile vase).

Let's call those possible events relevant to a simple disposition attribution ' a is D ' the $D(a)$ -tests. Intuitively, not all objects can be put in the same testing conditions: e.g. I might try to crush a fragile marble in my hand, but a skyscraper could not be subjected to the same fragility-test. The union of $D(x)$ -tests for any x whatsoever we'll call the D -tests (e.g. the fragility-tests =_{df} the union of fragility(x)-tests for any x). A D -test in which the response is given, we'll call a *D-manifestation*; a D -test in which the response is *not* given, we'll call a *failed D-test*. I will use these terms to describe some adequacy conditions on analyses of dispositional predicates. Bear in mind that my usage of terms like 'D-test,' 'fragility-response,' etc. is

in accordance with the stipulative definitions above in terms of *relevance*. However, as the notion of relevance is grounded in our intuitive conceptions of the test conditions for and manifestations of a disposition, these adequacy conditions will ultimately derive from our intuitions about cases.

A crucial question for providing an analysis that avoids the counterexamples is this: are possibilities in which a disposition is finkish or masked failed D-tests or not D-tests at all? If these fink and masking cases are *relevant* to disposition ascriptions then they should be ‘included’ in the possibilities relevant to the analysans. Are reverse-fink and mimicking cases D-manifestations or failed D-tests?

Analyses since the Martin and Johnston counterexamples can be divided into two categories, depending on their response to these questions. Here’s a simplified example to illustrate. Suppose a certain fragile glass x breaks in nine worlds out of ten total in the inner sphere of stressed(x)-worlds, but fails to break in the tenth where its disposition is masked by Johnston packing. According to the (SCA), the glass is not fragile, since, ‘stressed(x) $\Box \rightarrow$ breaks(x)’ only holds if *all* of the stressed(x)-worlds in the inner sphere are breaks(x)-worlds. There are two broad approaches to correcting this shortcoming. One – *the exclusionary approach* – will get that anomalous possibility out of the worlds relevant to whether or not x is fragile, e.g. by building into the conditional antecedent the qualification that the object be *stressed without being masked*. With such a qualification in place, the anomalous possibility is no longer in the inner sphere of T(x)-worlds, and thus no longer relevant to the analysans (given Lewis’s semantics for counterfactuals.) The new analysis sidesteps the counterexample by excluding the masking case from the relevant possibilities. On the other approach – *the tolerant approach* – the anomalous possibility is not excluded from the worlds relevant to the analysans but rather is *tolerated as an exception*. The subjunctive conditional, on Lewis’s semantics, is intolerant of exceptions in the following sense: ‘ $(\phi \Box \rightarrow \psi)$ ’ is non-vacuously false if *any* of the ϕ -worlds in the inner sphere are not ψ -worlds, even if *most* or *almost all* of ϕ -

worlds in the inner sphere are ψ -worlds. We might pursue the tolerant approach by abandoning the Lewis semantics for a semantics on which, for instance, $\lceil (\varphi \Box \rightarrow \psi) \rceil$ is true iff at least 90% of the inner-sphere of φ -worlds are ψ -worlds. The anomalous possibility is still relevant to this analysis, but it is now tolerated as an exception: since 90% of the inner-sphere stressed(x)-worlds are breaks(x)-worlds, ‘stressed(x) $\Box \rightarrow$ breaks(x)’ holds, and thus (on our hypothetical analysis), x is fragile.

While it is possible to mix exclusionary and tolerant strategies to deal with different sorts of counterexamples, in practice philosophers have exclusively opted for one or the other. Let’s call the thesis that finks and masking cases are not failed tests but rather *not* disposition test events at all (and thus should be excluded from the analysis-relevant test possibilities) and that mimicking cases are failed tests rather than manifestations *the exclusion thesis*. As far as I am aware, no one has ever argued for or against the exclusion thesis. Despite this, the exclusionary approach has been dominant until very recently, and intuitions seem to fall pretty heavily in support of exclusion. The electro-fink scenario Martin describes is intuitively one which is *irrelevant* (in my sense of *relevance*) to whether or not the wire is live. A (mimicking) scenario in which a brick breaks when stressed because an attached stress-activated mine detonates is intuitively one which is *irrelevant* to whether or not the brick is fragile.

Here’s a more detailed example which is an intuition pump for the exclusion thesis. Consider two poisonous substances, *Discreet Poison (DP)* and *Indiscreet Poison (IP)*, both of which cause death at similar lethal doses and via relatively similar modes of action. *DP* and *IP* have almost exactly the same effects, and as a result, behave similarly in non-masking poison-test cases. They also both have readily available antidotes. However, there is one important difference between *DP* and *IP*: exposure to *IP* immediately causes victims’ skin to turn bright purple all over, while exposure to *DP* does not. As a result, *IP* exposure is almost always quickly identified and the antidote administered; *DP* exposure is much less frequently identified, and as a result, the effects of *DP* are much less frequently masked by the antidote. Are the antidote cases relevant to

‘poisonous’-ascriptions to *DP* and *IP*? If the antidote cases are irrelevant, then *DP* and *IP* have similar distributions of D-responses among the D-tests; this would suggest that they are similarly poisonous. If the antidote cases are relevant, then *DP* would have a much higher proportion of D-responses among the D-tests than *IP*; this would suggest that *DP* is much more poisonous than *IP*.

I think we have relatively robust intuitions that *IP* and *DP* are roughly equally poisonous: after all, they are *only* different in that one turns people purple. Does this really make it more poisonous? It would be hard to swallow the consequence that they are not equally poisonous, but this seems to be a result for any analysis taking the tolerant strategy.

Thus, it seems that the antidote cases in which poisonousness is masked are *not* relevant to ‘poisonous’-ascriptions to *DP* and *IP*. This example provides good *prima facie* evidence for the broader claim that masking and fink cases are generally irrelevant to disposition ascriptions and should not be relevant to the analysans.

For these reasons, my sympathies are with the exclusionary approach. However, I’m not going to lean on the exclusion thesis as a premise in the arguments against the analyses below; it will become clear that the failure to exclude masks and finks from relevant test-possibilities can be exploited to produce *prima facie* counterexamples (like the *IP/DP* case) to non-exclusionary analyses.

I’m going to turn now to some recent analyses of dispositional predicates¹⁶ that have been proposed in the revisionary period since the counterexamples. The first family of analyses (§4) pursue the exclusionary approach and retain the counterfactual conditional as the central mechanism of analysis. It is generally agreed that finks, masks, and mimics cannot be excluded negatively with a clause in the conditional antecedent specifying that the object is tested in the absence of finkish/masking/mimicking circumstances, so I won’t consider this

¹⁶ Some of the analyses target *explicit disposition attributions* (‘*x* is disposed to *R* when *T*’d’) rather than dispositional predicates; though I am strictly most interested in potential extensions of these analyses to dispositional predicates, my criticisms will be on their own turf (i.e. explicit disposition attributions).

strategy.¹⁷ The first of these analyses I'll consider (in §4a) practices *exclusion by precisification*, following David Lewis's suggestion that masks and mimics could be excluded from the analysans by revising the rough-and-ready associated conditions with precisifying corrections. The second (§4b) practices *exclusion by qualification*, appending a qualification like 'in normal conditions' or 'in ideal conditions' to the counterfactual conditional, motivated by the idea that finks/masks/mimics only occur in *abnormal* or *non-ideal* conditions. After working through some issues with each of these accounts, I'll present in detail a problem for both of them in §4c. I'll then consider one final conditional account in §4d, which Sanford Shieh has suggested to me avoids the problem presented in §4c. In §4e, I'll consider some general issues with the theoretical utility of these accounts.

The second set of analyses eschew the Lewis conditional for other logical machinery which is *tolerant* of fink/masking/mimicking cases as exceptions. Michael Fara (§5) appeals to habitual sentences and David Manley and Ryan Wasserman (§6) appeal to direct less-than-universal quantification over possibilities in analyzing disposition ascriptions. I'm going to argue that none of these analyses is extensionally adequate, providing counterexamples in each case.

4. The Lewis conditional analyses

a. The precisified conditional analyses

Though the main attraction of David Lewis's 1997 paper 'Finkish dispositions' is his *reformed conditional analysis*, which is intended to fink-proof the (SCA), I'm interested here in his strategy for excluding masks and mimics, which involves what has been called the *two-step analysis* of dispositional predicates.

¹⁷ See for instance Martin 1994, p. 6 and Bird 1998, p. 231 for criticisms – the upshot is that such analyses are circular, since identifying which cases count as finks, masks, and mimics depends upon an antecedent understanding of the dispositional predicate being analyzed. However, see Mellor 2000 for a discussion of the merits that such an analysis, though circular, might have.

A dispositional concept is the concept of being disposed to give such and such response to such and such stimulus. So the first problem we face in analyzing any particular dispositional concept, before we can turn to the more general questions that our particular example was meant to illustrate, is the problem of specifying the stimulus and the response correctly.¹⁸

Rather than directly analyzing dispositional predicates like ‘fragile,’ ‘poisonous,’ and ‘soluble,’ Lewis suggests that we first draw an equivalence between each dispositional predicate and an *explicit disposition attribution* – a statement of the form ‘ x is disposed to R when T ’d’. Once each dispositional predicate has been associated with an explicit disposition attribution, the only remaining step is to give a general analysis of explicit disposition attributions. For example:

First step – ‘ x is fragile’ is true iff x is disposed to break when stressed;

Second step – ‘ x is disposed to R when T ’d’ is true iff $(x \text{ is } T\text{'d} \square \rightarrow x \text{ R's})$

All together – ‘ x is fragile’ is true iff $(x \text{ is stressed } \square \rightarrow x \text{ breaks})$

‘Finkish dispositions’ is primarily a *second step* paper: Lewis abandons the ‘simple’ second step thesis that x is disposed to R when T ’d iff $(x \text{ is } T\text{'d}) \square \rightarrow (x \text{ R's})$ for a fancy general analysis of explicit disposition attributions. While fink-exclusion is the target of Lewis’s fancy second step, he suggests that masks and mimics can be excluded in the *first step*, and it is this claim that concerns me here. Lewis does not carry the first step out in detail for specific dispositional predicates, but he suggests that there is no general impediment to doing so:

We might offhand define a poison as a substance that is disposed to cause death if ingested. But that is rough: the specifications both of the response and of the stimulus stand in need of various corrections. To take just one of the latter corrections: we should really say ‘if ingested without its antidote.’ Yet the need for this

¹⁸ Lewis 1997, pp. 152-153.

*correction to the analysis of 'poison' teaches no lesson
about the analysis of dispositionality in general.*¹⁹

The idea is simple: the analyses of 'poisonous' as 'disposed to cause death when ingested' and 'fragile' as 'disposed to break when stressed' are merely 'rough' and 'offhand,' and the masking and mimicking counterexamples arise only because of this roughness. If Lewis is right, then we should be able to weed the mask and mimic counterexamples out of our relevant possibilities by providing precisifications like 'if ingested without its antidote.'

Lars Bo Gundersen and Sungho Choi have recently developed accounts that aim to exclude not only masks and mimics but also finks via this precisification method, eliminating the need for Lewis's fink-proofing in the *second step*.²⁰ The criticisms that follow concern only the *first step*, so the *second step* differences between Lewis's reformed conditional analysis and Gundersen and Choi's precisified (SCA) will not be discussed here. I'll refer to these collectively as *precisified conditional analyses* (PCA).

The key agendum for precisified conditional analyses of dispositional predicates is to get down to brass tacks and carry out the *first step*: provide mask and mimic-excluding precise associated conditions for some dispositional predicates. The fact that Lewis doesn't bother to carry out the first step, and tells us that it 'teaches no lesson about the analysis of dispositionality in general' suggests that there shouldn't be any particular difficulty in producing these corrections.

Though ultimately I think that the task of producing these precisifications while remaining faithful to intuitions about disposition test scenarios and disposition manifestations is not only *not easy* but in fact *impossible* (as I'll argue in §4c), it will be useful (to illustrate how the precisification step might proceed, and to flag some serious difficulties that arise with the project) to dive into the nitty-gritty of producing the required precisifications. Since Lewis gets the ball rolling on the

¹⁹ Lewis 1997, p. 153.

²⁰ Gundersen 2002 and Choi 2008.

precisification of ‘poisonous’-associated conditions with his suggestion that ‘we should really say ‘if ingested without its antidote,’ I’ll pick up where he left off and attempt to draw an equivalence between ‘poisonous’ and a precise explicit disposition ascription.

What is wrong with the simple first step analysis of poisonousness as the disposition to cause death when ingested? It’s susceptibility to finks, masks, and mimics is one problem. Let’s focus on the masking cases. Quite often a poison is ingested but does not cause death because the causal process by which the poison causes death is prevented or interfered with by an antidote. Since both Lewis and his (PCA) descendants Gundersen and Choi propose *second step* analyses of explicit disposition ascriptions using counterfactual conditionals, we can explain why the masking cases are *prima facie* counterexamples to their analyses by looking to the semantics of the conditional. Cases in which $D(a) \& T(a) \& \neg R(a)$ are, by the counterfactual semantics, counterexamples to the analysis-schema $\lceil \Box(\Phi(x) \leftrightarrow \Psi(x) \Box \rightarrow \Pi(x)) \rceil$ (with Φ , Ψ , and Π taking D , T , and R as their values, respectively, and a assigned to ‘ x ’) by the strong centering schema $\lceil \varphi \& \neg \psi \rceil \vDash \lceil \neg(\varphi \Box \rightarrow \psi) \rceil$.

The (PCA)-adherent needs to choose test and response conditions which preclude the possibility of $D(x) \& T(x) \& \neg R(x)$ cases. The uncorrected *first step* analysis allows many such cases, e.g. those in which x poisonous, x is ingested, and x does not cause death because an antidote prevents the poison from causing death. Thus, Lewis’s first revision – x is poisonous iff x is disposed to cause death when ingested *without its antidote* – is intended to keep antidote cases out of the relevant test worlds. However, as Lewis acknowledges, he hasn’t finished the job: there are still $D(x) \& T(x) \& \neg R(x)$ counterexamples to the antidote-corrected analysis.

One class of remaining counterexamples exploits the problem of *dose*. In many ordinary cases, a poison is ingested without its antidote but fails to cause death because the quantity of poison ingested is not sufficient to cause death. We need another precisification which restricts the $T(x)$ -cases to those events in which a *sufficient dose* of a poison is ingested.

However, different poisons have different lethal doses, so it doesn't seem that we can just choose some value n and say that poisonousness = the disposition to cause death when at least n milligrams are ingested without antidote. After all, if we choose a very high n , the analysis will incorrectly predict that every substance – water, potatoes, acetaminophen, whatever – is poisonous, since just about every substance would cause death were a sufficiently large amount is ingested. If we choose a moderate or low n , we have the converse difficulty of not including *enough* events as $T(x)$ -cases. Suppose we say that poisonousness = the disposition to cause death when at least 10 milligrams are ingested without antidote. According to this analysis, a case in which 9 milligrams of botulinum is ingested is *not* a poisonous-test, but this doesn't seem right – ingesting 9 milligrams of botulinum without antidote will cause death, and this seems to be a perfectly ordinary manifestation of poisonousness. It would be ad hoc to exclude these small quantity cases from the poison-tests; intuitively they *do* count as poison-tests.

What it seems that we want, instead of a single poison-test dose threshold value n , is a *token reflexive* condition, like Lewis's antidote correction 'if x were ingested without *its* antidote': something like 'if x were ingested in *its* lethal dose.' However, to put such a condition to non-trivial use in our analysis, we can't just leave *lethal dose* unanalyzed, since *every substance* (water, potatoes, acetaminophen included) would kill were a lethal dose ingested without its antidote, on the intuitive conditional interpretation of a substance's 'lethal dose' as the dose at which that substance would kill were it ingested. This interpretation trivializes the analysis, resulting in poisonousness being a disposition that all substances have. So we need to secure a non-trivial interpretation of 'lethal dose.' But herein lies the problem: how could we possibly interpret 'lethal dose' in a way in which there are *no cases whatsoever* in which a substance is poisonous, is ingested in its lethal dose without antidote, and causes death? The analysis requires a specification of lethal dose on which there are no such cases, since even a single one constitutes a counterexample to the analysis. But what interpretation of 'lethal dose' could possibly do

this? ‘Lethal dose’ is typically characterized in terms of the distribution and frequency of lethal and non-lethal test-cases – e.g. the standard scientific measure of lethal dose is LD_{50} which is defined as the smallest dose which will cause death in 50% of subjects. We need a way of specifying $T(x)$ such that $D(x) \& T(x)$ suffices for $R(x)$ *every single time*, since a single $D(x) \& T(x) \& \neg R(x)$ case will be a counterexample to the analysis. There doesn’t seem to me to be any non-trivial way to interpret lethal dose such that it is *impossible* that someone ingests a lethal dose of a poison but doesn’t die.

The (PCA)-adherent has to find some way or other to specify *dose* in the testing condition for ‘poisonous’ that addresses the following three desiderata: [1] the dose specification should not be so high that substances which are not poisonous but which cause death when ingested in large quantities count as poisonous, [2] the dose specification must preclude the possibility of *any* $D(x) \& T(x) \& \neg R(x)$ cases, and [3] the dose specification can’t be *ad hoc* in the following sense: we can’t just pick a low dose in order to satisfy desiderata [1] and [2] but which gives the result, contrary to intuitions, that cases in which poisons cause death when ingested below that dose are excluded from the poison-tests. But desiderata [2] and [3] are in direct conflict: we want to count cases in which a tiny quantity of a highly lethal poison causes death as poison-tests, but to avoid the counterexamples, we need to exclude cases in which much larger quantities of less potent (but nonetheless poisonous) substances are ingested but do not cause death.

This tension is indicative of a general tension in the selection of cases which are to count as disposition tests (and disposition responses): we want to *exclude* the myriad cases of masks and finks and thus are driven toward *specificity* in characterizing test conditions, and yet, we want to *include* all of those cases which are intuitively disposition tests and responses, which drives us toward *generality* in characterizing test and response conditions. We can emphasize this tension with the metaphor of *modal space*. There is a certain portion of modal space that comprises the poison tests, which we’ll call *PT*. The Simple Conditional Analyst seeks to

designate *PT* by way of the condition ‘*x* is ingested’, but this picks out much too rough a portion of modal space, since it includes, for instance, fink and masking cases (which are intuitively not in *PT*). So the Precisified Conditional Analyst, seeing this roughness, attempts to shave off these undesirable bits of excess with her precise conditions. Unfortunately, the exception intolerance of the Lewis conditional means that the (PCA)-adherent needs to get rid of *every single* fink and masking case, since even a single $D(x) \ \& \ T(x) \ \& \ \neg R(x)$ case is a counterexample to the analysis. However, some of these undesirable bits of excess are so tricky to get rid of – for instance, the cases in which poisons are ingested but don’t cause death because the dose is too low – that it seems the only way to get the job done is to shave off *too much*. For instance, in eliminating our cases in which poisons are ingested but don’t cause death because the dose is too low, we might appeal to the precise condition ‘at least *n* milligrams of *x* is ingested’ for some small *n*. But now consider the cases in which $< n$ milligrams of some poison is ingested and causes death. These cases have now been cut out of the test-space by our precise conditions. But they are intuitively poison tests! After all, it seems clear that they are *manifestations* of poisonousness; but how could poisonousness manifest in cases in which it is not even tested? In short, the descriptions available to conditional analysts – vague or precise – don’t seem up to the task of *excluding* all of the fink and mask cases without *overexcluding* some relevant cases as well.

There are surely strategies I haven’t considered for getting around this problem with dose. I don’t intend this example to be decisive; I’m more interested in illustrating the sort of pressures that arise when we actually undertake the first step of Lewis’s proposed two-step analysis. These complications only exponentiate when considered in connection with myriad other factors unaccounted for by the (SCA)’s rough-and-ready associated conditions: various modes of poison administration (inhalation, absorption through the skin, etc.), what manners of causing death are admissible (not anything that causes death when ingested is poisonous), antidotes (and antidote dose-related complexity), the varieties and degrees

of harm that count as non-lethal manifestations of poisonousness, and so on. This exercise so far has only shown that the precisification step is quite difficult, and gives rise to problems which are not easily solved. Lewis's rhetoric suggests that he doesn't worry about the precisification step because he doesn't think it presents any interesting problems. As we have seen, whether or not it 'teaches a lesson about dispositionality in general', the precisification step is certainly not trivial.

Ultimately I think it can be definitively shown that the precisification step is in principle impossible to reconcile with intuitive truth conditions, but since the argument applies to *exclusion by qualification* as well, I'll first turn to this style of analysis, and save that argument for §4c.

b. The qualified conditional analysis

Though it is widely agreed that the (SCA) cannot be fink/mask/mimic-proofed with a negative qualification in the antecedent along the lines of 'in the absence of finks, masks, or mimics,' another strategy involving qualifying the conditional has been explored. Stephen Mumford, for instance, writes the following:

What is usually implied by a true disposition ascription is that there are background conditions, let us call these 'ideal conditions', in which such manifestations do follow from the stimulus. A disposition ascription thus invokes the following 'conditional conditional':

[Df_i] *if C_i, then (if Fx, then Gx)*

*where C_i represents the ideal conditions, F and G represent stimulus and manifestation respectively, and both conditionals have subjunctive force.*²¹

Wolfgang Malzkorn argues for *normal* rather than *ideal* conditions.²² Let's call the broad strategy under which both Malzkorn's and Mumford's analyses fall the *qualified conditional analysis* (QCA).

²¹ Mumford 1998, p. 88.

I'm not going to examine the (QCA) in much detail at all; I'd just like to get a rough idea how it is supposed to work and mention a few worries I have before turning to the argument against it and the (PCA).

In order to make sense of the (QCA) we need to understand how the qualifier is supposed to interact with the conditional. The reading that takes Mumford's comments at face value – i.e. something like 'normal/ideal conditions $(x) \square \rightarrow (T(x) \square \rightarrow R(x))$ ' – seems to raise the problem that x might have a disposition which it normally lacks, or lack a disposition which it normally has.

Suppose, for instance, that a certain wire is normally live, but currently, as a result of rolling blackouts – abnormal conditions! – is not live. Though the wire is not live, it is true that if the wire were in normal conditions, it would be live. Thus, if the wire were in normal conditions, if it were touched by a conductor, electricity would flow to the conductor.

To get around this problem, let's presume that 'normal conditions' never takes us to possibilities in which the target object is intrinsically different – it is the conditions *extrinsic* to the object which can either be normal or abnormal, ideal or non-ideal.

I have two additional concerns about the (QCA) worth mentioning here. The first is that it practices *exclusion by over-exclusion*: if things go right, the qualifier keeps finks and masks out of the test-possibilities, but it also keeps out lots of other intuitively relevant though 'abnormal'/'non-ideal' possibilities. Intuitively, a case may be a disposition test despite taking place in abnormal conditions; if aliens on some remote planet subject some extraterrestrial middle-sized dry goods to moderate physical stress, this is no less a fragility test for its abnormality.

Over-exclusive analyses have the following shortcoming: they run contrary to a certain highly natural account of the relation between relevant possibilities (in my sense of *relevant*), D-tests, and manifestations. On the highly natural account, [a] all and only the possibilities in which x is subjected to the D-test condition are the possibilities relevant to the truth of disposition ascription 'x is D' and [b]

²² Malzkorn 2000, pp. 457-458.

all of x 's D-manifestations are cases in which x has been subjected to the D-test condition. *Over-exclusive* analyses must commit to one of the following highly unintuitive theses: [c] some D(x)-tests are not relevant to whether or not x is D or [d] some manifestations of D by x are cases in which x is *not* subjected to the D-test condition.

The other worry I have about this analysis is that it does not seem to give the correct funny results for some *qualified disposition ascriptions*. Disposition ascriptions (both explicit disposition ascriptions and dispositional predicate ascriptions) can be qualified (e.g. 'Ryan is irascible on Mondays.'). Some qualifiers – most obviously, 'in abnormal conditions' and 'in non-ideal conditions' – take us away from normal conditions. But how are disposition ascriptions embedded in these qualifiers to be analyzed? The truth conditions of 'in abnormal conditions, x is fragile' and 'in abnormal conditions, in normal conditions, were x stressed it would break' are *prima facie* different (although it is a little hard to make sense of the latter). Here's an example to make the point vivid. In conditions in which humans lack lactase in their intestines –circumstances which are nomologically possible, though abnormal and non-ideal – milk is disposed to cause death when ingested. Thus 'In conditions in which humans lack lactase in their intestines, milk is disposed to cause death when ingested' is true. However, on the (QCA), this is analyzed as 'In conditions in which humans lack lactase in their intestines, in normal conditions, were milk ingested it would cause death.' The presence of multiple qualifiers make things a bit confusing, but if we use subjunctive conditionals as Mumford suggests, we get the following: (*lack lactase*-conditions $\square \rightarrow$ (normal conditions $\square \rightarrow$ (ingested(milk) $\square \rightarrow$ cause death(milk)))). The outside counterfactual takes us to the closest *lack lactase*-worlds, and then the middle conditional inside takes us to the *normal conditions*-worlds closest to these. But since (presumably) there aren't any *normal conditions* & *lack lactase*-worlds (since humans lacking lactase in their intestines constitutes a very abnormal condition), the *normal conditions* *qualifier* takes us to some \neg (*lack lactase*)-worlds, where the inside conditional does

not hold. The doubly qualified conditional is false, though the qualified disposition ascription is true.

I'm not sure either of the above two problems are particularly serious, but they are certainly hurdle for the (QCA).

The above should give us a sense for the thrust of the (PCA) and (QCA): now let's consider a problem for both.

c. The poison problem

The motivating idea of these Lewis conditional accounts is that the (SCA)'s extensional inadequacy stems from fink and mask cases, and that this could be remedied by being a little more careful about specifying the test and response conditions. Lewis's two-step strategy and the qualification strategy have such an allure because they hold an elusive promise: if only we were to supply the *correct precisifications* or the *correct qualifications* we could sidestep the counterexamples. Almost any counterexample or difficulty can be chalked up to a *particular* precisification or qualification, leaving the general strategy intact.

Even if we could produce precisifications or qualifications which keep fink and masking cases out of the test possibilities, the two analyses still seem to fall prey to the *closeness counterexamples*. Manley and Wasserman argue this point at length in their 2008 paper 'On linking dispositions and conditionals,' so I won't pursue it in detail here. It just bears mentioning that the (PCA) and (QCA) have yet to address the closeness counterexamples, which are unaffected by the qualification or precisification of associated conditionals.

In addition to these two familiar counterexamples, I'd like to present a new problem that I think weighs seriously against both sorts of the conditional accounts. The problem does not appeal to fink, mask, or mimic cases at all. An extensionally adequate conditional account requires eliminating not only fink and masking $D(x) \ \& \ T(x) \ \& \ \neg R(x)$ cases, but 'ordinary' $D(x) \ \& \ T(x) \ \& \ \neg R(x)$ cases (which Manley and Wasserman call 'reverse Achilles heels').

Consider the case of ‘fragile.’ The problem begins with the following platitudinous observation: fragile objects need not give the fragility-response in every intuitive test case. By ascribing fragility, one does not ascribe the disposition to break in response to every possible stress whatsoever. A fragile glass may be disposed to break in response to a diverse variety of stresses, yet fail to break when stressed *just so*. These sorts of cases – e.g. when a fragile object is stressed *in just the right way so it doesn’t break* – appear to be straightforward $D(x) \ \& \ T(x) \ \& \ \neg R(x)$ counterexamples to any viable conditional analysis. However, the instinct behind the precisified and qualified conditional analyses is to find a principled way to exclude such cases from the relevant possibilities. What I’d like to argue here is that, for several paradigmatic dispositional predicates, this strategy is demonstrably irreconcilable with the intuitive truth conditions of the disposition ascriptions being analyzed.

When we consider just a *single* object and its dispositions, it is easy to fudge the conditions in a way that excludes these ordinary failed D-tests – ‘If this very fragile vase doesn’t break in response to these stresses *a*, *b*, and *c*, then I suppose they must not be fragility-tests.’ However, if we look at *multiple objects* which have *contrary* reverse Achilles’ heels, we are unable to exclude all of the objects’ reverse Achilles’ heels test conditions while maintaining a plausible antecedent condition. Keeping in mind multiple objects with contrary reverse Achilles’ heels helps illustrate that this method of overexclusion is not only ad hoc but unworkable: it can be shown that any antecedent condition which excludes all $D(x) \ \& \ T(x) \ \& \ \neg R(x)$ cases from a Lewis conditional analysis will have properties which make it unsuitable to capture the intuitive truth conditions of the disposition ascriptions being analyzed.

I’ll give an unrealistic but illustrative example of the problem for the term ‘poisonous,’ and then follow with a version of the same problem for ‘fragile.’

Poisons can be administered in a number of different ways (not just via ingestion). A substance that causes death via a certain mechanism when *inhaled* is no less a poison than one which causes death via a certain

very different mechanism when *ingested*. But not every poison kills via every mode of administration! Here's an unrealistic example: suppose we have two substances, *IngestoVenom* and *DermaToxin*, which are quite different poisons. If even a drop of *IngestoVenom* is ingested, it will kill instantly; however, external dermal exposure to *IngestoVenom* is completely harmless. *DermaToxin* is just the opposite: you can drink it like water without a tummy ache, but a tiny drop on the finger is fatal.

How are we going to analyze 'poisonous' with a single Lewis conditional (for a proposal that appeals to a disjunction of Lewis conditionals rather than a single conditional, see §4d) in a way that correctly analyzes both *IngestoVenom* (*IV*) and *DermaToxin* (*DT*) as poisonous? The consequent isn't any particular trouble – 'x causes death', which I'll abbreviate $K(x)$ – but what antecedent condition $C(x)$ could we possibly provide such that $C(IV) \Box \rightarrow K(IV)$ and $C(DT) \Box \rightarrow K(DT)$? The poison problem shows that any antecedent condition that satisfies the requisite conditionals (a necessary condition for an extensionally adequate account) will correspond to a grossly unintuitive analysis of the target disposition ascription.

Here's how the problem goes. Assume that C is a condition such that both conditionals, ' $C(IV) \Box \rightarrow K(IV)$ ' and ' $C(DT) \Box \rightarrow K(DT)$ ', hold. Assume further that the 'closest' intuitive poison test cases for *IV* are *administered dermally(IV) & ¬ingested(IV)*-worlds and the closest intuitive poison tests for *DT* are *ingestion(DT) & ¬administered dermally(DT)*-worlds. In other words, the closest intuitive poison tests for *IV* and *DT* are their reverse Achilles' heels. It is important to emphasize that there is nothing illicit about this assumption; it is just a further description of our hypothetical possibility. Suppose, for instance, that in some possible state of the world w , people quite commonly, and with the greatest care, apply *IV* dermally because of its moisturizing effects on the skin, and drink *DT* as a laxative.

Now consider what sorts of cases count as satisfying condition C – i.e., according to our hypothetical analysis, which cases are poison tests? We don't want these nearest *administered dermally(IV) & ¬ingested(IV)*

and *ingestion*(DT) & \neg *administered dermally*(DT) cases to satisfy C, since these are cases in which the disposition doesn't manifest (and would thus be $D(x)$ & $T(x)$ & $\neg R(x)$ counterexamples). What we want is for *administered dermally*-cases to count as 'poisonous'-tests for DT and for *ingested*-cases to count as tests for IV. But if the conditional account provides a qualitative condition C such that $C(x) \square \rightarrow K(x) \leftrightarrow D(x)$, then it should be the case that, for any test condition *T*, $T(a)$ will satisfy condition C just in case $T(b)$ satisfies condition C. This is the further assumption that a certain test scenario '*x* is tested in manner *T*' either satisfies condition C or not, independently of the object assigned to the variable. Nothing prevents the conditional analyst from using some token-reflexive condition C which violates this principle, in the sense that cases in which different objects in otherwise qualitatively identical situations may diverge in whether or not they satisfy C. However, no such account has been proposed (so far as I know) and intuitively such an account would seem to share many of the problems raised in the next section for a disjunctive account (since we can think of such a conditional analysis with a token-reflexive antecedent condition like a disjunctive analysis with different antecedents for different objects). For a conditional account which satisfies the above assumption, in order to prevent all $D(x)$ & $T(x)$ & $\neg R(x)$ counterexamples we are going to have to provide a condition for which there are no cases in which *DT* does not give the 'poisonous'-response in the condition and no cases in which *IV* does not give the 'poisonous'-response in the condition. But since *DT* manifests *only* in the *dermal application*-events and *IV* manifests *only* in the *ingestion*-events, a condition which satisfies the relevant conditionals must be such that *x* is ingested *and* *x* is dermally applied in all cases satisfying C. But any analysis of 'poisonous' in terms of a conditional with an antecedent like that will be grossly unintuitive, since the test condition for 'poisonous' certainly does not require simultaneous ingestion and dermal application.

Here's an example for the dispositional predicate 'fragile.' After I describe the case, I will attempt to schematize the argument to make it as

transparent as possible how the problem might apply to other dispositional predicates.

Suppose that there are two materials, *StrongStuff* and *WeakStuff*, which have the following properties: *StrongStuff* is practically indestructible, while *WeakStuff* crumbles in response to the slightest stress. Now suppose there is an object, which we'll call *LeftStrong*, which is composed entirely of *StrongStuff* on its left side and *WeakStuff* on its right side. *LeftStrong* is intuitively quite fragile – after all, it will crumble in response to the slightest stress on its left side! Suppose that there is also an object, *RightStrong*, which is composed of *StrongStuff* on the right side and *WeakStuff* on the left side. Both *LeftStrong* and *RightStrong* are fragile, but they manifest fragility in very different circumstances. Cases in which *LeftStrong* is stressed on the left side and in which *RightStrong* is stressed on the right side appear to be troublesome $D(x) \ \& \ T(x) \ \& \ \neg R(x)$ cases, since they are intuitive fragility-tests, the objects are fragile, and the objects do not give the fragility-response in the cases. This puts the following constraint on an extensionally adequate conditional analysis: the antecedent of the conditional can never take a *LeftStrong*-like object to a *left side is stressed*-type fragility test, and can never take a *RightStrong*-like object to a *right side is stressed*-type fragility test. But every possible stress test is either an event in which an object is stressed on the left side only, or the right side only, or both sides. If some cases of being stressed on the left side only or the right side only count as fragility tests, then presumably some cases in which a fragile object with *LeftStrong*-like properties is stressed on the left will count as a fragility test and will stand as a counterexample (and *mutatis mutandis* for *RightStrong*). Thus, it appears that any viable antecedent in a conditional analysis of 'fragile' must satisfy the following constraint: only events in which an object is stressed *both* on the left and the right side count as fragility-tests. But this is a constraint that no plausible conditional account satisfies, for requiring stress on both sides diverges greatly from the intuitive notion of a disposition test for fragility.

Here is the general form of the argument: suppose that there is a disposition *D* (fragility), object property complex \mathbf{P}_1 (being composed of *StrongStuff* on the left side and *Weak Stuff* on the right side), and event property complex \mathbf{E}_1 (being an event of stress to the left side of an object) such that [a] for any object, having properties \mathbf{P}_1 entails being *D* (being composed of *StrongStuff* on the left side and *WeakStuff* on the right side entails being fragile), [b] an event having \mathbf{E}_1 entails that event being an intuitive *D*-test (being an event of stress to the left side of an object entails being a fragility-test), and [c] an object with \mathbf{P}_1 would not give the *D*-response to a *D*-test with \mathbf{E}_1 (an object composed of *StrongStuff* on the left side and *Weak Stuff* on the right side would not break were it stressed on the left side). Such a scenario puts the following constraint on antecedent condition *C* for a conditional analysis: $C(x)$ can never take an object x : $\mathbf{P}_1(x)$ to a *D*-test e : $\mathbf{E}_1(x)$, since such a case would constitute a $D(x) \ \& \ T(x)$ & $\neg R(x)$ case. This already requires *overexclusion*: events which are \mathbf{E}_1 are intuitive *D*-tests but cannot be counted as *D*-tests by the analysis on pain of counterexamples. We can make the pressure more acute, however, by considering *contrary cases*, like the *LeftStrong* and *RightStrong* cases, where we have an *exhaustive* set of event property complexes $\mathbf{E}_1, \mathbf{E}_2 \dots \mathbf{E}_n$ (being stressed on the left side only, right side only, or both sides) such that each *D*-test satisfies at least one of the properties and at most one of the properties, and all but one of the properties can be shown to be incompatible with the manifestation of the disposition by an object with the ‘corresponding’ property from the set $\mathbf{P}_1, \mathbf{P}_2 \dots \mathbf{P}_n$. By doing so, we have placed a constraint on antecedent condition *C* – e.g. that $C(x)$ must always take us to events that have the property of involving stress on both the left and right side of the target object. But this property that the antecedent of any adequate account must have is a grossly unintuitive and over-restrictive one: no plausible analysis of ‘fragile’ could possibly require that fragile tests all involve stress on both the left and right side of an object.

This argument depends crucially on the multiple realizability of fragility and poisonousness in objects with quite different categorical bases

and quite different distributions of manifestations among test events. Perhaps not all dispositional predicates can be correctly ascribed to objects with such variegated patterns of D-responses; however, it is a basic criterion of adequacy for an analysis of dispositional predicates that it should be able to handle such paradigmatic instances as ‘fragile’ and ‘poisonous.’ For this reason, I take the problem to militate against any Lewis conditional analysis of dispositional predicates. The poison problem is not necessarily a problem for conditional analyses of any dispositional predicate whatsoever – particularly for more natural dispositions (e.g. ‘soluble’) for which the distribution of D-responses seems far less gerrymandered.

The conditional analyst may try to rescue her account by appealing to context-sensitivity. In this section I have targeted a context-insensitive version of the conditional analyses; I will turn to context-sensitivity and dispositional predicates in Part II.

d. A disjunctive conditional analysis

Sanford Shieh has suggested to me a variation on the Lewis strategy which avoids the poison problem. The suggestion is that a dispositional predicate like ‘poisonous’ should not be analyzed using a *single* counterfactual conditional, but rather via a *disjunction* of *many* counterfactual conditionals.

The proposal grew out of the poison problem as follows. What I aim to show with the poison problem is that for any poison, though some conditional like ‘(x is ingested) $\square \rightarrow$ (x causes death)’ or ‘(x is inhaled) $\square \rightarrow$ (x causes death)’ will presumably hold, different conditionals of this sort will hold for the different poisons (depending on which modes of administration cause death). For this reason, providing a satisfactory *general* antecedent condition C for which ‘C(x) $\square \rightarrow$ (x causes death)’ holds for any poison whatsoever appears not to be possible.

Shieh’s suggestion is that we should rest content with these more specific conditionals, which plausibly track more *natural* properties – being poisonous_{inhaled} (i.e., roughly, being disposed to cause death when

inhaled), being poisonous_{ingested}, and so on. On his proposed analysis, x is poisonous *simpliciter* iff $(\exists m)(x \text{ is poisonous}_m)$ where the m ranges over something like modes of administration, and x is poisonous _{m} iff $m(x) \Box \rightarrow (x \text{ causes death})$. Since the proposal can also be expressed with the disjunction ‘ x is poisonous iff $(m_1(x) \Box \rightarrow (x \text{ causes death})) \vee (m_2(x) \Box \rightarrow (x \text{ causes death})) \vee \dots \vee (m_n(x) \Box \rightarrow (x \text{ causes death}))$ ’, let’s call this strategy the *disjunctive conditional analysis* (DCA).

As far as I know, no proposal like this has been defended in the literature on the semantics of dispositional predicates, although the proposal may sound familiar since many philosophers have defended a similar metaphysical claim – viz. that dispositions are *second-order properties* that apply to an object x just in case x has at least one of some set of first-order properties. Though the semantic analysis and the metaphysical claim are independent of each other, they do go well together: it is natural to pair the claim that dispositional properties are second-order properties with a ‘second-order’ semantic analysis according to which x is D iff at least one counterfactual (corresponding to a first-order property) in a big disjunction holds.

The first thing to observe is that almost all of the arguments against the other conditional accounts can be run against the (DCA) as well. The poison problem is exceptional in this regard. Like the (PCA) and (QCA), the (DCA) does nothing to address *the closeness and centering counterexamples*.

It will be useful to begin with the question of exception tolerance: recall that any actual or possible case in which $D(x) \ \& \ T(x) \ \& \ \neg R(x)$ provided a counterexample to the analysis ‘ $D(x)$ iff $T(x) \ \Box \rightarrow \ R(x)$ ’. For our analysis

$$\begin{aligned} \text{‘}x \text{ is } D\text{’ is true iff} \\ (T_1(x) \ \Box \rightarrow \ R(x)) \ \vee \\ (T_2(x) \ \Box \rightarrow \ R(x)) \ \vee \ \dots \ \vee \\ (T_n(x) \ \Box \rightarrow \ R(x)) \end{aligned}$$

a case in which $\neg(T_n(x) \ \Box \rightarrow \ R(x))$ for a single n will not constitute a counterexample, since there may still $(\exists m)(T_m(x) \ \Box \rightarrow \ R(x))$. A

counterexample will have to show that either that $D(a) \ \& \ \neg(\exists m)(T_m(a) \ \square \rightarrow R(a))$ (analysandum holds, analysans does not) *or* that $\neg D(b) \ \& \ (\exists m)(T_m(b) \ \square \rightarrow R(b))$ (analysans holds, analysandum does not).

The tight-rope act of balancing *generality* and *specificity* mentioned in §3a primarily arises in the (DCA) with the choice of how fine-grained we want our disjuncts to be – do we go for a longer disjunction with more fine-grained conditions, or one with fewer disjuncts and ‘broader’ conditions? The poison problem forced us from a single testing condition to a disjunction of multiple testing conditions, so our disjunct-T’s should at least be fine-grained enough to avoid the poison problem! For example, if we subsume the dermal exposure and ingestion testing conditions under a single conditional disjunct, we can run the poison problem with DermaToxin and IngestoVenom just as before. We haven’t chosen fine-grained enough test conditions to avoid the poison problem.

On the other hand, if we get too fine-grained with the testing conditions in the disjuncts, we open ourselves up to $\neg D(b) \ \& \ (\exists m)(T_m(b) \ \square \rightarrow R(b))$ counterexamples. For instance, in sketching the poison problem, in the ‘poisonous’ case I appealed to poisons which differ in terms of which *modes of administration* were lethal. In the parallel ‘fragile’ case, I appealed to objects which differ in terms of which *varieties of stress* resulted in breaking. Suppose then, that our disjunctive analysis of ‘fragile’ uses extremely fine-grained test conditions like ‘x is stressed in *just this way*’ (e.g. dropped from such-and-such a height onto such-and-such a surface). There will certainly be objects which are not the least bit fragile but for which one or even a few of these very specific conditionals hold (Manley and Wasserman call these ‘Achilles’ Heels’). The disjunctive analysis is quite susceptible to this sort of counterexample, since satisfaction of even one disjunct of the very long disjunction by an indisposed object will suffice.

The difficulty we face in choosing which disjuncts exactly we want in the analysans should not be underestimated. The case of ‘poisonous’ is quite felicitous for the (DCA). ‘Modes of administration’ – whatever exactly these might be – are the sort of thing such that, for any poison,

there will be at least one mode of administration such that (modulo defeating conditions like antidotes) if the poison were to be administered in that way, it would cause death, and *this suffices for object to be poisonous*. That is, ‘poisonous’ suggests a natural choice of disjunct antecedents – the modes of administration – since satisfying any one of the disjuncts suffices for being poisonous. Compare, however, ‘fragile’ and ‘modes of stress.’ Anything which is fragile surely would break in response to some mode of stress or other, but breaking in response to a single mode of stress – unlike causing death via a single mode of administration – certainly does *not* suffice for having the disposition. Intuitively, an object is only fragile if it breaks in response to a sufficient *variety* of modes of stress. ‘Poisonous’ is uniquely well-suited to a disjunctive analysis: tell me that *one specific conditional disjunct* like ‘(a is ingested) $\square \rightarrow$ (a causes death)’ holds and I can tell you that *a* is poisonous. But telling me that there’s a mode of stress M_1 for which ‘ $M_1(a) \square \rightarrow$ (a breaks)’ holds won’t in general suffice as evidence for the claim that *a* is fragile.

For this reason, it isn’t at all clear how to run the disjunctive conditional account for dispositional predicates other than ‘poisonous’. The most natural suggestion to my mind involves each disjunct being comprised of not a single counterfactual conditional, but a *conjunction* of counterfactual conditionals. We can imagine running this sort of analysis on ‘fragile’ where the modes of stress are symbolized with ‘S’ and ‘x breaks’ is symbolized ‘B(x)’:

$$\begin{aligned}
 &\text{‘x is fragile’ is true iff} \\
 &(S_{1,1}(x) \square \rightarrow B(x) \wedge S_{1,2}(x) \square \rightarrow B(x) \wedge \dots \wedge S_{1,n}(x) \square \rightarrow B(x)) \vee \\
 &(S_{2,1}(x) \square \rightarrow B(x) \wedge S_{2,2}(x) \square \rightarrow B(x) \wedge \dots \wedge S_{2,m}(x) \square \rightarrow B(x)) \vee \dots \vee \\
 &(S_{p,1}(x) \square \rightarrow B(x) \wedge S_{p,2}(x) \square \rightarrow B(x) \wedge \dots \wedge S_{p,q}(x) \square \rightarrow B(x))
 \end{aligned}$$

The gist of the proposal is this: having a single mode of stress S such that ‘ $S(a) \square \rightarrow B(a)$ ’ holds does not suffice for *a*’s being fragile, but having a whole bunch of different modes of stress $S_{i,i}$ for which all of the conditionals ‘ $S_{i,i}(a) \square \rightarrow B(a)$ ’ hold does suffice for *a*’s being fragile. There may not be any particular mode of stress for which breaking in that mode of stress is necessary or sufficient for being fragile: rather, being fragile is a

matter of breaking in response to a sufficiently large and diverse plurality of modes of stress.

It still isn't clear how to extend this analysis to other dispositional predicates. I have trouble conjecturing with any generality what a (DCA) adherent will want her disjuncts to look like, bearing in mind the variegated array of dispositional predicates from 'soluble' and 'poisonous' to 'brave,' 'irascible,' and 'loquacious.' I think it's improbable that a single (DCA) could be carried out in practice for any dispositional predicate – even 'poisonous' – given the immensity of the project of providing all of the disjuncts and the complications that arise in connection the issues with dose, antidotes, etc. (see §4b).

It's worth mentioning in particular the threat of $\neg D(b) \ \& \ (\exists T)(T(b) \ \Box \rightarrow R(b))$ counterexamples posed by mimics, accidental-closeness, and strong-centering. Presumably the (DCA) will use (PCA)-style precise conditions or (QCA)-style qualified conditionals to exclude finks, masks, and mimics (although I think we have seen that, independently of the poison problem, it is still not clear either of these strategies get the job done). But the (DCA) appears to be even *more* susceptible to closeness counterexamples: rather than a single conditional 'target' $T(x) \ \Box \rightarrow R(x)$ for which an actual $\neg D(x) \ \& \ T(x) \ \& \ R(x)$ case will suffice as a counterexample, we have *a whole bunch* of conditionals $T_n(x) \ \Box \rightarrow R(x)$ for which *any* $\neg D(x) \ \& \ T_n(x) \ \& \ R(x)$ case will suffice as a counterexample. In other words, while going disjunctive avoids the poison problem, it makes *analysans holds/analysandum does not hold* counterexamples easier to come by, loosening the requirement for a counterexample from $\neg D(x) \ \& \ T(x) \ \& \ R(x)$ to $\neg D(x) \ \& \ (\exists T)(T(x) \ \Box \rightarrow R(x))$ (with the quantifier ranging over the different counterfactual antecedents in all of the disjuncts). For this reason, I think we should be quite skeptical about the (DCA)'s prospects for extensional adequacy.

e. Other semantic considerations

I think it is also worth bringing some different sorts of criticisms against the complex conditional accounts, since these accounts reflect an

unfortunate tendency to chase extensional adequacy while neglecting myriad other desiderata for semantic theories.

Here are the sorts of questions to which we might want to put to use a semantic theory of dispositional predicates: What is involved in lexical competence with specific dispositional predicates? What is it that unites fragility or poisonousness as a conceptual category? How do we account for and model the vagueness and context-sensitivity of dispositional predicates? How do we analyze explicit disposition ascriptions which do not specify test conditions ('Mary has a disposition to get wild,' 'John has a disposition to boogie,' 'Max is disposed to blow up in people's faces')?

The (SCA) was a proposal that had natural applications to these questions which appeal to rough-and-ready associated conditions. Perhaps the vagueness and context-sensitivity of a dispositional predicate track the vagueness and context-sensitivity of the particular associated conditions. The (SCA)-adherent can point out how particular examples of the context-sensitivity of 'fragile' seem to track the context-sensitivity of 'is stressed' and 'breaks' (see Part II §2). She can tell a story about lexical competence that involves mastery of the associated conditions and our general capacity for counterfactual reasoning. Her theory has internal mechanisms that can be applied to a number of questions about the semantics of dispositional predicates.

However, the move from the (SCA) to the (PCA), from the (PCA) to the simple 'poison'-style (DCA), and from the 'simple' (DCA) to the even more complex disjunction-of-conjunctions (DCA) represents a slippery slope of diminishing theoretical utility.

The first point to observe is that whereas with the (SCA) we had a *specific analysis of a particular dispositional predicate* – 'x is fragile iff stressed(x) $\square \rightarrow$ breaks(x)' – with the (PCA) and (DCA) we have more abstract analysis-schemata: for some remarkably difficult choice of precise conditions, some counterfactual with conjunctions of those on each side of the conditional (in the case of the (PCA)), or disjunction of conditionals (in the case of the 'simple' (DCA)), or disjunction of conjunctions of conditionals (in the case of the 'fancier' (DCA)) will provide an

extensionally adequate analysis of a particular dispositional predicate. No longer can the conditional analyst appeal to the specific conditions in her explanations of the above phenomena. The (PCA) and (DCA) adherents haven't given any specific conditions in their analyses, and the prospects for doing so are slim given the incredible complexity in filling out the analysis-schemata with a specific choice of conditions which excludes masks, mimics, and finks, keeps intuitively relevant possibilities relevant, and addresses all of the complexities particular to a certain dispositional predicate (e.g. the dose, antidote, mode of administration worries for 'poisonous' I mentioned in §4b).

Whereas the natural, plausible picture of lexical competence on the (SCA) appealed to a general capacity for counterfactual reasoning and knowledge of two vague associated conditions, the analogous picture on the (PCA) or (DCA) is one which requires that speakers know huge conjunctions or disjunctions or conjunctions of disjunctions of very precise conditionals. Though not indefensible, this explanation of lexical competence seems to be far less plausible and cognitively economical than the one available to the (SCA)-adherent.

Nor do these accounts, with their *precise* conditions, account for the *vagueness* of dispositional predicates.

Nor do these accounts offer a particularly plausible account of the *conceptual unity* of dispositions. *Fragility*, for instance, seems to be a concept which categorizes certain entities which have something in common; but what fragile objects have in common does not seem to be the disposition to give a highly *specific* response to a highly *specific* test (given that not only vases and cups but also persons, ecosystems, nation-states, and so on can be fragile). The (DCA) suffers even more on the score of conceptual unity, since two objects which are fragile on this analysis may share nothing in common on the analysis: one satisfies one disjunct, while the other satisfies a different disjunct, with the only commonality being that both disjuncts are included in the long disjunction.

So much the worse for the prospect of a semantic theory of dispositional predicates, the (PCA) and (DCA) adherents may say. Perhaps

the extensional failure of the (SCA) and need to regiment it with precise conditions and complex constructions is indicative of a lack of unity among test- and response conditions, a lack of a characterizable mechanism for context-sensitivity, and a lack of a simple account of lexical competence for dispositional predicates. The complex accounts may still have some explanations to offer. Shieh has suggested to me that the (DCA)'s explanatory potential seems to be much greater in connection with the metaphysical problems connecting dispositional properties and their (presumably disjunctive) causal bases than in giving semantic explanations of the phenomena mentioned above. But the topic here is the semantics of dispositional predicates, and connecting dispositions to causal bases is not part of that project.

In response to this pessimism about the prospects for a widely applicable, explanatory semantic theory of dispositional predicates, I only offer a promissory note here: I believe an analysis of dispositional predicates can achieve extensional adequacy while still giving informative explanations of a class of semantic phenomena particular to and characteristic of dispositional predicates, and I believe that I provide such an analysis in Part III of this paper. This analysis is (I'll argue) both extensionally adequate *and* provides the groundwork for a theory with applications to a wide range of issues concerning dispositional predicates (lexical competence, vagueness, context-sensitivity, conceptual unity). We do not have accept theoretical inutility in chasing extensional adequacy. But we are still a long way away from a complete understanding of the data to be explained by such a theory; I hope to present and analyze these data throughout Part I and Part II.

It's worth summarizing some lessons learned from the initial consideration of the extensional inadequacy of the conditional analyses. The (SCA) is susceptible, we observed, to two particular sorts of counterexamples: finks, masks, and mimics on the one hand, which intuitively are not relevant test cases (in the case of finks and mask) and not manifestations (in the case of mimics), and accidental closeness and strong-centering counterexamples on the other, which intuitively *are*

relevant test possibilities, but are *exceptions* – $D(x) \ \& \ T(x) \ \& \ \neg R(x)$ or $\neg D(x) \ \& \ T(x) \ \& \ R(x)$ possibilities which, though relevant, are not indicative of the broader pattern of x 's behavior in response to the disposition test outside the fluky inner sphere.

The complex conditional accounts were proposed to exclude finks, masks, and mimics. We observed that both the precisification and qualification strategies appeared to face an uphill battle in excluding finks, masks, and mimics, and compounded the problems of specifying the test- and response conditions besides. We thus have no adequate proposal available for moving from the vague test- and response- conditions of the (SCA) to a more perspicuous specification of T and R that excludes finks, masks, and mimics. None of the more complex analyses proposed a response to the accidental closeness and strong centering counterexamples, which arise because the possibilities relevant to the counterfactual conditional comprise a relatively small portion of modal space (the inner sphere of T-worlds). This is intuitively a much smaller portion of modal space than is relevant to the disposition attributions, which suggests that analyses which abandon the Lewis conditional or revise its semantics to take a broader range of possibilities (beyond the inner sphere) to be relevant may harbor greater potential. Perhaps the most important lesson came from the poison problem, in which we observed that providing an antecedent which excludes *all* possible $D(x) \ \& \ T(x) \ \& \ \neg R(x)$ cases could only occur in an analysis with a grossly unintuitive antecedent condition, e.g. one requiring that fragility tests must involve stress on both sides of an object. We also observed that the move from the (SCA) to more complex analysis schemata like the (PCA) and especially the (DCA) involved a loss of theoretical utility, since the resources internal to the complex theories (myriad unspecified precise conditions + lots of logical machinery) are less suited to explanation of phenomena like lexical competence, vagueness, context-sensitivity, and conceptual unity than the simpler resources of the (SCA) (rough-and-ready associated conditions + minimal logical machinery). This provides a sketchy agenda for semantic theorists interested in dispositional

predicates: find a way to preserve the theoretical and explanatory utility of the (SCA) while excluding finks, masks, and mimics from relevant possibilities and including in the relevant possibilities D-tests outside the inner sphere. My account in Part III will roughly follow this proposed agenda. However, I'll first turn to two recent analyses of dispositional predicates that go in a different direction, abandoning the exclusionary approach and Lewis conditionals for fink/mask/mimic-tolerance using habitual sentences (§5) and proportional quantification over test- and response-satisfying possibilities (§6).

5. Fara's habitual analysis

In recent years, some philosophers have moved away from conditionals as a mechanism for the analysis of dispositional predicates and disposition attributions. For instance, here's Michael Fara:

Conditionals are simply not suited to the task of stating the truth conditions of disposition ascriptions. It is a constitutive feature of conditionals that modus ponens is a valid rule of inference for them. Yet what the phenomenon of masking shows is that an analogous rule of inference is not valid for disposition ascriptions: from the fact that N is disposed to M when C, together with the fact that C, it does not follow that N M's.²³

In these next two sections I'm going to examine two recent proposals that use non-conditional machinery: Michael Fara's habitual analysis, and David Manley and Ryan Wasserman's proportional account.

Michael Fara has proposed an analysis according to which disposition ascriptions are constructions in which an unpronounced dispositional operator embeds a habitual statement. Fara first observes that explicit disposition ascriptions – 'x is disposed to R when T'd' – belong to a class of constructions all of which embed infinitival complements: 'x is reluctant to R when T'd,' 'x is happy to R when T'd,' 'x is able to R when T'd,' 'x is required to R when T'd,' 'x is known to R when

²³ Fara 2008, p. 61.

T'd,' and so on (each of which embeds the infinitival complement *to R when T'd*).²⁴ Fara then appeals to 'a widely accepted principle of syntax' – that infinitival clauses must have subjects – and a highly plausible principle about these subjects – that 'the implicit subject... is constrained to corefer with the main noun phrase' – to reach the conclusion that the truth conditions of 'x is disposed to R when T'd' can be represented as 'x is disposed: (x R's when T'd).' ²⁵

After establishing that 'the truth conditions of disposition ascriptions are stateable as 'N is disposed to M when C' is true iff DISP(N Ms when C)', where 'DISP' is the disposition operator embedding the habitual, Fara turns to the crucial remaining question: 'How should the operator DISP be interpreted?'²⁶ He first claims on the basis of a few examples that the operator is *factive*: DISP(x R's when T'd) entails x R's when T'd.²⁷ Here is how he motivates the claim:

*Consider how odd it would sound to assert the denial of an instance of the factivity of DISP. "I'm disposed to pay my taxes late, but I don't pay my taxes late"; "elastic bands are disposed to stretch when pulled, and they don't stretch when pulled." Sentences such as these do not just sound odd, or inappropriate; they sound trivially false, absurd. The factivity of DISP explains this.*²⁸

Fara then goes on to observe on one more feature of disposition ascriptions, which brings us to his complete analysis:

When we attribute to an object a disposition to do something, therefore, we are claiming not only that it does that thing, but also that the fact that it does it depends, in some substantial way (although not entirely), on how that object is. There is something about the object, and not just about its relations to other things, that makes it the case that it behaves in the way it does. To put the point another way: to attribute to an object a disposition to do so-and-so is to say not just

²⁴ Fara 2008, pp. 62-63.

²⁵ Fara 2008, pp. 62-63.

²⁶ Fara 2008, p. 69.

²⁷ Fara 2008, p. 69.

²⁸ Fara 2008, p. 69.

that it does so-and-so, but that it has some intrinsic property in virtue of which it does so-and-so. In full, the account of disposition ascriptions that I am proposing is:

The Habitual Account

“N is disposed to M when C” is true iff N has an intrinsic property in virtue of which it Ms when C.²⁹

Recall that one of the big problems with the Lewis conditional accounts was their failure to *tolerate exceptions* – a single case in which $D(a) \& T(a) \& \neg R(a)$ was sufficient to sink the analysis ‘ $\Box(D(x) \leftrightarrow (T(x) \Box \rightarrow R(x)))$ ’. The greatest strength of Fara’s account is that habituals tolerate exceptions. Neither ‘That vase is disposed to break when struck’ nor ‘That vase breaks when struck’ entails that in every possible circumstance in which the vase is struck it breaks. The exception-tolerance of habituals gives Fara’s account great promise.

However, I’d like to suggest that the account is still susceptible to a specific type of masking counterexample. The counterexamples violate the purported factivity of the disposition operator: they are cases in which an object has a disposition to R when T’d, but doesn’t R when T’d.

Here’s an example. Suppose Mark is disposed to smoke when he gets home from work. In fact, Mark is disposed to smoke at any time whatsoever. However, he has a very health-conscious cardiologist fiancée, Mary, who does not approve of his smoking. So whenever Mark arrives home from work and pulls out a cigarette to light up, Mary takes his cigarettes from him and throws them in the garbage. Mark can’t smoke without any cigarettes around and Mary keeps a close eye on him, so he doesn’t smoke when he gets home from work. The disposition ascription analysandum, ‘Mark is disposed to smoke when he gets home from work’ is true, but the analysans, ‘Mark has an intrinsic property in virtue of which he smokes when he gets home from work’ is false (since he *doesn’t* smoke when he gets home from work.)

This case is a counterexample to the habitual account just in case [1] the disposition ascription (‘Mark is disposed to smoke when he gets home

²⁹ Fara 2008, p. 69-70.

from work') is true and [2] the habitual ('Mark smokes when he gets home from work') is false. Thus, there are two ways to defend the analysis against the purported counterexample: deny the falsity of the habitual, or deny the truth of the disposition ascription. I'll consider each strategy in turn.

Denying the falsity of the habitual seems to be the less plausible strategy. It is quite counterintuitive to say that 'Mark smokes when he gets home from work' is true, since Mark *never* smokes when he gets home from work. As Fara points out, a habitual statement 'x R's' or 'x R's when T'd' may be true even though x has never R'd – e.g. 'Mary handles the mail from Antarctica' may be true even though Mary has *never* handled *any* mail from Antarctica (e.g. because there has never been any mail from Antarctica).³⁰ However, affirming the truth of 'Mark smokes when he gets home from work' requires something much stronger, viz. that 'x R's when T'd' may be true even though x *frequently* is T'd, but *never* R's when T'd. This seems highly implausible. We can compound this intuition by considering the fact that many true descriptions of the scenario seem to be incompatible with the truth of 'Mark smokes when he gets home': 'Mark doesn't smoke when he gets home,' 'Mark never smokes when he gets home,' 'Mary always prevents Mark from smoking when he gets home,' 'Mark never smokes when his cigarettes are in the trashcan, and his cigarettes are always in the trashcan when he gets home from work' and so on. So it doesn't appear to be a plausible way out of the counterexample to deny the falsity of 'Mark smokes when he gets home from work.'

The more plausible strategy is to deny the truth of the disposition ascription and claim that 'Mark is disposed to smoke when he gets home from work' is false. I think that as I've described the situation, this is already quite unintuitive, but I also think that we can again appeal to various considerations to compound this intuition.

In the first place, consider Mark's disposition to smoke (at any time). Intuitively, this is not a disposition he loses and gains throughout the day – I have described him as *always* being disposed to smoke. Thus,

³⁰ Fara 2008, pp. 65-66.

it seems implausible to say that he *loses* this disposition after work; he is just as disposed to smoke as ever, but intuitively a mask (Mary) prevents the disposition from manifesting. We might differentiate two different sorts of temporal disposition attributions: the temporally qualified disposition ‘*x* is disposed to R at regular time *t*’ (e.g. ‘Mark is disposed to smoke when he gets home from work’) and the temporally indexed simple disposition attribution, ‘at *t*, *x* is disposed to R’ (e.g. ‘at time *t* after work, Mark is disposed to smoke.’) The former case is of the form ‘*x* is disposed to R when T’d,’ where the T-condition is Mark arriving home from work. The latter is a disposition attribution without an explicit test condition – ‘*x* is disposed to R.’ I should be very clear that I intend the counterexample to involve the *former* ascription with the T-condition ‘Mark arrives home from work.’ The habitual that the *latter* version (but not the former) embeds is presumably true, since ‘Mark smokes’ is true for any time at which it is true that Mark habitually smokes.

We can leverage intuitions even further here by appealing to highly plausible theses about intrinsic duplicates that Fara appears to accept. Fara is among a small but growing number of philosophers who do not take dispositions to be intrinsic properties; nonetheless, Fara seems to be committed to a weaker intrinsicness thesis: objects which are intrinsic duplicates have the same dispositions (even if some of these are extrinsic dispositions). Fara denies the stronger intrinsicness thesis because of cases of *relational* disposition attributions, e.g. ‘Shoemaker’s key, which has the extrinsic disposition to fit perfectly when inserted into Lock.’³¹ This disposition can be lost without any intrinsic change in the key – specifically, because the disposition may be lost because of a change to *Lock*. Nonetheless, Fara claims, ‘if there were another key, intrinsically just like Shoemaker’s... we would be justified in assuming that it, too, had that extrinsic disposition.’³² Thus I take Fara (at least in his dissertation) to endorse the thesis that intrinsic duplicates share dispositions.

Now suppose Mark has an intrinsic duplicate, Mart. Just like Mark, Mart is disposed to smoke all the time, but unlike Mark, Mart doesn’t have

³¹ Fara 2001, p. 106.

³² Fara 2001, p. 106.

a health-conscious fiancée. So when Mart gets home from work, he smokes, unlike Mark, whose is prevented from smoking after work by Mary. I don't think we can reasonably deny that Mark's intrinsic duplicate Mart is disposed to smoke when he gets home from work; but by the thesis that intrinsic duplicates share dispositions, we can conclude that Mark is also disposed to smoke when he gets home from work.

It seems that we can explain the discrepancy straightforwardly in terms of *masking*: Mark and Mart share the disposition to smoke after work, but Mark has the disposition chronically masked by Mary. Remove the extrinsic mask from the situation, and Mark's disposition will presumably manifest – if Mark comes home one day and Mary is gone, and she doesn't throw out his cigarettes, is there any reason to think that he won't smoke?

Counterexamples like this one can be produced for a wide variety of cases: roughly, to any case where an object disposed to R when T'd doesn't R when T'd because it's disposition is chronically masked. For instance, suppose a vase that is disposed to break when stressed is protected by Johnston's internal packing. Putting the packing in the vase doesn't remove its disposition to break when stressed, but it does falsify the habitual – the vase is still disposed to break when struck, but it no longer breaks when struck.

I have criticized Fara on the basis of his own commitments regarding intrinsic duplicates. While we both agree that 'a is disposed to break when struck' and 'b is disposed to break when struck' will either both be true or both be false for intrinsic duplicates *a* and *b*, I argued that 'a breaks when struck in virtue of an intrinsic property' and 'b breaks when struck in virtue of an intrinsic property' may diverge quite widely in truth conditions because of differences between the two objects' environments. For instance, if *a*'s disposition is chronically masked but *b*'s is not, the *a*-habitual will come out false and the *b*-habitual true though the disposition ascription analysanda will both be true. It follows that Fara's view are internally inconsistent.

Aside from this inconsistency, I think the following can be said about intrinsicness: since (in general, with ‘vulnerable,’ for instance, being a much-discussed exception) intuitions about disposition ascriptions ‘respect’ intrinsicness – i.e. intuitions about the truth value of a disposition ascription to an object x are usually invariant between cases in which x is intrinsically unchanged – the machinery used in analyzing dispositional predicates should also ‘respect’ intrinsicness. It is worth returning for a moment to Lewis conditional analyses to see how the accidental-closeness and strong-centering counterexamples result from a failure of the Lewis conditional to respect intrinsicness. Presumably whether or not object a is fragile does not vary between possibilities in which a is intrinsically the same but extrinsic circumstances change. A flower pot is fragile (or not) regardless of whether it sits on a twentieth-floor windowsill or in a bed of pillows. But whether or not ‘stressed(*flower pot*) $\square \rightarrow$ breaks(*flower pot*)’ holds *does* vary between possibilities in which the flower pot is intrinsically unchanged: the inner sphere of *stressed(flower pot)*-possibilities relative to w_1 in which the flower pot is located on the apartment windowsill may include only *dropped twenty stories(flower pot)*-possibilities and thus *breaks(flower pot)*-possibilities, making the counterfactual true, though the inner-sphere of *stressed(flower pot)*-possibilities relative to w_2 in which the flower pot has Johnston packing inside may all be \neg *breaks(flower pot)*-possibilities, making the counterfactual false. Diagnosis: the disposition analysandum displays invariance of truth value (and presumably truth conditions) across possibilities with no intrinsic change, but both habituais and counterfactuals can vary in truth value across these possibilities. In this respect, habituais seem no better suited to the analysis of disposition ascriptions than subjunctive conditionals.

6. Manley and Wasserman’s proportional account

One big problem with the Lewis conditional analyses arose because a remarkably narrow range of possibilities are relevant to the conditional analysans – the *inner sphere* of test condition-worlds. This narrow focus

resulted in the closeness and centering counterexamples. Another big problem arises from the failure of the Lewis conditional to tolerate exceptions – *all* inner sphere T-worlds must be R-worlds for ‘ $T(x) \square \rightarrow R(x)$ ’ to hold.

David Manley and Ryan Wasserman have proposed an account that addresses these two major problems. First, in evaluating ‘ x is disposed to R when T ’d’, instead of looking solely at the inner sphere of T-worlds, Manley and Wasserman look at *all* nomologically possible events in which x is T’d and in which x is not intrinsically changed. Call these the T-events. Second, they do not require that $R(x)$ in *all* of the T-events (as the Lewis conditional does) but only ‘in some suitable proportion’ of them.³³ Thus, all together: ‘ x is disposed to R when T ’d’ is true iff $R(x)$ in a suitable proportion of T(x)-events.

Manley and Wasserman’s account is not susceptible to the accidental closeness and strong centering counterexamples that plagued the Lewis conditional accounts. In the familiar example from Manley and Wasserman, a certain brick counts as fragile according to the (SCA) because the inner sphere of T-worlds just so happens (by a sort of fluke) to be R-worlds even though most T-worlds outside the inner sphere are not R-worlds. Since Manley and Wasserman’s analysis looks at all of the T-events rather than just the inner sphere, their account presumably gives the correct result (since, assuming the brick is not fragile, $\neg R(\text{brick})$ in the overwhelming majority of T(brick)-events).

The fink/mask/mimic counterexamples also depended upon semantic features of the Lewis conditional. These possibilities are not excluded from consideration in Manley and Wasserman’s analysis, but rather (as in Fara’s account) are tolerated as exceptions. According to Manley and Wasserman, the T-events ‘can [and presumably will] include masking and finkish cases; these need not in any sense be ‘ignored’ or ruled as irrelevant by context.’³⁴ Possibilities in which x ’s disposition to R when T ’d is finkish or masked will drag down the proportion of R-worlds among the T-events, and possibilities in which y mimics a disposition to R

³³ Manley and Wasserman 2008, p. 76.

³⁴ Manley and Wasserman 2007, p. 74.

when T'd will raise the proportion of R-worlds among the T-events. The idea, however, is that while even a single one of these would refute any of the Lewis conditional analyses, those possibilities only comprise a tiny wrinkle of the relevant modal space in Manley and Wasserman's account and thus can be tolerated as exceptions.

A recent paper by Daniel Bonevac, Joshua Dever, and David Sosa (2009) proposes a counterexample intended to exploit the *mimic-inclusion* of Manley and Wasserman's analysis via a mechanism which Manley and Wasserman call 'Achilles' heels.' Take a brick that isn't fragile: Manley and Wasserman remind us that among all of the fragility-tests it could possibly be subjected to, there may very well be some – if you dropped it *just so* – in which it would break. Bonevac, Dever, and Sosa's alleged counterexample to Manley and Wasserman's analysis is a case in which a sturdy brick which has an Achilles' heel (which we'll call test condition *AH*) and is attached to an 'Apollo Machine', which would subject the brick to *AH* were it subjected to *any* fragility-test, causing it to break. As a result, Bonevac, Dever, and Sosa claim, in any possible situation in which the brick is fragility-tested, it gives the fragility-response, since the Apollo Machine causes it to be subjected to *AH* (in which it breaks). Thus, since *all* of the T-events are R-events, and 100% is surely a 'suitable proportion' in any context, the brick is fragile.

I do not think the counterexample is decisive against Manley and Wasserman's account. They would seem to have the following line of response available (to which I am sympathetic): though the brick breaks in all of the cases in which it is tested while attached to the Apollo Machine, there are also many cases – in fact, the overwhelming majority of T-events – in which *the brick isn't attached to the Apollo Machine* (or anything like an Apollo Machine), and thus doesn't break when stressed. Bonevac, Dever, and Sosa are correct that all of the T-events in which the brick is attached to the Apollo Machine are R-events, but surely the overwhelming majority of T-events are cases in which the brick isn't attached to the machine. The Apollo Machine is essentially a fancy mimic; the brick

mimics fragility just as it would if a more banal mimic like a stress-activated mine were attached.

However, I do agree with Bonevac, Dever, and Sosa that Manley and Wasserman's fink-, mask-, and mimic-inclusion can be exploited to generate counterexamples to the proportional analysis. Finks, masks, and mimics are presumably tolerated because they aren't supposed to make much of a dent in the relevant proportions of T- and R-events. But I don't think that this is case. Let me illustrate.

Suppose we wonder whether two vases in a (nomologically possible) world w are fragile. In world w , there is actually quite a lot of sophisticated fink and masking technology. People frequently use moleculo-finks and moleculo-masks to prevent fragile things from breaking when stressed. If a certain object is plugged into a moleculo-fink it will rigidify its internal microstructure when stressed, rendering it (temporarily) not at all fragile. If a certain object is plugged into a moleculo-mask, the mask will produce counterbalancing shockwaves whenever the object is stressed that prevent the object from breaking. But here's the rub – being able to plug into a moleculo-fink or -mask requires having a special internal circuitry and a USB port. But our two vases differ in the following respect: one has the requisite internal circuitry and USB port for the moleculo-fink and -mask, but the other, an antique, does not. Now suppose our two vases break in exactly the same proportion of non-fink/masking cases. I think as a result we would be strongly inclined to that they are equally fragile. However, when the finkish and masking cases are included, the proportions get thrown off – the vase with the special circuitry will break in significantly smaller proportion of T-events, and thus according to Manley & Wasserman's account is significantly less fragile; but as they have been described in the example, the two vases, if my intuition is right, are equally fragile.

Or consider the following case, which I first offered in §3 in support of the exclusion thesis: *DP* and *IP* are poisons which are exactly the same in almost every way (same lethal dose, same lethal modes of administration, similar modes of action) and both of which take a long

time to cause death. However, *IP* has a highly visible side effect –it turns victims’ skin purple – which *DP* does not. There are antidotes widely available for both *IP* and *DP*. Now, since *IP* takes a long time to cause death and has the visible side effects, presumably the T(*IP*)-events will include quite a lot of masking cases in which $\neg R(IP)$. However, since *DP* doesn’t have these visible side effects, people will take antidotes which mask its poisonousness in far fewer cases. Thus, the two will diverge greatly in the proportions relevant to Manley and Wasserman’s account. If Manley and Wasserman’s analysis is correct, then *p* is far more poisonous than *q*, but intuitively, the two seem to be equally poisonous. How could the side effect alone result in a dramatic divergence in poisonousness?

There are both positive and negative lessons to be learned from Manley and Wasserman’s account. The positive lesson is that direct quantification over T- and R-events is a big step in the right direction from the Lewis conditional accounts, for two reasons: first, appealing to a broader rather than narrower portion of modal space prevents the closeness counterexamples that plagued the Lewis conditional accounts, and second, requiring that an object give the response in a ‘suitable proportion’ of relevant T-events rather than in *all* inner sphere T-worlds allows the account to tolerate some exceptional cases in which an object is disposed and tested but does not give the response. By broadening the range of relevant T-events beyond the inner sphere to all nomologically possible, no-intrinsic change T(*x*)-possibilities, Manley and Wasserman have brought the set of *analysans-relevant* events much closer to the set of intuitively *analysandum-relevant* worlds than the Lewis conditional accounts. Their account is not threatened by the possibility of counterexamples owing to fluky facts about an object’s behavior in a small region of modal space.

On the other hand, the failure of Manley and Wasserman’s account even in light of this giant leap forward suggests the non-negotiability of the exclusion of fink and masking cases from the T-events, and mimics from the R-events. Two objects appear to be similarly disposed iff they display the same or similar patterns of the relevant response behavior in the broad

modal space of D-tests *excluding finks, masks and mimics*; but this may be the case even if the two objects differ widely in susceptibility to finks, masks, and mimics. Susceptibility to finks, masks, and mimics – e.g. what separates the wired vase from the non-wired vase or *Indiscreet Poison* from *Discreet Poison*– appears to be largely irrelevant to intuitions about whether or not an object is disposed, and to what degree it is disposed.

Thus we have an agenda for revising Manley and Wasserman’s account: supplement it with some mechanism which excludes fink and masking cases from the T-events and excludes mimicking cases from the R-events. However, since the mechanisms for exclusion proposed by conditional analysts – *precisification* and *qualification* – appear not to be up to the task, there is no obvious way to repair Manley and Wasserman’s analysis by regimenting it with test- and response- specifications which exclude finks, masks, and mimics. In Part III, I will propose a mechanism for fink-, mask-, and mimic-exclusion – *resemblance to paradigmatic test- and response conditions* – but first, in Part II, I will consider recent arguments in the dispositions literature concerning the context-sensitivity of dispositional predicates.

PART II: CONTEXT-SENSITIVITY ARGUMENTS

7. The context-sensitivity of dispositional predicates

A number of philosophers have advanced the thesis that dispositional predicates (as well as explicit disposition attributions) are context-sensitive. Recently, several authors have used the context-sensitivity of dispositional predicates and explicit disposition attributions to argue for their own analyses by suggesting that their analysis provides an adequate treatment of this context-sensitivity. Michael Fara (2001, 2005) was (to my knowledge) the first to use this sort of context-sensitivity argument for his analysis of dispositions:

Disposition ascriptions, like the habitual sentences they embed, are context-dependent to a certain extent... Once this context-dependence of disposition ascriptions is recognized, we see further reason to prefer the Habitual Account to others, since the context-dependence of disposition ascriptions seems to run in exact parallel to that of corresponding habitual sentences.³⁵

Manley and Wasserman further the dialectic in their (2008), claiming that ‘no one has suggested a plausible mechanism for the context dependence of dispositional predicates,’³⁶ and arguing that their own proportional account does provide such a mechanism. Manley and Wasserman compare the context-sensitivity of dispositional predicates to that of gradable adjectives and argue that this context-dependence can be treated semantically by the context-dependent expression ‘suitable proportion’ in their analyses. Bonevac, Dever, and Sosa suggest in a recent paper that the context-sensitivity of dispositional predicates ‘outruns that of the comparative’ and that Manley and Wasserman’s account thus fails to give a complete mechanism for the context-sensitivity of dispositional predicates.³⁷

³⁵ Fara 2001, p.45. For similar remarks in his (2005), see p. 76.

³⁶ Manley and Wasserman 2008, p. 72.

³⁷ Bonevac, Dever, and Sosa 2009, footnote 11.

8. Context-shifting arguments

These authors argue for the context sensitivity of dispositional predicates and explicit disposition attributions using context-shifting arguments. They describe multiple scenarios in which the same disposition sentence is tokened, but in which the disposition term's contribution to the intuitive truth conditions of what is said seems to differ.

Here's an example. Sasha and Malia are in pottery class and have each sculpted pots, which we'll call *S* and *M*. Sasha picks up *M* and Malia says 'Be careful! That is very fragile.' Let's call this utterance u_1 . Later that class, the sisters' teacher, Mr. Potter, goes around examining the students' work one last time before firing the pots in the kiln. He picks up *S* and furrows his brow: 'Sasha, this pot is too fragile to fire as it is – see this weak spot where the clay is thin? It would almost certainly shatter in the kiln.' Malia asks – 'Mr. Potter, is my pot fragile as well?' Mr. Potter examines *M*. 'No, Malia. This pot isn't fragile at all. It will be fine in the kiln.' Let's call this utterance u_2 . Intuitively, Malia and Mr. Potter could both have spoken the literal truth by uttering u_1 and u_2 , respectively; however, one does so by saying that *M* is fragile, while the other one says that *M* isn't fragile.

To get at the content expressed by 'fragile' in these contexts, we can paraphrase the two utterances without using the dispositional predicate: Malia seems to have said (roughly) that the pot is disposed break if it were dropped, or grasped too tightly, or something like that, and Mr. Potter seems to have said (roughly) that the pot is not disposed to shatter or deform in response to being fired in the kiln. In the context of u_1 'fragile' appears to express the property of being disposed to break if subjected to moderate physical stress, and in the context of u_2 'fragile' seems to express the property of being disposed to break in response to heat stress. All of this provides *prima facie* evidence for the claim that 'fragile' is context-sensitive: that is, the content semantically expressed by the term may vary

between tokenings of that term in different utterances. We will observe that other dispositional predicates show similar context-dependence: ‘brave,’ for instance, can intuitively express different properties in contexts in which the following are salient (respectively): a dangerous military mission, fighting fires, corporate whistleblowing, facing a painful medical treatment, eating fugu, and so on.

I say *prima facie* evidence because context-shifting arguments are quite controversial, and have recently received especially trenchant criticism from Herman Cappelen and Ernie Lepore.³⁸ One thrust of their criticisms is that intuitions about *what is said* (which may vary with the context of utterance) are not reliable guides to the *semantic content expressed* by a sentence, which is invariant between contexts. Other philosophers have suggested that context-shifting arguments are insufficient to establish context-sensitivity.³⁹ Frequently, critics of so-called ‘radical contextualism’ – i.e. the sort of contextualism we get if we take content semantically expressed by utterances to track the intuitions invoked in context-shifting arguments – argue that it is antithetical to the project of producing systematic, compositional semantic theories for natural language.⁴⁰ It is not my purpose in this paper to defend the methodology of context-shifting arguments or to make general claims about the pervasiveness or proper analysis of context-sensitivity in natural languages. As I have mentioned, a number of philosophers⁴¹ are (tacitly) committed to this ‘radical’ contextualism in taking context-shifting arguments to provide sufficient evidence for context-sensitivity. My aim is just to carry these arguments to their dialectical end: assuming that the content expressed by dispositional predicates does vary contextually in the way that context-shifting arguments used by Fara, Manley and

³⁸ Cappelen and Lepore 2005.

³⁹ E.g. see Zoltan Szabó (2006).

⁴⁰ For instance: ‘Not only is [the postulation of rampant context-sensitivity] ‘the lazy man’s approach to philosophy,’ it undermines systematic theorizing about language. The more we believe context can influence semantic content, the more we will find ourselves at a loss when it comes to explaining ordinary communication.’ (Szabó 2006, p. 32)

⁴¹ Fara (2001), Manley and Wasserman (2007) and (2008), Bonevac, Dever and Sosa (2006) and (2009).

Wasserman, and Bonevac, Dever, and Sosa suggest, are any of their proposed mechanisms for this context-sensitivity adequate? I will argue that they are not. Those skeptical of ‘radical’ contextualism and context-shifting arguments may view these remarks as an in-house debate. Additionally, by probing intuitions in this area, I hope to accumulate a body of data about the sort of context-sensitivity dispositional predicates exhibit if context-shifting arguments bear the dialectical weight we put on them; this body of data may prove useful as a case study for more foundational disputes.

How prevalent is the context-sensitivity of dispositional predicates (if we take context-shifting arguments at face value)? Quite ubiquitous, it seems to me. First, I’ll claim that a few token dispositional predicates show potential for great contextual variation (rather than, say, a small range of different contents that can be expressed which we might chalk up to mere lexical ambiguity or simple polysemy.) Second, I’ll gesture at context-shifting arguments that might be fleshed out in greater detail for a wide variety of dispositional predicates, suggesting that the phenomenon of context-sensitivity isn’t confined to a few aberrant dispositional predicates. My interest here is to investigate some proposed mechanisms for the context-sensitivity of dispositional predicates. *The methodological soundness of using context-shifting arguments to establish semantic context-dependence is assumed.* The general methodology I take – not only in evaluating the various analyses of context-sensitivity, but also in discussion of their extensional adequacy in Part I, and in further remarks in Part III about lexical competence, vagueness, and conceptual unity – is well summarized by Cappelen and Lepore:

*A theory of semantic content is adequate just in case it accounts for all or most of the intuitions speakers have about speech act content, i.e. intuitions about what speakers say, assert, claim, and state by uttering sentences.*⁴²

⁴² Cappelen and Lepore 2005, p. 53.

I hold fast to this thesis as a sort of methodological bedrock in this dialectical context, though I am more than willing to entertain the possibility – in a different dialectical context – that Cappelen and Lepore are apt to title it ‘the Mistaken Assumption.’

So much for methodological preliminaries.⁴³ Here’s a story intended to show that ‘brave’ has expresses quite different properties in different contexts. Let’s follow the brave (and sometimes not so brave) exploits of best friends and lifelong companions Johnny and Max. In their teenage years, Johnny is a wild one, spending his days indulging in the thrills of extreme sports and his nights eating fugu, while Max abstains. ‘I’m just not brave, Johnny,’ he’d say, ‘but you sure are’ (u_1). But when the two both join the military, things are a little different. Johnny ends up deserting his unit the night before combat while Max valiantly leads the attack. Their unit commander says ‘Max sure is brave, but that Johnny, the deserter, he doesn’t have a lick of brave in him’ (u_2). The ascription is again inverted when Max and Johnny both end up side-by-side in the field hospital being treated for malaria (Johnny has been found since deserting). The war has gotten so bad that supplies are short and quinine needs to be rationed, and a field doctor says ‘I’m going to give the quinine to Johnny, who has been brave in facing his treatment, unlike Max’ (u_3). After their military careers, Johnny and Max enter the high-stakes world of junk bond sales, where their Johnny is far more willing to meet with big, well-informed clients and maintain risky accounts. This leads their boss to say ‘Johnny is the bravest junk bond salesman I’ve ever known, but Max doesn’t seem to be very brave at all’ (u_4). But again the ascription is inverted, when Max becomes a whistleblower, publishing a signed account of the bond agency’s fraudulent practices in the newspaper while Johnny is

⁴³ Well, here’s one last one: I appeal to *the property expressed* by dispositional predicates uttered in context in what follows because it is easy to make sense of contextual variation by rephrasing the property expressed as an explicit disposition ascription. That is, it’s much easier to make clear how what ‘fragile’ expresses in context c_1 differs from c_2 by pointing out that one token expresses the disposition to break in response to moderate physical stress and the other expresses the disposition to break in response to heat stress. This is merely a convenience; what follows does not depend seriously upon adopting this way of speaking over other idioms for talking about semantic content (extensions, intensions, Fregean senses, possible worlds, truth conditions, *et cetera*).

complicit and works to cover up the scandal so he doesn't lose his job. A story on the scandal describes Max as 'a brave whistleblower who has never cowered in the face of corporate greed' and Johnny as 'the very opposite of brave, a bureaucrat who would sacrifice the integrity of the entire financial system in order to keep his job' (u_5).

Here we have five different utterances, u_1 - u_5 , each of which, as the scenario is described, seems to be literally true, each ascribing a property using the term 'brave', three of which ascribe bravery to Johnny and not to Max, two of which ascribe bravery to Max and not to Johnny. Assume further that Max and Johnny do not particularly change in their 'brave'-relevant characteristics throughout their lives: throughout his life, Johnny is always disposed to eat the fugu and brave the extreme sports, always disposed to wimp out in the military, to respond diligently to the doctor's orders, to take the riskiest junk bond accounts, and to try to save his own butt in the corporate scandal; Max, throughout his life, is disposed to do just the opposite. Who is brave and who isn't? Intuitively, in u_1 'brave' expresses the disposition to fearlessly engage in reckless behavior; in u_2 , the disposition to fight valiantly in combat; in u_3 , the disposition to courageously endure difficult medical treatment; in u_4 , the disposition to engage in high-risk business dealings; in u_5 , the disposition to face up to corporate abuse of power. The very same word, 'brave,' seems to express five different properties in the five different contexts.

Here's a similar (but shorter and sketchier) context-shifting argument for 'fragile.' Suppose there is a widget-factory that has a number of parallel assembly lines **A**, **B**, **C**... **Z** each of which produces a different widget part. Each assembly line begins with the same input object, a ball of widget-gunk, but each subjects the ball of widget-gunk to different operations: **A** involves heat pressure, **B** involves cutting and stretching, **C** involves pressing into a dense glob, and so on. Suppose the response of a ball widget-gunk to each assembly line process varies with properties of the particular ball of widget-gunk: some balls of widget-gunk survive the heat pressure and become **A**-parts, but others deform and crack; some balls of widget-gunk survive the cutting and stretching and become **B**-

parts, but others disintegrate; some balls of widget-gunk survive being pressed into a glob and become **C**-parts, but others become misshapen or crumble; and so on.

Different properties are desirable for different assembly lines: for instance, a ball of widget-gunk with lots of air bubbles in it might be well suited to successfully survive assembly line **B**'s cutting and stretching, but to deform in response to line **A**'s heat pressure. Imagine a conversation the individual assembly line managers might have around the pile of widget-gunk balls: the **A**-manager says of an air-bubbly ball 'This ball of widget-gunk looks pretty fragile,' while the **B**-manager dissents, saying 'That ball isn't the least bit fragile!' The **C**-manager says of a certain ball susceptible to crumble under pressure 'It's a fragile one, no doubt about that,' but the **A**- and **B**-managers have their doubts. We can imagine each assembly line manager speaks the literal truth using the term 'fragile' to pick out the balls of widget-gunk which have properties conducive to a bad result on her own assembly line. Thus we get as many properties expressed by 'fragile' as there are assembly lines. It appears that the property expressed by 'fragile' can differ on different tokenings in a way that depends on contextual differences related to the interests of each particular manager.

I've now sketched a few context-shifting arguments in depth and concluded that 'brave' and 'fragile' can be used in different contexts to express different properties. At this point, I'll cease drawing the elaborate examples, and merely gesture at context-shifting arguments for other dispositional predicates. The question now is whether or not the context-sensitivity that we saw exhibited by 'fragile' and 'brave' generalizes to other dispositional predicates. For though the property expressed by 'fragile' and 'brave' would seem to differ quite significantly on different occasions of tokening, perhaps they are exceptional cases.

It does not appear, however, that they are exceptional cases. Think about 'irascible,' and 'loquacious.' In one context, it seems like 'irascible' might express the disposition to blow up in response to minor annoyances, but in another context express the disposition to blow up in response to

major catastrophes. We can also imagine that in some cases, only those who respond to *certain* salient sorts of stresses in the right way count as irascible, while in another context, only those who respond to *different* sorts of stresses count. The property expressed by ‘loquacious’ might vary depending on whether classroom discussion-hogging or cocktail hour conversation-starting is salient.

Manley and Wasserman observe that dispositional predicates generally display *gradability*: dispositional predicates order objects as *more D* or *less D* than others, but vary between contexts in *how high along the ordering of D-ness* the threshold of ‘D’ falls. In a *cooking mushrooms*-context, the threshold for ‘poisonous’ will likely be much lower than in an *assassination attempt planning*-context; e.g. in the former the ‘shaggy parasol’ mushroom, which causes mild upset stomach, may count as poisonous, though in the latter it likely will not. In the examples, ‘fragile’ and ‘brave’ seemed to far outrun this sort of context-sensitivity: confusion over Johnny and Max’s bravery can’t be resolved by appealing to different thresholds in different contexts (since sometimes Max comes out above the threshold and Johnny below, but in other cases just the opposite). I think that the same is true for even for more ‘natural’ dispositional predicates: they show context-sensitivity that far outruns gradability. Consider ‘poisonous’: one parameter of contextual variation has to do with mode of causing death: in a *worried parent*-context, ‘poisonous’ may include drain cleaner (a corrosive) in its extension, although in a *toxicology lab results*-context, it would not. What variety and degree of non-lethal harm counts as a poisonous response would seem to vary as well. Prima facie, ‘solubility’ displays context-sensitivity that plausibly tracks what counts in context as *dissolving*.

In general, the context-dependence of dispositional predicates corresponds to a difference not only in the property expressed by the dispositional predicate, but also in what counts as a D-test or D-response. In the Johnny and Max story, each tokening not only corresponds to a different property, but also to different ‘bravery’ test- and response conditions: in the context of u_i , bravery-tests are events which demand a

certain reckless fearlessness or bravado; in the context of u_2 , bravery-tests include military battles and events which resemble these in certain relevant ways; in the context of u_3 , bravery-tests include events of facing illness; and so on.

In exploring the context-sensitivity of various dispositional predicates, we can ask ourselves how the conditions which count as D-tests and D-responses depend upon features of context. In Part I, it was implicitly assumed that there was a canonical, context insensitive set of D-test and D-response events for each dispositional predicate, but context-shifting arguments suggest that the possibilities which count as D-tests or -responses in one context of utterance may not count as D-tests or -responses in another.

I want to make one final remark before getting on with the analyses and their potential ‘mechanisms’ for context-sensitivity. Fara, Manley and Wasserman, and Bonevac, Dever and Sosa seem to agree in their discussions of the context-sensitivity of dispositional predicates that the *context sensitivity mechanism*-desideratum for an analysis of dispositional predicates is *different* from the *extensional adequacy*-desideratum. Fara’s context-sensitivity claim appears to introduce a *novel* criterion by which to evaluate analyses of dispositional predicates. Bonevac, Dever and Sosa make the somewhat ambivalent claim that ‘*perhaps* [it] is to be lamented’ (my italics) that ‘no one has suggested a plausible mechanism for the context dependence of dispositional predicates’; presumably they think it is *certainly* to be lamented that no one has provided an extensionally adequate analysis of dispositional predicates.⁴⁴

However, once we go in for the radical context-sensitivity of dispositional predicates, accounting for context-dependence *just is* a matter of providing an extensionally adequate account of the truth conditions of sentences involving these predicates in context. If we take dispositional predicates to be context-sensitive, it is not the case that we have our old agendum – extensional adequacy – as well as a new one – ‘providing a mechanism’ for context-sensitivity. There is no longer any

⁴⁴ Bonevac, Dever, and Sosa 2009, p. 7.

need to provide an extensionally adequate analysis of the *context-invariant content* expressed by dispositional predicates, since dispositional predicates *aren't* context-insensitive. Rather, for an analysis to be extensionally adequate, it must account for the truth-conditions of dispositional predicate ascriptions in context. The context-sensitivity argument – ‘we should prefer *this* account to *that* one because this one ‘provides a mechanism’ for context-sensitivity while that one doesn’t’ – simply *is* the argument that, given the context-dependence of dispositional predicates, *this* analysis is extensionally adequate (in correctly predicting the truth values of dispositional predicate ascriptions *in context*) while *that* analysis is *not* extensionally adequate (since it fails to predict the truth values of dispositional predicate ascriptions in context).

As we’ll see when we turn to the specific accounts, we can easily modify superficially context-insensitive analyses with supplemental ‘mechanisms’ for context-sensitivity into analyses which (aim to) give an extensionally adequate account of the truth-conditions of dispositional predicate ascriptions in context.

I now turn to the analyses of dispositional predicates examined in Part I. I’ll argue that none of those analyses provide a plausible account of the context-sensitivity of dispositional predicates.

9. The conditional analyses

- a. The simple conditional analysis and mechanisms for context-sensitivity

None of the conditional analyses – (SCA), (PCA), (QCA), or (DCA) – were explicitly formulated with the context-sensitivity of dispositional predicates in mind. No adherents of any of these accounts, as far as I am aware, have made the claims that Fara and Manley and Wasserman have made in connection with their accounts, viz. that the analysis provides a ‘mechanism’ for the context-sensitivity of dispositional predicates. We should probably expect that the prospects for happening upon such a mechanism buried in these analyses are not very good. It should come as

no surprise, then, that I'll argue that there is no such mechanism in any of the conditional analyses. However, since there are lessons about the context-dependence of dispositional predicates to be learned by taking a close look at these analyses – especially the (SCA) – it should be a worthwhile effort.

The (SCA), which analyzes dispositional predicates in terms of vague, rough-and-ready conditionals, has the following going for it: in some cases, the context-sensitivity of the dispositional predicate appears to track the context-sensitivity of the associated conditions. For instance, intuitively 'fragile' can be used in some contexts to ascribe the property of being disposed to *physically disintegrate* in response to stress, but in others to ascribe the disposition to *cease functioning* in response to stress. Some recalcitrant teenagers going down to the junkyard with baseball bats to blow off steam by smashing old cars to bits will use the term 'fragile' to ascribe the *physical disintegration* property; on the other hand, if my five year old nephew is running wildly around the house with my laptop in his arms, I might try to avoid disaster by using 'fragile' to ascribe the *ceasing to function* property. Luckily enough, this contextual difference corresponds nicely to a difference in the content expressed by our rough-and-ready associated response condition 'breaks' in context: 'x breaks' can be used to express something like 'x physically disintegrates' in some contexts, and something like 'x ceases to function' in others.

It looks like a plausible move to appeal to the context-dependence of the associated rough-and-ready condition 'breaks' to explain the context-dependence of 'fragile': what 'x is fragile' expresses in a context is what the translation sentence 'were x stressed, x would break' would express were it uttered in that context. In the *junkyard*-context, the teenagers ascribe the *physical disintegration*-property because if they were to say 'That car over there looks like one which would break were it stressed!', 'break' would express something like the property of *physically disintegrating*; in the *wild nephew*-context, I ascribe the *ceasing to function*-property because were I to say 'Be careful! The laptop would break were it stressed!', 'break' would express something like the property of *ceasing to function*.

Supposing this generalizes, the (SCA)-adherent has a plausible mechanism for the context-sensitivity of dispositional predicates: the context-sensitivity of dispositional predicates tracks and is explained by the context-sensitivity of the rough-and-ready associated test and response conditions.

We're now in a position to say why an analysis with a 'mechanism' for the context-sensitivity of dispositional predicates is really different than an extensionally adequate analysis of dispositional predicates in context. What does it mean to say the following?:

[SCA + CS] 'x is fragile' is true iff were x stressed, x would break, and the context-sensitivity of 'stressed' and 'break' provides a mechanism for the context-sensitivity of 'x is fragile.'

It seems that what [SCA + CS] gestures at is something like this:

[SCA2] 'x is fragile' is true in context c iff 'were x stressed, x would break' is true in context c (and 'stressed' and 'breaks' are context-sensitive.)

The first bit of [SCA + CS] (up to 'and') seems to give an *extensionally inadequate* analysis of the target sentence 'x is fragile', since 'stressed' and 'break' are tokened in context (the context of the analysis) and in that context, semantically express some specific properties. But the appeal to 'mechanism' for context sensitivity depends on the *different* properties expressed by 'stressed' and 'break' in *different* contexts. Given that the terms can't express all of these different properties on a single tokening in context, the analysis ("x is fragile' is true iff were x stressed x would break') fails extensionally for contexts in which 'stressed' and 'break' express *different* properties, despite the mechanism for context-sensitivity in the second part of the analysis.

This illustrates what I had in mind at the end of §2 in saying that the *extensional adequacy*-agendum and *mechanism for context-dependence*-agendum are one and the same: namely, providing an analysis that correctly predicts the truth conditions of the target sentence as it occurs in context.

All of the potential ‘mechanisms’ for context-sensitivity we’ll encounter work in this same way: they identify some context-sensitive expression in the *analysans* and peg the context-sensitivity of the *analysandum* to the context-sensitivity of those expressions. But instead of *using* these context-sensitive expressions in the analysis – which will result in their context-sensitivity ‘going to waste’, so to speak, since they semantically express some single property in the context of the analysis – we should tie what is expressed by the *analysandum* expression in context to what would be expressed by the *analysans* expression in context. The easiest way of doing this involves the move to [SCA2]: mentioning the *analysandum*-sentence and the *analysans*-sentence, and analyzing what is expressed by the *analysandum*-sentence in some context *c* in terms of what is expressed by the *analysans*-sentence in *c*.

This same move can be made for the subsequent proposed mechanisms for context sensitivity: e.g. when Fara says that sentence ‘*x* is disposed to *R* when *T*’d’ is true iff *x* *R*’s when *T*’d in virtue of an intrinsic property, and this habitual (‘*x* *R*’s when *T*’d’) provides a mechanism for the context-sensitivity of the *analysandum*, we might reconstruct his analysis as something like ‘sentence ‘*x* is disposed to *R* when *T*’d’ is true in context *c* iff ‘*x* *R*’s when *T*’d in virtue of an intrinsic property’ is true in *c*.^{45,46}

We have a candidate mechanism for context-sensitivity in the (SCA)—the context-sensitivity of dispositional predicates tracks the

⁴⁵ Sanford Shieh has suggested to me that this isn’t the *only* way to recast mechanisms for context sensitivity as extensionally adequate analyses of dispositional predicates in context. For instance, we might say the following: ‘*x* is fragile’ is true in *c* iff, in *c*, if *x* were stressed, *x* would break. He pointed out that it doesn’t follow from the fact that the relevant terms are used in the context of analysis that they don’t express the properties that would be expressed in the context via a comparison to cross-world reference: we say that ‘Aristotle is fond of dogs’ is true with respect to a counterfactual state of the world *w* just in case, in *w*, Aristotle is fond of dogs (*using* rather than *mentioning* ‘Aristotle’ to denote the person designated by ‘Aristotle’ in *w*). Thus, the ‘metalinguistic’ reconstruction is not obligatory.

⁴⁶ Taking contexts to be *concrete* circumstances of utterance, ‘imagining’ utterances into contexts in which they do not occur is tricky business: what does it mean to say ‘ $\lceil \varphi \rceil$ is true in *c*’ when $\lceil \varphi \rceil$ isn’t actually tokened in *c*? This is a serious question in need of a non-trivial answer, but addressing the issue would take us quite far afield of our topic here. Since the problem does not seem particularly pressing for our purposes – I think we are able to make good sense of the analyses without solving this problem – I leave it aside here.

context-sensitivity of their associated conditions. We can *predict* the context-dependence of the analysandum by looking to the context-sensitivity of the analysans, and we can *explain why* dispositional predicates are context-sensitive to begin with – because of the context-sensitivity of their associated test- and response conditions. Let's put the analysis to use in our pottery classroom case: u_1 expresses something like 'M is disposed to break when subjected to moderate physical pressure' because, given what is salient in the context of u_1 , 'is stressed' would express (roughly) the property of being subjected to moderate physical pressure; u_2 expresses something like 'M is disposed to break when subjected to heat stress' because, given what is salient in the context of u_2 , 'is stressed' would express (roughly) the property of being subjected to heat stress characteristic of the sort undergone when fired in a kiln. This seems to be a satisfactory and elegant explanation of the context-sensitivity in this case.

So far so (surprisingly!) good. Unfortunately, while the proposed mechanism looks good in these few cases, it is inadequate to account for *all* (or even *most*) of the context-sensitivity of dispositional predicates. In most cases, there simply aren't any test- or response-characterizing expressions in the analysans which display the sort of context-sensitivity to be accounted for in the analysandum.

Take 'poisonous.' I suggested above that in some contexts (e.g. *toxicology results*) 'poisonous' applies only things which kill via a certain chemical mode of action at the cellular level (e.g. excluding corrosives), but in other contexts (e.g. *worried parents*), also applies to things which kill via a broader range of modes of action (e.g. including corrosives). If the proposed mechanism is adequate to account for all of the context-sensitivity of dispositional predicates, then there will be some expression in the response-characterization of our analysis of 'x is poisonous' which displays this sort of context-sensitivity: it should express the response condition *causes death via a chemical mode of action at the cellular level* in some contexts, and the different response condition *causes death via a certain (more inclusive) mode of action* in other contexts. The problem is

that there just *isn't* any term in English that shows anything like this sort of context-sensitivity; there is no word that denotes the property of causing death via a chemical mode of action at the cellular level in all and only the contexts in which such an event would count as a 'poisonous'-response, but expresses something different in all and only the contexts in which this property is the test condition for 'poisonous'.

In a few cases, like 'fragile' and 'is stressed' and 'breaks', and 'soluble' and 'dissolves', there seems a strong connection between the context-sensitivity of the dispositional predicate and the rough-and-ready associated condition expressions. But in most cases, there just isn't any such connection: there aren't terms in the analysans which are context-sensitive in the right sort of way. For many dispositional predicates (take the 'poison' test condition, or almost any dispositional predicate denoting a human behavioral disposition, like 'brave,' 'irascible,' 'loquacious,' and 'stubborn') we have enough trouble providing a simple, extensionally approximate rough-and-ready characterization of the test and response conditions – are we really to suppose that there exist some characterizations not only up to that task, but which also precisely parallel and ground the context-sensitivity of these dispositional predicates? It seems quite implausible.

The associated conditions mechanism is a start in explaining the context-sensitivity of dispositional predicates, but is inadequate to explain *all* of the context-dependence that dispositional predicates and disposition ascriptions show. At the very least, we will need a further mechanism to explain what's going on in the cases that the associated conditions mechanism can't account for.

b. The complex conditional analyses

The (QCA) seems to offer little potential for improvement upon the (SCA): 'normal conditions' and 'ideal conditions' are certainly context-sensitive expressions, but how can they explain the context-sensitivity that the (SCA) could not? It seems pretty clear that the qualifier isn't going to

pick up all the slack where the associated conditions could not account for context-dependence; for this reason, I'm going to assume that the (QCA) fares much the same as the (SCA) in respect of providing a mechanism for context-sensitivity.

The (PCA) and (DCA) are actually *less* equipped to provide a mechanism for the context-sensitivity of dispositional predicates than the (SCA) and (QCA). Providing a mechanism for the context-sensitivity of dispositional predicates involves picking some context-sensitive expressions in the proposed analysans (e.g. the rough-and-ready associated conditions) and claiming that the context-sensitivity of dispositional predicates is determined by and 'tracks' the context-sensitivity of these expressions. The (SCA)-adherent can point to the context-sensitivity of associated conditions, the (QCA)-adherent to the context-sensitivity of expressions of associated conditions and the context-sensitivity of 'normal,' Fara to the context-sensitivity of habitual sentences, and Manley and Wasserman to the context-sensitivity of 'suitable proportion.' These mechanisms can be transformed into analyses (for example, metalinguistic analyses):

- (SCA) 'x is fragile' is true in context c_i iff 'T(x) $\square \rightarrow$ R(x)' is true in c_i
- (QCA) 'x is fragile' is true in context c_i iff 'In normal conditions, T(x) $\square \rightarrow$ R(x)' is true in c_i
- (FARA) 'x is disposed to break when stressed' is true in context c_i iff 'x breaks when stressed in virtue of an intrinsic property' is true in c_i
- (M&W) 'x is disposed to break when stressed' is true in context c_i iff 'x breaks in a suitable proportion of nomologically possible worlds in which x is intrinsically unchanged' is true in c_i .

I argued that (SCA) does quite well for instance in predicting and explaining the 'stress'-tracking and 'breaks'-tracking context-sensitivity of 'fragile,' and we will see that (M&W) does a pretty good job of accounting for the gradability of dispositional predicates.

Similar moves from (PCA)- or (DCA)-adherents would result in something like these unlovely constructions:

- (PCA) For precise test conditions $T_1(x), T_2(x) \dots T_n(x)$ and response conditions $R_1, R_2, \dots R_m$ associated with dispositional predicate 'D', 'x is D' is true in context c_i iff
 $(T_1(x) \wedge T_2(x) \wedge \dots \wedge T_n(x)) \square \rightarrow (R_1(x) \wedge R_2(x) \wedge \dots \wedge R_m(x))$ is true in c_i .
- (DCA) For precise test conditions $(T_{1,1}(x), T_{1,2}(x), \dots T_{1,p}(x)), (T_{2,1}(x), T_{2,2}(x), \dots T_{2,q}(x)), \dots (T_{r,1}(x), T_{r,2}(x) \dots T_{r,s}(x))$ and response condition $R(x)$ associated with dispositional predicate 'D', 'x is D' is true in context c_i iff
 $(T_{1,1}(x) \wedge \dots \wedge T_{1,p}(x)) \square \rightarrow (R(x) \vee \dots \vee (T_{r,1}(x) \wedge \dots \wedge T_{r,s}(x)) \square \rightarrow (R(x))$ is true in c_i .

The (SCA)-, (QCA)-, (FARA)-, and (M&W)-adherents could propose a mechanism for context sensitivity by picking some expressions in the analysans which the context-sensitivity of dispositional predicates allegedly tracks. But in the case of the (PCA) and (DCA), no expressions in the analysans appear to be very useful for this purpose. Does the context-sensitivity of the myriad precise conditions provide a mechanism for the context-sensitivity of dispositional predicates? That seems incredibly implausible. How could a conjunction of precise 'fragile'-associated conditions display the same sort of context-sensitivity that is reasonably modeled by 'is stressed'? What about a massive disjunction of conditionals? Disjunction of conjunctions of conditionals? I can't imagine that anyone would defend the (PCA) or (DCA) on this context-sensitivity point, since there just doesn't appear to be anything in their big conjunctions and disjunctions of precise conditions (which haven't even been specified) which even remotely tracks the context-sensitivity of dispositional predicates. If dispositional predicates are context-sensitive in the ways we've supposed, then it seems hard to see how the (PCA) or (DCA) could be extensionally adequate (in the sense of providing the correct truth values for disposition sentences in context.)

10. Fara's habitual analysis

a. The vandal and the specialists

Recall Fara's habitual analysis: 'x is disposed to R when T'd iff x has an intrinsic property in virtue of which it R's when T'd.' Fara claims that the context-sensitivity of disposition attributions 'run[s] in exact parallel to that of corresponding habitual sentence[s] [they embed]' and that this is a good reason to suppose that some habitual account is correct.⁴⁷ If this context-sensitivity claim is right, then that would provide evidence for the claim that *some* habitual analysis is feasible for dispositional predicates (even if, as I've argued, Fara's particular (2001)/(2005) analysis is not). In this section, I'll argue that the context-sensitivity of habituals is actually quite different from the context-sensitivity of dispositional predicates and that this constitutes evidence against any account (not just the Fara (2001)/(2005) version) according to which dispositional ascriptions embed habitual sentences.

Fara sets up his claim by telling a story 'which brings out a fact about habituals to which [he] has not yet called attention, namely that they exhibit a measure of context-dependence.'⁴⁸ Here's how the story goes. Imagine that a certain relic – let's say a sixteenth century Venetian glass cup – is in the possession of museum officials, who know that *it breaks when struck* (that is, they come to know a certain habitual fact about the cup). They want to put the glass on display in the museum, but it is too fragile to handle, so one of the officials comes up with the visionary idea of protecting the glass with an internal support structure, i.e. Johnston packing. The officials put the Johnston packing in the cup and put the cup on display in the museum.

One night, however, a vandal breaks into the museum and steals the cup. The vandal is hell-bent on destroying the cup, but refuses to remove the Johnston packing to do so:

The vandal spends weeks hitting the cup with various large objects, attempting to break it. She could, of course, simply remove the packers' structure and easily smash the thing; but that would violate her scruples. Instead, she is intent on getting one up on the packers

⁴⁷ Fara 2005, p. 76.

⁴⁸ Fara 2005, p. 75.

by smashing the cup while it continues to be protected. Her attempts are in vain, however, and in the end she gives up, frustratedly exclaiming:

(30) The damn cup just doesn't break when struck!⁴⁹

Now, Fara observes, we seem to have a problem: the vandal's utterance, (30) is intuitively true, but didn't we say in setting up the story that the cup *does* break when struck?

But the appeal to context-sensitivity resolves this problem, Fara claims, and not only that – it also shows that the context-sensitivity of disposition ascriptions 'is entirely parasitic upon that of their embedded habitual sentences.'⁵⁰ He illustrates this by adding another element to the story. It's a little tricky to reconstruct his argument in detail without the text, so I'll reproduce the relevant passage in its entirety here:

Suppose museum specialists have been spying on the vandal, watching her efforts to smash the cup, efforts culminating in her exasperated utterance of (30). The museum specialists then saunter into the room, and proudly respond:

(31) Oh, but it does break when struck; that's why we protected it in the first place. We did such a good job, though, that you won't be able to break it.

It seems that their utterance of (31) is true; and given that it embeds [(21) 'The cup breaks when struck'], their utterance of that sentence must also be true. So we shouldn't just be asking, as we have been until now, whether the sentence (21) is true. We should be asking, instead, whether an utterance of that sentence expresses a truth in a given context. And as for (21), so for habitual sentences quite generally. Recall now the ascription

(19) The cup is disposed to break when struck.

Since the vandal would speak falsely if she were to utter the habitual (21), the Habitual Account requires that she would also speak falsely if she were to utter (19). On the other hand, the cup does have an intrinsic

⁴⁹ Fara 2005, p. 75.

⁵⁰ Fara 2005, p. 76.

property—say its weak molecular bonding—in virtue of which, if the museum specialists were to utter (21), they would speak truly. And so the Habitual Account predicts that they would also speak truly if they were to utter (19). What does this show? It certainly does not show that according to the Habitual Account Johnston’s cup both is and is not disposed to break when struck, any more than Fred’s true utterance of “I am happy” and Ted’s false utterance of that same sentence show that I am both happy and unhappy. What it shows, rather, is that disposition ascriptions, like the habitual sentences they embed, are context-dependent to a certain extent. The disposition ascription (19) expresses a falsehood in the context of the vandal’s utterance of (30), while in the context of the museum specialists’ utterance of (31) it expresses a truth. This context-dependence of disposition ascriptions runs in exact parallel to that of the corresponding habitual sentences, and so we have in the Habitual Account an explanation of the context dependence of disposition ascriptions—it is entirely parasitic on that of their embedded habitual sentences.⁵¹

To make things a little easier to follow, I’m going to rename these sentences and utterances in a way that makes it easy to keep them straight: we have the sentences (my names on the left, Fara’s on the right)

[DIS]	The cup is disposed to break when struck.	(19)
[HAB]	The cup breaks when struck.	(21)
¬[HAB]	The damn thing doesn’t break when struck!	(30)

and context *V* corresponding to the *vandal’s* context of utterance and *M* corresponding to the *museum officials’* context of utterance. According to the habitual analysis, [DIS] embeds [HAB]: ‘The cup is disposed to break when struck’ is analyzed as ‘The cup breaks when struck in virtue of an intrinsic property.’ Since [DIS] embeds [HAB] under a factive operator, (according to the habitual analysis) [DIS] → [HAB] and ¬[HAB] → ¬[DIS].

Fara claims (last paragraph in the block quote) that ‘since the vandal would speak falsely if she were to utter the habitual (21) [HAB], the Habitual Account requires that she would also speak falsely if she were to

⁵¹ Fara 2005, pp. 75-76.

utter (19) [DIS].’ To simplify, let’s drop the counterfactual and suppose that the vandal *does* say these things. We have the following:

First datum: The vandal speaks *falsely* in *V* by saying ‘The cup breaks when struck.’ ([HAB] is false in *V*).

First commitment: The vandal speaks *falsely* in *V* by saying ‘The cup is disposed to break when struck.’ ([DIS] is false in *V*).

I call these the *first datum* and *first commitment* since the former is Fara’s intuition about the truth value of the vandal’s utterance (which seems correct and robust) while the latter corresponds to *what the Habitual Account requires*, i.e. a commitment of Fara’s analysis.

Then Fara claims that ‘if the museum specialists were to utter (21) [HAB] they would speak truly. And so the Habitual Account predicts that they would also speak truly if they were to utter (19) [DIS].’ Again dropping the counterfactual we get

Second datum: The specialists speak *truly* in *M* by saying ‘The cup breaks when struck.’ ([HAB] is true in *M*).

Second commitment: The specialists speak *truly* in *M* by saying ‘The cup is disposed to break when struck.’ ([DIS] is true in *M*).

Fara then jumps to the conclusion that habituals explain the context-sensitivity of disposition ascriptions, but before getting there, we should ask whether each of these four theses is true. Keeping it short and sweet, we have the following four theses:

First datum: [HAB] is false in *V*.
First commitment: [DIS] is false in *V*.
Second datum: [HAB] is true in *M*.
Second commitment: [DIS] is true in *M*.

I think that Fara is correct on three out of the four theses: I think the first datum, second datum, and second commitment are true. But my intuition is that the *first commitment* is *false*: it seems to me that the vandal speaks *truly*, by ascribing the disposition to break when struck to the cup. After all, we've already said that the cup has that disposition; unless we suppose that it lost the disposition when the Johnston packing was placed inside, there doesn't seem to be any obvious reason to suppose that it isn't disposed to break when struck.

What is it about the context of the vandal that gives Fara the intuition that [DIS] is *false* in *V* despite being true in *M*? Unfortunately Fara doesn't defend the first commitment; he just says (correctly) that his account requires that it be true.

Let's ask this: *what property is it that the vandal attributes to the cup using [DIS] that it doesn't have?* Fara doesn't make any claims about this. I think it will be very useful to address this, since, though I *disagree* with Fara that the two tokenings of [DIS] ascribe different enough properties to make [DIS] true in *M* but false in *V*, I *agree* with Fara that the analysans sentence 'The cup breaks when struck in virtue of an intrinsic property' would be true in *M* but false in *V*. In the next two sections I'll explore two explanations of what is responsible for this context-sensitivity (both of which are couched in Fara's own semantics for habituals) and argue that on both explanations, the context-sensitivity of habitual sentences is *not* inherited by disposition ascriptions.

b. Fara's semantics for habituals and two explanations of the context-sensitivity of habituals

Presumably, the difference between the property expressed by 'The cup breaks when struck in virtue of an intrinsic property' in context *V* and the property expressed by the sentence in context *M* rests upon the difference in what is expressed by the habitual 'The cup breaks when struck' in the two contexts. I think Fara would surely agree with that, since he is committed to the claim that the habitual *does* express something

different in *V* and *M* (this follows from *the first datum* and *the second datum*.)

The question specifically concerns the context-dependence of habitual sentences. What makes the habitual ‘The cup breaks when struck’ true in *M* but false in *V*? Fara says little about the context-sensitivity of habituals (except that it explains the context-sensitivity of disposition ascriptions) in the (2005) paper, but much more in his (2001) dissertation.

I want to rule out one moderately plausible explanation which does not appeal to the context-sensitivity of habituals: we might suppose that the salient difference has to do with the denotation of ‘the cup’ in the different contexts. Perhaps the vandal is saying that a certain object, *x*, which includes *both the cup and the support structure*, breaks when struck, while the specialists are saying that a *different* object, *y*, which includes *only the cup and not the support structure*, breaks when struck. This is one way to make good sense of the fact that the vandal speaks falsely and the specialists speak truly, but, since it has nothing to do with the context-sensitivity of habitual sentences, it is almost certainly not what Fara has in mind. Let’s take this option off of the table; we send in our intrepid field semanticist for clarification, and both the vandal and specialists agree that they intend to refer to the same object – *just the cup without the support structure*.

What is it about the context-sensitivity of habitual sentences that makes the vandal’s habitual attribution false and the specialists habitual attribution true? I’d like to propose at least a partial explanation of the context-sensitivity of habituals. The explanation will have two elements: [1] I’ll give some context-shifting arguments that I think illustrate a pattern of context-sensitivity displayed and try to characterize that pattern, and [2] I’ll consider Fara’s (2001) semantics for habituals and suggest that this pattern of context-sensitivity can be expressed using the machinery of his semantics.

Fara (2001) advocates a quantificational semantics for habitual sentences, according to which habitual sentences like ‘The cup breaks when struck’ have the same logical form as adverbial quantification

sentences like ‘The cup always breaks when struck’ and ‘The cup often breaks when struck.’ Unlike sentences involving adverbial quantifiers, generics have an *unpronounced* quantifier, ‘meaning something like *generally* or *normally*,’ which Fara symbolizes ‘**GEN**.’⁵² He proposes the following analysis of ‘*N M*’s when *C*’:

GEN(If *s* is a situation in which *C*, then *N M*’s in *s*)

*where GEN is the quantifier associated with the unpronounced adverb, and the ‘if’-clause does not signal a conditional, but instead serves to indicate the restriction of the quantifier GEN.*⁵³

Fara gives as an example the analysis of ‘Tim hits the target when he shoots at it’ as **GEN**(If *x* is a situation in which Tim shoots at the target, then *x* is a situation in which Tim hits the target). Since Fara is explicit that the ‘if-then’ is a *restriction* of the quantifier rather than a conditional, I’ll use a conventional symbolism for restricted quantification to express the same thing as Fara:

[**GEN** *x*: Tim shoots at the target in *x*](Tim hits the target in *x*)

As Fara points out, there is a familiar problem for quantificational semantics for habituals: the sort of quantification that **GEN** expresses seems to vary drastically depending upon the habitual sentence. Fara points out, for instance, that while ‘Mary handles the mail from Antarctica’ may be true even though Mary has never handled mail from Antarctica, ‘Mary smokes’ seems to be true only if Mary has done quite a bit of smoking. Fara makes a very nice observation in motivating the quantificational account: he points out that quantifiers need not be *numerical* (where **Q** is a numerical quantifier just in case the *number* of *x*’s which are ϕ determines the truth conditions of $\ulcorner (\mathbf{Q}x)(\phi(x)) \urcorner$) nor *proportional* (where **Q** is a proportional quantifier just in case the *proportion* of *x*’s which are ϕ is determinative of the truth conditions of $\ulcorner (\mathbf{Q}x)(\phi(x)) \urcorner$). He gives some examples of non-numerical, non-

⁵² Fara 2001, p. 65. Fara’s italics.

⁵³ Fara 2001, p. 65.

proportional quantification, like ‘*a significant number of children wore shoes*’ and ‘*too many athletes take steroids*.’⁵⁴

Fara suggests that **GEN** is such a non-numerical, non-proportional quantifier. His semantics for **GEN** center crucially on the notions of *permissible* and *impermissible* exceptions to a habitual. Fara considers the habitual ‘Jeff goes skiing when the snow is fresh.’ This habitual may be true even though there have been exceptional cases in which the snow is fresh and Jeff *didn’t* go skiing; however, Fara claims, these exceptions must be *permissible*. Fara characterizes permissible exceptions to [**GEN** $x : \varphi(x)$]($\psi(x)$) as $\varphi(x) \ \& \ \neg\psi(x)$ cases c which have some feature **F** such that

- [1] Given that c has feature **F**, regardless of whether or not $\varphi(x)$, $\neg\psi(x)$.
- [2] Given that $\varphi(x)$, were it the case that c lacked **F**, it would be the case that $\psi(x)$.
- [3] **F** is not a normal feature of $\varphi(x)$ -cases (where ‘normal features’ may vary between different contexts of utterance).⁵⁵

Feature **F** is essentially preventive interference: the first rule stipulates that whenever **F** is present, it doesn’t matter to the habitual behavior whether or not the habitual trigger occurs, and the second rule stipulates that the habitual behavior would have occurred had **F** been absent. Fara explains the need for [3] in connection with the following case: take the habitual ‘Sally drives her BMW when it’s raining.’ Suppose Sally’s BMW is broken and almost never starts; as a result, Sally never drives the BMW, rain or not. This feature satisfies [1] – given that the car is broken, raining or not, Sally won’t drive it; and if we suppose that Sally would otherwise drive the BMW in the rain, it satisfies [2] as well. But the habitual intuitively isn’t true, since Sally never drives the BMW when it’s raining, even though the habitual trigger occurs all the time. This is because, Fara argues, permissible exceptions cannot arise because of a feature that is part of the normal background conditions – that would make the exceptions quite unexceptionable!

Returning to the case of the vandal and the specialists, one plausible explanation why [HAB] is true in M but false in V appeals to this ‘normal

⁵⁴ Fara 2001, p. 59.

⁵⁵ Fara 2001, p. 76.

feature' restriction. In both contexts, there are the same exceptions to the habitual ('The cup breaks when struck') – the many cases in which the vandal strikes it but it does not break. In both contexts, there is a feature which satisfies [1] and [2] – the presence of the Johnston packing. However, it does seem plausible that the different contexts vary with respect to whether or not the Johnston packing is a normal feature. *Prima facie*, the Johnston packing constitutes a more 'normal' background condition in the context of the vandal's utterance than in the context of the specialists: after all, the packing has been in the cup for as long as the vandal has been in contact with it and he hasn't made any attempt to take it out; on the other hand, the specialists first encountered the cup without the Johnston packing and then chose to put the Johnston packing in it. Thus, it doesn't seem like a stretch to suppose that Fara's system gives a nice explanation of why [HAB] is true in *M* but false in *V*: in *M* the failure of the cup to break on the occasions of the vandal striking it are *permissible exceptions*, since the Johnston packing is a feature of those cases satisfying [1], [2], and [3], while in *V* these cases are *impermissible exceptions*, since the Johnston packing fails to satisfy [3] in light of the fact that it is a *normal* background condition from the point of view of the vandal.

However, though this application represents a potential explanatory success of Fara's habitual semantics, it is not a good result for his habitual analysis of dispositional predicates. Fara singles out condition [3] as an explanation of the context-sensitivity of habituals. But disposition ascriptions do *not* show this kind of context-sensitivity: whether or not a *mask* (e.g. the Johnston packing in this case) is *normal* or *abnormal* seems to make no difference. Compare the example from Fara that I appealed to in my argument against his analysis in Part I: 'Mark is disposed to smoke after work' and 'Mark smokes after work.' This case looks like another which is good for Fara's habitual semantics but bad for his disposition semantics: though criterion [3] looks like a good explanation of why the preventative interference of Mark's health-conscious fiancée is not an acceptable exception-permitting feature, in

light of the fact that it is *normal* for situations satisfying the habitual trigger ('after work'), the fact that this interference is normal in no way pulls intuitions (at least mine) toward denying the truth of the disposition ascription.

Fara was actually (to my knowledge) the first to point out that masks can be *quite normal*:

*Dispositions of objects are being masked all the time. I'm disposed to go to sleep when I'm tired; but this disposition is sometimes masked by too much street noise. Cylinders of rubber are disposed to roll when placed on an inclined plane; but this disposition can be masked by applying a car's brakes.*⁵⁶

Though Fara intends to highlight that masks can be normal in the sense of not being exotic, his comments also suggest that they might sometimes be *normal* in the sense of 'normal' in condition [3].

For example, I am disposed to sleep when I'm tired (disposition ascription), and thanks to the absence of street noise in Middletown, I sleep when I'm tired (habitual). However, if I move to New York, street noise will become a normal feature of situations in which I'm tired, and will accordingly result in my no longer sleeping when I'm tired (negation of the habitual). This can be explained nicely by criterion [3]. However, does the move to New York really result in my *losing* the disposition to sleep when I'm tired? Intuitively not: I'm just as disposed to sleep when I'm tired as I ever was, though this disposition is unfortunately chronically masked. Additionally, since this disposition seems *not* to be one of those (relational dispositions) that Fara thinks violate *intrinsicness*, I shouldn't lose the disposition merely because of an extrinsic change. But if the disposition truly embeds the habitual, then when the street-noise disruption becomes a normal feature of situations in which I'm tired, not only will the habitual be false, but the disposition which embeds it will also be false.

A second dimension of context-sensitivity exhibited by habituals has to do with *generality* and can be modeled in Fara's habitual semantics by

⁵⁶ Fara 2005, p. 50.

supposing that the domain of situations bound by the quantifier **GEN** can vary between contexts.

To illustrate this variety of context-dependence, I'll use two examples from Fara. For the first example, consider the sentence 'Jeff goes skiing when the snow is fresh.'

Jeff is an avid skier, and he lives in Tahoe during the winter months so that he can ski every day should he want to. Jeff is particularly fond of freshly fallen powder snow, and thus [the sentence 'Jeff goes skiing when the snow is fresh' is true]... But let us alter the example. Suppose that because of an accident Jeff has been in hospital for the past three weeks suffering from a torn knee ligament. Unfortunately, then, Jeff is unable to go skiing, whatever the condition of the snow. In such a context, it seems, an utterance of ['Jeff goes skiing when the snow is fresh'] would be false. And our semantics predicts exactly that. For although the exceptions to ['Jeff goes skiing when the snow is fresh'] have the feature of being situations in which Jeff's knee is injured... given the context just described this is a normal feature for situations to have, and so the third Permissibility Constraint is unsatisfied.⁵⁷

I am somewhat puzzled by Fara's claim here. He certainly intends *contexts* to be *contexts of utterance*; just prior to this example, he says that 'whether or not a feature is normal will depend, in part, on the context in which a given habitual is *uttered*'⁵⁸ (my italics). However, in sketching the example, Fara *doesn't describe any contexts of utterances* – he just describes what has happened to Jeff. Radical context-sensitivity can depend upon all sorts of highly specific contextual factors: conversational dynamics and conversational goals, the common knowledge of conversational participants, salient objects in the environment, and so on. But Fara doesn't specify this sort of context, so he doesn't seem to be in a position to say what is expressed by an utterance of the sentence in context.

I think that even after the accident, the habitual sentence 'Jeff goes skiing when the snow is fresh' can express a truth in some contexts of

⁵⁷ Fara 2001, p. 77.

⁵⁸ Fara 2001, p. 76.

utterance. Here's an example. Suppose Jeff is a famous celebrity and a reporter calls him up for an interview. The reporter asks him about his hobbies. 'I like skiing a lot,' Jeff says. The reporter, also an avid skier, asks Jeff how particular he is about snow quality when skiing. 'Very particular in fact; I go skiing when the snow is fresh.' It seems to me that Jeff could speak the literal truth by uttering that habitual sentence. And yet, as Fara claims, he could also speak falsely by saying 'I go skiing when the snow is fresh, because lately he hasn't been skiing much, regardless of the freshness of the snow.'

The contextual difference between two cases seems to have to do with *how long a period of time* Jeff intends speak about. We can observe similar context-sensitivity of habituais having to do with *location* rather than time. 'It snows in the winter' is true if a speaker is talking about Connecticut but false if she's talking about Mexico. In these examples, the very same habitual sentence may be true or false depending on which cases count – just the time since the accident? Jeff's life up to and after the accident? Winters in Mexico? Winters in Connecticut?

There's a very nice way to model this on Fara's analysis: what varies contextually is the *domain* of situations bound by the **GEN** quantifier. In the interview case, quantification is over situations which include those before the accident; in the winter snow cases, the domain ranges over situations in the winter in Mexico and situations in the winter in Connecticut, respectively.

Here's another example from Fara.⁵⁹ Suppose there is a king whose kingdom has fallen on hard times. There is a famine so bad that even the king's diet is disrupted. Before the famine, the king ate fine fish and meat; since the famine, he's been eating potato stew. Consider two habitual sentences: 'The king eats fine fish and meat when he's hungry,' and 'The king eats potato stew when he's hungry.' It seems to me that either of these two sentences – and either of their negations – could express a truth, given the right sort of context of utterance. Suppose one of the king's scouts runs into a friendly foreigner in the woods. 'What's your kingdom

⁵⁹ Fara 2005, p. 65.

like?’ says the scout. ‘Oh it’s very nice,’ says the scout. ‘The king eats fine fish and meat when he’s hungry, and even the peasants are well fed.’ The scout presumably could speak the literal truth with such a sentence – what is expressed is that the kingdom, over some broader period of time than the famine, is as the scout has characterized it. Suppose the friendly foreigner presses on: ‘Well that sounds quite lovely, but I’ve heard that a number of kingdoms have been ravaged by the drought. How have things been lately?’ ‘Not so good,’ the scout responds. ‘Things have gotten so bad that even the king eats potato stew when he’s hungry.’

It seems that this might be explained by supposing that the situation quantifier **GEN** ranges over different domains of situations in which the king is hungry: in the fish-and-meat case, over something like all of the situations during the king’s reign, but in the potato stew case, over the more restricted class of cases occurring since the start of the famine.

Returning to the vandal and the specialists, this offers another way we might characterize what is different between *M* and *V*: in *V*, the vandal uses the habitual to express something about the behavior of the cup since he’s come in contact with it, whereas the specialists express something about the behavior of the up at a higher level of diachronic generality. Here, the context-sensitivity of dispositional predicates *partially* parallels that of the habituals. Dispositional predicates can be used to attribute tensed properties (‘Marge is a little irritable *right now*’) but there are also tenseless constructions which ascribe properties to enduring entities. These constructions can vary in diachronic generality as habituals do in the Jeff the skier and the king’s famine cases: ‘Barry is quite loquacious’ can be used to ascribe to Barry the property of being a loquacious person (i.e. disposed to respond in a certain way to situations ranging over a long period of time) or of being loquacious tonight (i.e. disposed to respond in a certain way to situations ranging over a short period of time).

However, it doesn’t seem that dispositions can vary synchronically across changes in location as habituals can. There is a very simple explanation for this, which has cropped up several times in connection with Fara’s account: the truth conditions of disposition ascriptions are

invariant across changes *extrinsic* to the object ascribed the disposition. While whether an object x habitually R's when T'd *can vary* across scenarios without any intrinsic change to x , whether x is disposed to R when T'd *cannot*. This is the crux of why the habitual analysis of dispositional predicates fails.

On either of the above characterizations of the context-sensitivity of habituals, the first commitment – that [DIS] is false in V – is incorrect. Suppose the relevant difference between M and V is a difference in the intended level of diachronic generality; the disposition attribution could parallel this contextual variation, but unless the disposition of the cup is *lost* at some time, then [DIS] will still be true at in V . Intuitively, the disposition of the cup to break when struck is never lost, although it is *masked* by the Johnston packing. Suppose on the other hand that the *normality* of the masking in V results in the difference in [HAB] between M and V (as predicted by Fara's criterion [3]): the disposition ascription *doesn't* display this sort of context-dependence, so this would provide no reason for [DIS] to be false in V .

In short, Fara's claim that [DIS] is false in V is inconsistent with his own (highly plausible!) weak intrinsicness assumption – that *non-relational* dispositions are not lost or gained without intrinsic change, and are shared by intrinsic duplicates. For this reason, I reject Fara's strong claim that the *sole* source of the context-sensitivity of dispositional predicates is the context-sensitivity of embedded habitual sentences. Indeed, the arguments above suggest that we should be skeptical of weaker claims to the effect that even *some* of the context-sensitivity of dispositional predicates should be explained in terms of the context-sensitivity of habituals.

11. Manley and Wasserman's proportional account

Manley and Wasserman's analysis of the context-sensitivity of dispositional predicates consists of two claims: the first is that the context-sensitivity of dispositional predicates is exhausted by their *gradability*; the second is that the gradability of dispositional predicates tracks the

proportion of cases in which objects manifest the disposition in response to the test condition. I'll argue that both of these theses are incorrect: the context-sensitivity of dispositional predicates outruns gradability, and Manley and Wasserman's mechanism does not correctly predict the gradability of dispositional predicates.

Bonevac, Dever, and Sosa mention in a footnote to their discussion of Manley and Wasserman that the context-sensitivity of dispositional predicates outruns gradability.

Perhaps no one has suggested a plausible mechanism for the context dependence of dispositional predicates, and perhaps this is to be lamented... We note in passing that Manley and Wasserman's own positive account of the mechanism fails – context-dependency of dispositional predicates cannot be a matter of selecting a threshold for an underlying comparative, since the dimensionality of the context-sensitivity outruns that of the comparative. An aerospace engineer, when calling a material fragile, has in mind its behavior at high temperatures on re-entry, while a sculptor, when calling a material fragile, has in mind its tendency to shatter when chiseled. Neither engineer nor sculptor is using 'fragile' to mean more fragile than the other; they are instead focusing on different aspects of a complex range of fragility-relevant features.⁶⁰

This seems to me to be right. Context-shifting arguments can be given for context-sensitivity that appears to have little in common with the context-sensitivity of exemplary gradable adjectives. Manley and Wasserman might defend their view by claiming that, in point of fact, dispositional predicates do not show non-gradable context sensitivity, but it is not clear how such a move could be motivated on methodological grounds since both gradable and non-gradable context-sensitivity of dispositional predicates were argued for using the same sort of context-shifting arguments. Context-shifting arguments can be given for non-comparative context-sensitivity of dispositional predicates: Bonevac, Dever, and Sosa's engineer and sculptor example, the pottery class example, the story of Johnny and Max, other 'fragile' and 'poisonous' cases mentioned above,

⁶⁰ Bonevac, Dever, and Sosa 2009, p. 7.

and so on. Thus, the thesis that the context-sensitivity of dispositional predicates is merely comparative is not well supported by the evidence from context-shifting arguments.

My second claim is that Manley and Wasserman's mechanism fails to account even for the gradability of dispositional predicates. I used an example in Part I in arguing for the exclusion thesis that I would like to elaborate here as an argument against Manley and Wasserman's proposed mechanism for the gradability of dispositional predicates. Recall that Manley and Wasserman defend an analysis of dispositional predicates according to which x is more D than y iff x gives the D -response in a greater proportion of nomologically possible no-intrinsic-change D -tests than y . Crucially, mask and fink cases are *included* in the D -tests and mimics are *included* in the D -responses. Now recall the example involving two poisons IP and DP (*indiscreet poison* and *discreet poison*), similar in every 'poisonous'-relevant respect (same lethal dose, same time to cause death, same lethal modes of administration, etc.) *except* that IP -exposure has a certain highly visible side effect that DP -exposure does not: it turns the victim's skin purple! As a result of this side effect and the fact that masking cases are *included* in the T -events, $R(DP)$ -events will comprise a much *smaller* proportion of the $T(DP)$ -events than the proportion of $R(IP)$ -events among $T(IP)$ -events. A far greater number of the IP cases will be *antidote* cases, since the obvious side effect makes exposure to IP very easy to identify. But if Manley and Wasserman's proportional account is correct, then IP is much more poisonous than DP . This seems highly unintuitive.

Fortunately, it appears that the mechanism fails not because the proportional analysis of gradability is incorrect, but because finks, masks, and mimics have been incorrectly included in the proportion-relevant events. Both the extensional failure examined in Part I and the context-sensitivity problem here seem to result from Manley and Wasserman's insistence that the relevant possibilities included in the proportional

account ‘can [and should] include masking and finkish cases; these need not in any sense be ‘ignored’ or ruled as irrelevant by context.’⁶¹

On the other hand, the move from counterfactuals to direct quantification over test- and response-events seems to be a significant advance: it avoids the closeness and centering counterexamples and brings the set of possibilities relevant to the analysis closer to the set of possibilities intuitively relevant to the analysis (rather than focusing in on an unintuitively narrow slice – the inner sphere – as the Lewis conditional accounts do).

If things were simple, we could just take the good from Manley and Wasserman (direct quantification over possible test- and response-events) and ditch the bad (failure to exclude finks, masks, and mimics) by supplementing the account with a mechanism for fink-, mask-, and mimic-exclusion. However, we have seen that the mechanisms for fink-, mask-, and mimic- exclusion familiar from the conditional accounts – *precisification* and *qualification* – seemed to be inadequate to the task. Thus, I’ll begin Part III by proposing a new mechanism for fink-, mask-, and mimic-exclusion, which I will then put to use in an account which builds upon Manley and Wasserman’s proportional analysis.

⁶¹ Manley and Wasserman 2007, p. 6.

PART III: PARADIGMS AND INTERPRETATION

12. Toward an *exclusionary* proportional account

In Parts I and II, I critically studied six semantic analyses of dispositional predicates: the simple conditional analysis, the precisified conditional analysis, the qualified conditional analysis, the disjunctive conditional analysis, the habitual analysis, and the proportional analysis. In Part I, I argued that none of the analyses are extensionally adequate. In Part II, I argued that none of the analyses provides a complete mechanism for the context-sensitivity of dispositional predicates. In this part of the paper, I will propose a new analysis which models the interpretation of dispositional predicates in context, and argue that such an analysis is extensionally adequate and provides a mechanism for the context-sensitivity of dispositional predicates. I'll then argue that the account has applications to some other topics related to the semantics of dispositional predicates.

My account will build on the model provided by Manley and Wasserman's proportional analysis. Their analysis has two main advantages over the conditional analyses. First, in contrast to the unintuitively small slice of modal space relevant to conditional accounts – the *inner sphere* of worlds satisfying the disposition test condition – the portion of modal space relevant to Manley and Wasserman's analysis is intuitively sound: all nomologically possible, no intrinsic change, test condition-satisfying events count as T-events. Second, their account can handle the accidental closeness and strong centering cases that were counterexamples to the conditional analyses. The primary disadvantage of their account is their explicit commitment to *include* fink and masking cases in the T-events and mimics in the R-events. My initial idea is thus to modify Manley and Wasserman's proportional account to exclude finks, masks, and mimics.

Unfortunately, it's not clear *how* to do this, for two reasons in particular. The first is that Manley and Wasserman analyze *explicit disposition ascriptions* ('*x* is disposed to R when T'd') rather than

dispositional predicate attributions (though they seem to suggest that the account can handle these as well) and as a result do not explain how T-events and R-events get selected for dispositional predicates (which, unlike explicit disposition ascriptions, don't come with explicitly specified test and response conditions). The second reason is that both of the proposed mechanisms for fink-, mask-, and mimic-exclusion we've considered so far (in connection with the conditional analyses) – *precisification* and '*normal conditions*' qualification – haven't gotten the job done.

These two problems comprise two agenda for producing an exclusionary proportional account: first, associate sets of T-events and R-events with particular dispositional predicates, and second, do so in such a way as to exclude finks and masks from the T-events and mimics from the R-events.

There was one other independent problem with Manley and Wasserman's analysis: though they do give a plausible mechanism for the gradability of dispositional predicates (*modulo* fink-, mask-, and mimic-exclusion), they do *not* account for other, non-comparative dimensions of context-sensitivity of dispositional predicates. Though it is entirely possible that Manley and Wasserman have principled reasons for taking dispositional predicates to be gradable but not context-sensitive in non-comparative respects, there is good reason to draw a different conclusion from their papers: they use *context-shifting arguments* without qualification to establish context-sensitivity. The use of context-shifting arguments as sufficient evidence for context-sensitivity leads to conclusions requiring 'radical' context-sensitivity that far outruns gradability.⁶² I have followed Fara, Manley and Wasserman, and Bonevac, Dever, and Sosa in taking context-shifting arguments to be sufficient evidence for semantic context-dependence; this commitment leads to recognizing dispositional predicates' context-sensitivity well beyond the comparative.

Thus our three agenda are as follows: modify Manley and Wasserman's account in order to [1] specify T- and R-events for

⁶² See Cappelen and Lepore (2005), Chapter 3 for an excellent discussion of the instability of 'moderate' contextualism.

dispositional predicates, which, unlike explicit disposition ascriptions, do not have ‘built-in’ test and response conditions, [2] exclude finks and masks from the T-events and mimics from the R-events, and [3] account for the non-gradable context-sensitivity of dispositional predicates.

13. The basic model of interpretation

Though the basic model I’m proposing gives results similar to those of Manley and Wasserman’s proportional analysis, it takes a radically different form: it is a model of *the interpretation of dispositional predicates in context*. What I’m modeling is the semantic interpretation of the content expressed by dispositional predicates uttered in a sentence in context. If we can model how interpreters are able to use contextual information to interpret utterances containing dispositional predicates, then this should suffice to give us a relatively clear picture of their semantics.

Interpreters take utterances to (literally, semantically) express some content, and in doing so ‘choose’ between various *candidate* contents. I’ll focus in on the content expressed by dispositional predicates rather than the content expressed by the utterances as a whole. The *candidate interpretations* interpreters adjudicate between are contents that the interpreter may take the dispositional predicate to express. Candidate interpretations will have differing degrees of *interpretational eligibility*. The more eligible some interpretation it is, the ‘better’ it is as an interpretation. Low interpretational eligibility corresponds to being a *bad* interpretation; high eligibility corresponds to being a *good* interpretation.

Because dispositional predicates intuitively ascribe properties to give a certain response to a certain test condition, candidate contents will be individuated in terms of the choice of T-events (test cases) and R-events (response cases). For instance, if by saying ‘That vase is fragile,’ I ascribe to the vase the disposition to break in response to moderate physical stress, then this content can be identified with the pair $\langle T_\alpha, R_\alpha \rangle$ where T_α , the set of test events, is the set of nomologically possible, no-intrinsic-change events in which the vase is subjected to moderate physical stress,

and R_α , the set of response events, is the set of nomologically possible, no intrinsic-change events in which the vase breaks in response to moderate physical stress.⁶³

Two dimensions of constraints are represented within the final model: context-invariant lexical constraints, and interpretational constraints. I'll introduce the context-invariant lexical constraints first; the shortcomings of a one-dimensional model will quickly become clear.

The specific lexical constraints will be modeled using *paradigms* associated with each dispositional predicate. Paradigms are type-exemplars in the tradition of prototypes and stereotypes. However, rather than appealing to paradigmatically disposed objects (e.g. a paradigmatically fragile vase), I will use *paradigmatic events of manifestation* in the model. A paradigmatic manifestation of fragility is an event which is typical of manifestations of fragility. An event in which a certain fragile vase is knocked off of a mantel and falls onto a marble floor, shattering into lots of tiny shards is intuitively a paradigmatic manifestation of fragility: it involves a typical fragility test, moderate physical stress, and a typical fragility response, physically breaking in response to moderate physical stress. On the other hand, a manifestation of fragility in which a clay pot deforms in response to heat stress in a kiln is plausibly somewhat less paradigmatic, since heat stress is intuitively a less typical fragility-test than moderate physical stress.

Paradigmatic events of manifestation play an important role in determining which events count as T-events and R-events. In this model, rather than specifying T- and R-events non-relationally via a qualitative description of test and response conditions (e.g. the fragility tests for x are all of the nomologically possible, no- intrinsic-change events in which x is stressed) they are determined *relationally* by *resembling the paradigm* in the right ways. T-events are those which resemble the paradigm in respect of having similar *test conditions*, and R-events are those which resemble

⁶³ Almost entirely throughout, I will represent disposition contents using only two values, $\langle T_\alpha, R_\alpha \rangle$, although representing both Manley and Wasserman's complete analysis and my own requires adding a third value, representing either 'suitable proportion' (Manley and Wasserman) or 'suitable distribution' (in my version). I drop the third value often just for the sake of simplicity.

the paradigm in respect of having the same or similar *test conditions and response conditions*.

Let's label paradigmatic manifestations of a disposition 'e_P'; call the paradigmatic manifestation of fragility described above 'e_{P,F}.' The event e_{P,F} typifies a characteristic fragility test and response: an object is subjected to a typical fragility test (moderate physical stress) and gives a typical fragility response (breaking in response to moderate physical stress). 'Fragility'-associated T-events will be all and only those in which an object is subjected to a test stimulus appropriately similar to the test stimulus in the paradigmatic event; 'fragility'-associated R-events will be all and only those in which an object is subjected to a test stimulus appropriately similar to the paradigm and gives a response appropriately similar to the response given in the paradigmatic event.

How 'appropriate similarity' is interpreted will evolve somewhat as I develop the model. For now, let's consider the prospects of a relatively *restrictive* notion of appropriate similarity, such that only events of *moderate physical stress* will T-resemble the paradigmatic event e_{P,F} (which involves moderate physical stress on the vase) and only events of *breaking in response to moderate physical stress* R-resemble e_{P,F}.

Here's how the model goes. An interpreter has associated with each dispositional predicate 'D' a paradigmatic event of manifestation e_{P,D}. A candidate interpretation [α] of a dispositional predicate (in an utterance in context) will correspond to the candidate pair <T_α, R_α> where T_α is the set of nomologically possible, no-intrinsic-change-to-x T(x)-events and R_α is the set of R(x)-events (according to interpretation [α]). For instance, suppose someone says 'That vase is fragile.' Take two candidate interpretations: [α] is an interpretation according to which the speaker ascribes to the vase the disposition to physically break in response to *moderate physical stress*, and [β] is an interpretation according to which the speaker ascribes the disposition to physically break in response to *severe heat stress*. These two interpretations will be individuated on the basis of their associated candidate T- and R-event sets, <T_α, R_α> and <T_β, R_β> where T_α is the set of nomologically possible, no-intrinsic-change

events in which the vase is moderately physically stressed, T_β is the set of nomologically possible, no-intrinsic-change events in which the vase faces severe heat stress, R_α is the set of nomologically possible, no-intrinsic-change events in which the vase is moderately stressed and breaks in response, and R_β is the set of nomologically possible, no-intrinsic-change events in which the vase faces severe heat stress and breaks in response.

The model tells us that interpretation $[\alpha]$ is an eligible interpretation of the utterance ‘That vase is fragile,’ since the T-events all involve moderate physical stress on the vase and the R-events all involve the vase breaking in response to moderate physical stress. However, the model tells us that interpretation $[\beta]$ is not eligible, since T_β is a set of *heat stress* rather than *moderate physical stress* events. The model will give the correct result for utterances in which ‘fragile’ is used to ascribe the disposition to break in response to moderate physical stress; but ‘fragile’ can also be used to ascribe all sorts of other properties (like the disposition to break in response to heat stress).

This is the first major problem with the model: it does not account for the context-sensitivity of dispositional predicates. Presumably the sentence ‘That vase is fragile’ could be used to ascribe the $[\alpha]$ -property (the disposition to break in response to moderate physical stress) or the $[\beta]$ -property (the disposition to break in response to severe heat stress) in some context or other. I will address this problem with two moves. The first will involve loosening the interpretation of ‘appropriate similarity’ that events in the T- and R-event sets must bear to paradigmatic events to count as minimally lexically admissible. This will allow for many different minimally admissible pairs $\langle T_\alpha, R_\alpha \rangle$, $\langle T_\beta, R_\beta \rangle$, $\langle T_\gamma, R_\gamma \rangle$, and so on, for each dispositional predicate, corresponding, roughly, to the contents expressible by the dispositional predicate in some context or other.⁶⁴ The second move is the introduction of the second dimension of the model – general interpretational constraints – which allow an interpreter to pick up the slack between minimally lexically admissible contents and the actual

⁶⁴ ‘Roughly’ because as we will see, despite satisfying these minimal lexical requirements, some contents will be so interpretationally unfit for any context that they can’t actually be (literally, semantically) expressed by the predicate.

content expressed by using specific features of the context to adjudicate between these various minimally admissible interpretations.

The second major problem with the model is that it does not yet provide a mechanism for fink-, mask-, and mimic-exclusion. Since fink and mask cases fall under the same rough-and-ready test condition as the paradigmatic event, they might be counted as admissible T-events by the weak similarity requirement, and since mimicking cases fall under the same rough-and-ready response condition as the paradigmatic event, they might be counted as admissible R-events.

14. Interpretational constraints

I will first address the context-sensitivity problem. Recall the pottery classroom example: Sasha and Malia are in pottery class and have each sculpted pots. Sasha picks up Malia's pot, M , and Malia says 'Be careful! That is very fragile.' This is utterance u_1 . Later, the teacher, Mr. Potter, goes around examining students' work one last time before firing the pots in the kiln. He picks up Sasha's pot and furrows his brow: 'Sasha, this pot is too fragile to fire as it is – see this weak spot where the clay is thin? It would almost certainly shatter in the kiln.' Malia asks – 'Mr. Potter, is my pot fragile as well?' Mr. Potter examines M . 'No, Malia. This pot isn't fragile at all. It will be fine in the kiln.'

Now if an interpreter were to use the sort of reasoning reconstructed in the preliminary model above, she would interpret 'fragile' in both u_1 and u_2 as expressing the disposition to break in response to moderate physical stress. While that is the correct interpretation for u_1 , it is *not* the correct interpretation for u_2 .

The problem is simple (although the solution is not): 'fragile' can be used in context to ascribe various properties, like the property to break in response to moderate physical stress (as it does in u_1) or the property to break in response to severe heat stress (as it does in u_2). Test events of the former property resemble $e_{P,F}$ but test events of the latter don't (according to our 'strict' account of resemblance), since $e_{P,F}$ involves moderate physical stress but not severe heat stress. With a strict notion of

resemblance in place, $e_{P,F}$ is only an exemplar of ‘moderate physical fragility,’ one of an indefinite number of properties that can be expressed by the dispositional predicate ‘fragile.’ How can a paradigmatic manifestation of moderate physical fragility serve in the interpretation of ascriptions other properties using the term ‘fragile’, like the ascription of the disposition to break in response to severe heat stress in u_2 ?

The answer is twofold: first, we must *loosen* the similarity constraint so that both the *moderate physical stress* and *severe heat stress* T- and R-sets are admissible, and second, we must introduce the second, interpretational dimension in order to explain how interpreters adjudicate between multiple lexically admissible interpretations in a specific context. I’ll go into more detail later on this revised similarity constraint (which I’ll call *sortal-independent category resemblance*). For now I’ll leave the constraint at a relatively intuitive level and illustrate by example: for ‘fragile,’ events of stress – physical stress, emotional stress, environmental stress, geopolitical stress, etc. – all belong to T-sets that would be counted as appropriately similar to any intuitively plausible paradigmatic manifestation of fragility, but events which do not involve any sort of stress do not.

There will be a number of minimally admissible contents determined by this vague appropriate similarity relation: in the case of fragile, the contents associated with any pair of event sets including only nomologically possible, no-intrinsic-change events of stress (for the T-event sets) and events of breaking in response to stress (for the R-event sets). Some contents will correspond to familiar interpretations like ‘moderate physical fragility’ and ‘heat stress fragility’; others will be gerrymandered, unnatural event classes which include various stresses lacking some unified underlying character. Choosing between all of these minimally admissible contents within a given context will require *interpretational constraints*: a dimension in the model that evaluates how good an interpretation is in context.

A familiar ‘master rule’ of interpretational goodness is the *principle of charity*:

*We make maximal sense of the words and thoughts of others when we interpret in a way that optimizes agreement (this includes room, as we said, for explicable error, i.e. differences of opinion.)*⁶⁵

In the pottery classroom case above, it is relatively easy to see why in u_1 ‘fragile’ expresses the property to break in response to moderate physical stress and in u_2 it expresses the property to break in response to extreme heat stress: in each case, those interpretations make greatest light of the speaker’s highly visible intentions and communicative goals. In the context of u_1 , Malia is trying to prevent Sasha from carelessly destroying her pot by dropping it or smashing it in her hands, so it makes most sense for her to be alerting her to the fact that the M is disposed to break in response to the sort of stress it would undergo in those events. In the context of u_2 , Mr. Potter intends to reassure Malia that her pot is ready to go into the kiln, and can do so by telling her that it *doesn’t* have a disposition to break in response to that sort of event – an event of severe heat stress.

There is nothing particularly mysterious about the use of contextual factors in interpretation in this specific case. However, providing a *general* and *systematic* account of the way context is used in the interpretation of speech acts is no small feat, and I will not attempt it here. Rather, I will lean on three principles in the spirit of the principle of charity that can be usefully applied to the case of dispositional predicates and which constitute a simple but powerful basic set of interpretational rules: *scrutability*, *interest*, and *plausibility*. Then I’ll consider a slightly more complex case in which we can apply the three rules.

Scrutability (for lack of a better term) is something of a misnomer which collapses two notions: a property which is scrutable is [1] the sort of property that can be identified and [2] the sort of property which is projectible and can be used to make informative generalizations.

Interest has to do with contextually salient factors and the attitudes of conversational participants. Roughly, an interpretation of a property ascription is interesting in a context if the conversational participants care about whether or not the object has the property.

⁶⁵ Davidson 1974/2006, p. 207.

A *plausible* interpretation of an utterance in context *c* is one which interprets the speaker as having said something which it is plausible (to the interpreter) that she say in *c*.

Here's an example to illustrate the interpretational constraints. A group of pilots are out on the tarmac preparing for some training exercises. They are talking about the emotional states of various other pilots in their cohort, and Jenny the pilot utters the following: 'Yeah, I'm worried about Max – he's fragile. I don't know if he's going to hold up after all of this grueling training. Actually, come to think of it, that plane over there [*gestures*] that he's flying is pretty fragile too – I don't know if we want either of them in combat.' Call the utterance in which Jenny the pilot says Max is fragile u_3 and the utterance in which she says Max's fighter jet is fragile u_4 . Consider three possible interpretations of 'fragile' (for either utterance), each of which is a content expressible by 'fragile' in some context: [α] ascribes the property *emotional fragility*, the disposition to break down emotionally in response to psychological stress; [β] ascribes the property *moderate physical fragility*, the disposition to break physically in response to moderate physical stress (e.g. of the sort that fragile vases might break in response to); [γ] ascribes the property *extreme physical fragility*, the disposition to break physically in response to extreme physical stress. The corresponding T- and R-event sets look like this:

[α]- u_3 : <*psychologically stressed*(Max)-events, *emotional meltdown*(Max)-events>
 [β]- u_3 : <*moderately physically stressed*(Max)-events, *breaks*(Max)-events>
 [γ]- u_3 : <*extremely physically stressed*(Max)-events, *breaks*(Max)-events>
 [α]- u_4 : <*psychologically stressed*(jet)-events, *emotional meltdown*(jet)-events>
 [β]- u_4 : <*moderately physically stressed*(jet)-events, *breaks*(jet)-events>
 [γ]- u_4 : <*extremely physically stressed*(jet)-events, *breaks*(jet)-events>

We can straightforwardly model the eligibility of these candidate interpretations of u_3 and u_4 by appealing to our interpretational constraints. Intuitively, [α] is a good interpretation of u_3 and an awful interpretation of u_4 ; [β] is not a very good interpretation of either utterance; and [γ] is a good interpretation of u_4 but an awful interpretation of u_3 .

Why is [α] a good interpretation of u_3 and an awful interpretation of u_4 ? It's simple with our interpretational rules in place: it would be wildly implausible to ascribe such an inscrutable emotional property to a jet; on the other hand, emotional fragility is a very scrutable property in persons, and of particular interest in the context. [γ] is just the opposite: all persons 'break' in response to extreme physical stress, so though the property is scrutable for persons, it isn't very interesting, and not very plausible (assuming the cooperativeness of the speaker). [β] would be of interest if either the jet or Max had that property, but since it's very implausible that either does, it isn't a very good interpretation.

In the pottery classroom example, both the *heat stress fragility* and *moderate physical fragility* interpretations are scrutable, but which property the conversational participants are interested in is very different between the context of u_1 and u_2 . In the context of u_1 , Malia is concerned that her sister will break her pot by subjecting it to moderate physical stress; in u_2 , Mr. Potter is concerned about the readiness of the students' pots for the kiln. This difference in interest seems to be a clear reason for the difference in the eligibility of the two interpretations in the two contexts.

Remember that these interpretational values correspond to a *content* represented by a pair $\langle T_\alpha, R_\alpha \rangle$; the content corresponds to the property of a suitable proportion of $T_\alpha(x)$ cases being $R_\alpha(x)$ cases. The interpretational dimension serves another purpose in addition to adjudicating between minimally lexically eligible T- and R-event sets: it also rules out irregular test and response pairs which lack underlying patterns of T- and R-events, like a T_α set which involves some but not all cases of heat stress, as well as some of minute physical stress, and some cases of extreme physical stress. The properties determined by these irregular pairs – dispositions to give a certain inscrutable pattern of response to a random set of tests – will be quite implausible, unnatural, and uninteresting candidate contents. Only those sets which are comprised of events with shared underlying characteristics will be likely to satisfy the constraints.

15. Fink-, mask-, and mimic-exclusion

I have argued at length that fink and masking cases should be excluded from the T-events and mimics should be excluded from the R-events; however, the preliminary version of the model, like Manley and Wasserman's proportional account, does not appear to exclude finks, masks and mimics. How can the model be revised for fink-, mask-, and mimic-exclusion?

Let's start with mimic-exclusion. Consider Johnston's original mimicking example:

*A gold chalice is not fragile but an angel has taken a dislike to it because its garishness borders on sacrilege and so has decided to shatter it when it is dropped. Even though the gold chalice would shatter when dropped, this does not make it fragile because while this dispositional conditional is not bare, i.e. the breaking when struck has a causal explanation, something extrinsic to the chalice is the cause of the breaking.*⁶⁶

Notice that Johnston is careful in specifying the *causal* features of the event. Mimicking cases are marked by certain causal features which make them unsuitable to count as R-events. A mimicking case is one in which an object is subjected to the D-test and nominally gives the D-response, but the D-test and D-response are not causally related in the appropriate way.

Disposition ascriptions do not ascribe mere regularities of test and response conditions; rather, they seem to ascribe something like a causal power. Mimicking cases are disposition tests in which the response is given not because of a power in the object, but because of irrelevant external factors (interfering angels, stress-activated mines, and so on).

Take another familiar mimicking case from Daniel Nolan⁶⁷: a certain Styrofoam cup would break if struck thanks to the Hater of Styrofoam. The Hater of Styrofoam (as you might have guessed) despises Styrofoam, and whenever he can locate it, he destroys it. He also has incredible hearing, so were some Styrofoam struck, he would hear it, leading him to destroy the Styrofoam. This is a mimicking case because

⁶⁶ Johnston 1992, p. 232.

⁶⁷ Lewis 1997, p. 153.

the causal relation between the cup, the stress, and the breaking is not appropriate to the disposition.

My proposal for dealing with mimics builds on David Lewis's response to the Hater of Styrofoam.

*There is a certain direct and standard process whereby fragile things most often (actually, nowadays, and hereabouts) break when struck, and the styrofoam dishes in the story are not at all disposed to undergo that process.*⁶⁸

The general lesson is that the meaning of dispositional predicates builds in *causal constraints*. Thus, for instance, the stress-activated mine case is excluded from 'fragile' R-events on the basis of the fact that it does not satisfy the causal constraints on 'fragile.' These causal constraints will generally conform to the interpretational constraints in the model: the mimic-free contents will be far more scrutable, interesting, and plausible to attribute in context than their mimic-inclusive counterparts, since the contents which include the mimics show much less underlying unity (in virtue of the test and response being connected by various disparate causal patterns). The causal constraints may vary between dispositional predicates or even for the same dispositional predicate between contexts – e.g. 'poisonous' may include only substances which cause death via a certain chemical mode of action at the cellular level (excluding corrosives) in the context of a toxicology lab but have more permissive causal constraints (including corrosives) in the context of worried parents debating where to put a bottle of drain cleaner.

So much for mimic-exclusion. Masks and finks also involve causal defects. The key consideration for the exclusion of finks and masks is that there are also constraints on *background conditions extrinsic to the object*. Physical stress on an object may fail to result in its breaking because the shockwaves are absorbed by Johnston packing or counteracted by a protective sorcerer. Exposure to poison may fail to result in death because an antidote has been taken. The reason these cases fail to 'count against' how disposed an object is has to do with the presence of *preventative*

⁶⁸ Lewis 1997, p. 153.

interference – there are factors present (the mask or fink) which ensure that the object won't give the disposition response, regardless of whether it is tested or not (and regardless of whether or not it has the disposition).⁶⁹ Like the case of mimics, the causal constraints on interference can be seen as a specific manifestation of general interpretational constraints: the fink- and mask-free dispositions are the more natural, scrutable, and plausible to ascribe in context than their fink- and mask-inclusive counterparts.

16. Two final revisions

Dispositional predicates can be used to ascribe very different properties to radically different categories of entities. For instance, we can make sense of property ascriptions using the word 'fragile' to a vase (disposed to break when physically stressed), a person (disposed to break down emotionally when psychologically stressed), an ecosystem (disposed to lose environmental equilibrium in response to environmental changes), a nation-state (disposed to give way to unrest or instability in response to political or social dynamics) and so on. Each of these different contents corresponds to different sorts of test and response conditions, but unlike more familiar cases (e.g. the pottery classroom example), this context-sensitivity seems to depend largely upon the *category* or *sort* of object the disposition is ascribed to. For this reason, let's call this dimension of context-sensitivity *sortal variance*.

How can a paradigmatic event involving an object of one sort be used in interpreting a disposition ascription to an object of a very different sort? For instance, how can we interpret a 'fragile(ecosystem)' ascription using a 'fragile(vase)' paradigmatic event of manifestation?

⁶⁹ This characterization is familiar from Fara's constraints on permissible exceptions: counting as a mask or fink case require the presence of features such that [1] given the features, the response will not be given, regardless of whether or not the object is tested, and [2] were mask or fink absent, the object would give the response when tested. As Fara also points out in connection with permissible exceptions, the *normality* of background conditions plays a role here as well: presumably the reason milk is not poisonous (rather than poisonous, though universally masked by lactase) is that the presence of lactase in the human digestive tract is assumed to be a 'normal' background condition rather than an interference.

In some cases, it is easy to model this resemblance with the idea of a paradigmatic event involving a test and response condition which are 'sortally restricted' by the sort of object involved in the paradigmatic event of manifestation, but which can be 'sortally detached' and 'reattached' to objects of different categories. For instance, we might suppose that an ascription of fragility to an *emotional* being ascribes a disposition to break *emotionally* in response to *emotional* stress resembles a paradigmatic manifestation of fragility by a *physical* object in which the object responds to *physical* stress by *physically* breaking. The two events do *not* resemble each other in a strict sense in that physical stress does not resemble emotional stress, physical breaking does not resemble emotional breakdown, and the causal process by which the physical test and response are connected does not resemble the causal process by which the emotional test leads to the emotional response. On the other hand, if we take the associated test and response (characterized by the paradigmatic event) to correspond to 'sortal independent' concepts of *stress* and *breaking* – whereby a *stress* of any sort resembles a *stress* of any other sort and a *breaking* of any sort resembles a *breaking* of any other sort – then this enables us to say that the emotional event *does* resemble the paradigm.

This kind of comparison can be made even when there is not a strict *lexical* category (*stress*, *breaking*) associated with a disposition. For instance, consider the case of 'poisonous.' Poisonous substances are those which (roughly) are disposed to cause harm or death when they get into the system of our body; but poisonous individuals seem to be those which are disposed to cause harm to a system when they enter the system. 'Poisonous,' whether applied to a substance disposed to cause death when ingested, an employee disposed to cause unrest in the workplace, or an athlete disposed to cause havoc in the locker room, seems to determine the sortal-independent test condition *coming into contact with* and the sortal-independent response *causing harm*, even though these various types of *coming into contact with* (being ingested, entering the workplace, joining a team) and *causing harm* (causing death, causing disruption in the

workplace, screwing up team chemistry) don't resemble each other independently of the sortal-independent test and response categories.

We should take 'appropriate similarity' to the paradigmatic event of manifestation to be this broad, sortal-independent resemblance rather than the narrower, sortal-dependent resemblance. Giving a precise characterization of these sortal-independent conceptual categories is very tricky, and I won't try to do that here. One area of literature where this broader sense of conceptual categorization is explored at length is discussions of *metaphorical* thought and language. Metaphorical use of adjectives (dispositional or not) seems to involve appealing to these broad conceptual resemblances across very different categories of things: beautiful people and beautiful theories are united by sharing certain sortal-independent features, though characterizing these features is no easy feat. I think that we have a good intuitive sense of the notion of taking a sortal-specific content – *physical* stress – and 'detaching' the concept from its specific category type to form a broad conceptual category – *stress* – which can then be reconfigured within a different sortal – *emotional* stress, *environmental* stress, *dialectical* stress.

My proposal is that *sortal variance* involves exactly this kind of narrow categorical 'detachment' and 'reattachment': we can recognize that a certain *narrow, type-specific category, emotional stress*, resembles a very different narrow, type-specific category, *physical stress*, in falling under the same *broad, type-independent category*. Thus, in order to handle the phenomena of sortal variance, we should modify our preliminary model by taking *paradigmatic resemblance in test-respects* and *paradigmatic resemblance in response-respects* to be resemblance in this *broader, type-independent* sense.

There is one final modification to make to the model, motivated by a worry that Manley and Wasserman's *proportional* approach to gradability is overly general and does not *quite* correspond to our actual grading intuitions. Take for instance the case of 'poisonous': I have suggested that our intuitions about 'more poisonous' tend to closely track *lethal dose*. Intuitively, if I were asked 'Which of these is more poisonous?', the

information I would need to make an informed judgment would primarily be information about lethal doses.⁷⁰ Lethal dose seems to be a good heuristic for proportion of manifestation –intuitively, a lower lethal dose will in general correlate with a higher proportion of manifestation – but if these two are in conflict, do our gradability intuitions really track proportion of manifestation? It seems to me that our actual use of dispositional predicates (and our dispositions to use them) will track gradability measures which we have easier access to. Consider two poisons, one of which consistently causes death at a lower lethal dose across different modes of administration, *except* for a single mode of administration, where some seemingly insignificant factor blocks the poison from causing death (e.g. poor absorption in the gastrointestinal tract results in higher lethal dose for ingestion, or poor membrane permeability results in higher lethal dose by injection at a certain site where the toxin will only become systematically available if it crosses that membrane). It seems that in such a case, one poison might cause death in a higher proportion of cases because of a big difference in a single mode of administration even though another is generally more lethal when delivered via every other mode of administration. Intuitions in this regard are sketchy; however, because it seems that there may be cases in which proportion of manifestation is not a perfect guide to grading, I’m going to suggest we speak of ‘a suitable *distribution* of R-events among T-events’ rather than ‘a suitable proportion of R-events among T-events.’ The revision is a minor one; it allows that in some cases, the suitable distribution may correspond to more salient and observable cues than proportion of manifestation.

17. The final model

Here is the final model. Recall that each dispositional predicate is associated by an interpreter with a paradigmatic event of manifestation e_P . The interpreter appeals to resemblance to e_P in determining the

⁷⁰ 21st century semantic fieldwork in action: if you Google ‘most poisonous substance,’ most top results cite lethal doses. Additionally, *time to cause death* seems to be a relevant consideration.

interpretational eligibility of a content, which I've individuated in terms of a pair of a set of T-events and R-events, $\langle T_\alpha, R_\alpha \rangle$. Now that I've introduced the notion of a 'suitable distribution' of R-events among T-events, we can more finely individuate interpretations using a quadruple: $\langle T_\alpha, R_\alpha, D, V_{\alpha,1} \rangle$, where $D(T_\alpha, R_\alpha)$ is the set of possible distributions of R_α -events among the T_α -events, and $V_\alpha: D(T_\alpha, R_\alpha) \rightarrow \{0,1\}$ is a valuation function on distributions of R_α -events among the T_α -events, with 1 corresponding to a suitable distribution and 0 corresponding to an unsuitable distribution. Contents with the same T- and R-event sets but different valuation functions correspond to interpretations with different suitable distributions and can be used to model the gradability of dispositional predicates.

Here's how the model works. Suppose an interpreter hears an utterance 'x is fragile' in context c_T . Candidate interpretations $\langle T_\alpha, R_\alpha, V_\alpha(d) \rangle$, $d \in D(T_\alpha, R_\alpha)$, $\langle T_\beta, R_\beta, V_\beta(d) \rangle$, $d \in D(T_\beta, R_\beta)$, ... , $\langle T_\omega, R_\omega, V_\omega(d) \rangle$, $d \in D(T_\omega, R_\omega)$ correspond to the different contents 'a V_α -admissible distribution of $T_\alpha(x)$ events are $R_\alpha(x)$ events', 'a V_β -admissible distribution of $T_\beta(x)$ events are $R_\beta(x)$ events' and so on.

The first dimension of evaluation is our broad, sortal-independent *T-resemblance* and *R-resemblance* of the paradigm. We can associate interpretations with *eligibility vectors* which correspond to the eligibility of the interpretation in each of our relevant dimensions. If T_α -events are characterized by *T-resembling* e_P (in the broad sense) and the R_α -events are characterized by *T-resembling* and *R-resembling* e_P , then the interpretation gets assigned the value 1 in the first of dimension of its eligibility vector, otherwise, the interpretation is assigned the value 0. Interpretations with the resemblance value 1 are called 'minimally lexically admissible.'

The second dimension of evaluation involves our *general interpretational constraints*. The more *plausible*, *scrutable*, and *interesting* the property corresponding to ' $R_\alpha(x)$ in a $V_{\alpha,1}$ -suitable distribution of $T_\alpha(x)$ events', the higher the interpretational value assigned to the eligibility vector. Since satisfaction of interpretational constraints,

unlike ‘sufficient’ broad categorical T- and R-resemblance of e_p , is a real-valued rather than binary, we might assign eligibility values to interpretations in this second dimension along the open interval $(0, 1)$ with 0 corresponding to a maximally unsound interpretation and 1 corresponding a maximally sound interpretation. Since for practical purposes, intuitions about satisfaction of interpretational constraints are not so incredibly robust as to provide real values, I’m going to assign only three values: ‘+’ for generally satisfying interpretational constraints, ‘*’ for moderately satisfying interpretational constraints, and ‘-’ for generally failing to satisfy interpretational constraints.

Thus, each interpretation $\langle T_\alpha, R_\alpha, V_\alpha(d) \rangle$, $d \in D(T_\alpha, R_\alpha)$ is now assigned an eligibility vector $\langle T$ - and R-resemblance value, interpretational value \rangle . The *best* interpretation of a dispositional predicate in context corresponds to the interpretation with the highest interpretational value among those with T- and R-resemblance values of 1.

18. The model at work

Here are some applications of the model to a number of different cases. I’ll first appeal to a case to show how the model works to exclude finks, masks, and mimics. I’ll then consider cases reflecting the various sorts of context-sensitivity dispositional predicates display.

a. Fink-, mask-, and mimic-exclusion

Consider the case of Fara’s cup in the possession of museum specialists. A new specialist joins the museum team and inquires about the cup: ‘Is it fragile?’ ‘Oh yes, it is fragile,’ one of the museum officials responds. Call this disposition ascription context c_1 . The specialist seems to speak the literal truth in saying that the cup is fragile; how are the officials to interpret this utterance?

Compare the following interpretations:

- [α]: $\langle T_\alpha, R_\alpha, V_{\alpha,1} \rangle$ where T_α includes all of the *non-masking* nomologically possible, no intrinsic change in which the cup is moderately physically stressed, and R_α includes all of the nomologically possible, no intrinsic change events in which the cup breaks in *non-mimicking* response to that stress
- [β]: $\langle T_\beta, R_\beta, V_{\beta,1} \rangle$ where T_β and R_β are as in [α], except T_β *additionally* includes *masking* cases in which there is preventative interference (e.g. the Johnston packing)
- [γ]: $\langle T_\gamma, R_\gamma, V_{\gamma,1} \rangle$ where T_γ and R_γ are as in [α], except that R_γ *additionally* includes mimicking cases
- [δ]: $\langle T_\delta, R_\delta, V_{\delta,1} \rangle$ where T_δ and R_δ are as in α except that they only include events which take place on Tuesdays, Wednesdays, and Thursdays
- [ε]: $\langle T_\varepsilon, R_\varepsilon, V_{\varepsilon,1} \rangle$ where T_ε includes events in which the cup has water poured into it and R_ε includes events in which the water doesn't leak out of the cup

and each valuation function determines similar proportions of manifestations as thresholds of admissibility. We want [α] to come out as the correct interpretation, and we want to find each of [β], [γ], [δ], [ε] lacking in some respects – [β] because of mask-inclusion, [γ] because of mimic-inclusion, [δ] for being having strange subsets of the correct T_α and R_α sets, and [ε] for ascribing a content which is not even minimally lexically admissible.

Consider first the *T-resemblance* and *R-resemblance* values: which of [α], [β], [γ], and [δ] have T- and R-sets which display characteristic test and response conditions which resemble (in the 'broad sortal-independent category' sense) our paradigmatic fragility manifestation e_P (which is, recall, an event in which a vase falls from a mantelpiece, hits the marble floor below, and shatters)?

T_α is a set of events characterized by *moderate physical stress*(cup) – these events resemble e_P in test-characterizing respects – i.e. in falling under the type-independent category *stress*. R_α is a set of events characterized by *moderate physical stress*(cup) and *breaking*(cup) – these events resemble e_P in both test- and response-characterizing respects – i.e.

in falling under the type-independent categories *stress* and *breaking*. Thus, $[\alpha]$'s eligibility vector gets a '1' for the first value – $E(\alpha)$: $\langle 1, ? \rangle$.

T_β is a set of events characterized by the same test condition as before, *moderate physical stress*(cup) – these events resemble e_P in test-characterizing respects. R_β is a set of events characterized by the same test and response as before, *moderate physical stress*(cup) and *breaking*(cup) – these events resemble e_P in both test- and response-characterizing respects. Thus, $[\beta]$'s eligibility vector gets a '1' for the first value – $E(\beta)$: $\langle 1, ? \rangle$.

T_γ is a set of events characterized by the same test condition, *moderate physical stress*(cup) – these events resemble e_P in test-characterizing respects. R_γ is a set of events characterized by the same test and response condition as before, *moderate physical stress*(cup) and *breaking*(cup) – these events resemble e_P in both test- and response-characterizing respects. Thus, $[\gamma]$'s eligibility vector gets a '1' for the first value – $E(\gamma)$: $\langle 1, ? \rangle$.

T_δ is a set of events characterized by the same test condition, *moderate physical stress*(cup) – these events resemble e_P in test-characterizing respects. R_δ is a set of events characterized by the same test and response condition as before, *moderate physical stress*(cup) and *breaking*(cup) – these events resemble e_P in both test- and response-characterizing respects. Thus, $[\delta]$'s eligibility vector gets a '1' for the first value – $E(\delta)$: $\langle 1, ? \rangle$.

T_ε is a set of events characterized by test condition *water poured into*(cup) – these events do resemble (even in the broad sense) e_P in test-characterizing respects, since they do not fall under the broad type-independent category *stress*. R_ε is a set of events characterized by *water poured into* (cup) and \neg *water leaks out of*(cup) – these events do not resemble (even in the broad sense) e_P in test-characterizing or response-characterizing respects – i.e. falling under the type-independent . Thus, $[\varepsilon]$'s eligibility vector gets a '0' for the first value – $E(\varepsilon)$: $\langle 0, ? \rangle$.

Thus, $[\varepsilon]$ is the only interpretation which does not satisfy the minimal lexical constraints on 'fragile' determined via broad T- and R-

resemblance of e_p . If $[\beta]$, $[\gamma]$, $[\delta]$ are bad interpretations of the utterance, then this will have to be reflected in the second dimension of evaluation, the *interpretation value*.

Let's examine the interpretation values of $[\alpha]$ - $[\varepsilon]$ comparatively. $[\alpha]$ interprets the utterance as ascribing the disposition to break when stressed in the absence of preventative causal interference via the familiar causal process by which physical objects break when stressed. $[\beta]$ interprets the utterance as ascribing the disposition to break when stressed, regardless of the presence of preventative causal interference, via the familiar causal process by which physical objects break when stressed. $[\gamma]$ interprets the utterance as ascribing the disposition to break when stressed, in the absence of preventative causal interference, though not via a specific familiar causal process by which physical objects break when stressed (since some causal chains in the R_γ -events are non-standard mimicking cases, e.g. in which the cup is stressed, breaks, but doesn't break in response to the stress but rather in response to the detonation of an attached stress-activated mine). $[\delta]$ interprets the utterance as ascribing the disposition to break when stressed on Tuesday, Wednesday, and Thursday, in the absence of preventative causal interference, and via the familiar causal process by which physical objects break when stressed. $[\varepsilon]$ interprets the utterance as ascribing the disposition not to leak when water is poured in.

Each of the interpretations except for $[\delta]$ ascribes a highly scrutable property – one which can be identified and inductively generalized. $[\delta]$ ascribes a less scrutable property because, although the disposition ascribed can be identified, there are not particularly useful generalizations about objects which have this disposition on Tuesday, Wednesday, and Thursday.

Each of the interpretations except for $[\delta]$, $[\gamma]$, and $[\varepsilon]$ ascribe a property which is interesting to the conversational participants. $[\delta]$ is not so interesting because the conversational participants don't care in this context about Tuesday, Wednesday, and Thursday-specific dispositions of the cup. $[\gamma]$ is less than maximally interesting because the conversational

participants don't care so much about the disposition to break via any causal process at all – they are more concerned with the disposition to break via the characteristic causal process by which physical stress on a fragile object causes the object the break. [ε] ascribes a property which might be of general interest to the conversational participants, but isn't of particular interest in this context in which the new specialist has asked about the *fragility* of the cup.

The interpretations vary significantly in their plausibility. [α] interprets the speaker as having attributed a property which the cup very plausibly has and which the speaker might plausibly ascribe to the cup. [ε] suffers somewhat in plausibility because it isn't clear why the speaker would attribute a disposition not to leak in response to a question about fragility. [β] is less plausible than [α] because [β] fails to impose the proper sort of constraints on background conditions in which the cup is stressed; why would the speaker attribute the property to break via a characteristic causal process *even* in cases of preventative interference rather than the property to break only in cases in which causal prevention is absent? [γ] suffers in the respect of plausibility in interpreting the speaker as ascribing a disposition not united by a salient, physical fragility-characteristic underlying causal process; why would the speaker attribute the broad disposition to break via some of a highly variegated bunch of causal processes? [γ] suffers in the respect of plausibility because it interprets the speaker as ascribing a disposition which, for no apparent reason, is specific to Tuesday, Wednesday, and Thursday.

In accordance with these various interpretational strengths and weaknesses, we might assign the following rough interpretational values: '+' to [α], and [β], though a 'higher' + to [α] than [β] and [ε]; '*' to [γ] and [ε]; and '-' to [δ].

Thus, we get the valuations

[α] <1, +>
 [β] <1, +> (though a lower '+' than [α])
 [γ] <1, *>
 [δ] <1, ->
 [ε] <0, *>

Thus, the model correctly predicts that $[\alpha]$ is the best interpretation, $[\beta]$ is a good but not maximally eligible interpretation, $[\gamma]$ is a not-very-good interpretation, $[\delta]$ is a bad interpretation, and $[\epsilon]$ is not even lexically permissible for ‘fragile’ in any context whatsoever.

We have seen in this section that if the T- and R-sets include events with the sorts of causal issues that arise in fink, mask and mimic cases, then this will result in some loss of general interpretational eligibility. The more a class of disposition manifestations is unified by causal constraints, the more eligible the corresponding interpretation will be. The more likely that a class of disposition manifestations requires certain background conditions, like the absence of certain kinds of interference, the more plausible interpretations will correspond to T-sets in which the object is tested in the absence of this interference.

b. Gradability

Consider the following three utterances. In the context of u_1 , two lab technicians are working with microbes, and one says ‘Be careful! That microbe is fragile!’ In the context of u_2 , I’m having dinner with my housemates and say, ‘Be careful with that cup; it’s fragile – I dropped it last week and the handle chipped off.’ In the context of u_3 , two aeronautical engineers are comparing hull materials for a spacecraft, and one says ‘We should use the platinum hull material rather than the titanium – the titanium is fragile and might not survive a rocky landing.’ Consider six candidate interpretations of each utterance:

$[\alpha_1]$: $\langle T_\alpha, R_\alpha, V_{\alpha,1} \rangle$ where T_α events involve miniscule to moderate stress, R_α events involve breaking in response to miniscule to moderate stress, and $V_{\alpha,1}$ assigns 1 to distributions in which a very high proportion of T_α events are R_α events and 0 to all others

$[\alpha_2]$: $\langle T_\alpha, R_\alpha, V_{\alpha,2} \rangle$ where $V_{\alpha,2}$ assigns 1 to distributions in which a moderate to high proportion of T_α events are R_α events, and to others

- [β 1]: $\langle T_\beta, R_\beta, V_{\beta,1} \rangle$ where T_β events involve moderate stress, R_β events involve breaking in response to moderate stress, and $V_{\beta,1}$ assigns 1 to distributions in which a very high proportion of T_β events are R_β events and 0 to all others
- [β 2]: $\langle T_\beta, R_\beta, V_{\beta,2} \rangle$ where $V_{\beta,2}$ assigns 1 to distributions in which a moderate to high proportion of T_β events are R_β events and 0 to all others
- [γ 1]: $\langle T_\gamma, R_\gamma, V_{\gamma,1} \rangle$ where T_γ events involve extreme stress, R_γ events involve breaking in response to extreme stress, and $V_{\gamma,1}$ assigns 1 to distributions in which a very high proportion of T_γ events are R_γ events and 0 to all others
- [γ 2]: $\langle T_\gamma, R_\gamma, V_{\gamma,2} \rangle$ where $V_{\gamma,2}$ assigns 1 to distributions in which a moderate to high proportion of T_γ events are R_γ events and 0 to all others

Each of these interpretations of each of the three utterances get a ‘+’ value for broad T- and R-resemblance of the paradigmatic event of manifestation e_P , since all of the test-characterizations fall under the broad, sortal-independent category of *stress* and all of the response-characterizations fall under the broad, sortal-independent category of *breaking*.

Let’s consider the interpretational values of each for utterance u_1 . Each content is scrutable both in material objects in general and in microbes. The α -interpretations are more *interesting* to the conversational participants since presumably the lab tech is not worried about her partner subjecting the microbes to extreme stress or even moderate stress – she’s worried that he might carelessly subject the microbes to very slight stresses which might nonetheless cause them to break. The γ -interpretations and [β 2] are highly implausible in light of these interests and the fact that it is utterly obvious that the microbes have the corresponding properties – assuming the cooperativeness of the conversational partner, we should not expect her to attribute properties to objects which all conversational participants already presuppose that they

have. Supposing that the microbes are evidently fragile enough to break in a high proportion of moderate stress cases and at least some miniscule stress cases, and $[\beta_1]$ is somewhat implausible for a similar reason. Suppose that it is presupposed by both conversational participants that the microbes break in at least a small proportion of miniscule stress cases; then $[\alpha_1]$ will be the more plausible ascription for the lab tech to make in light of the fact that $[\alpha_2]$ is less informative. If this is the case, then $[\alpha_1]$ will be the most eligible interpretation. On the other hand, if the interpreter thinks it is unlikely that the microbe is *so very* fragile as it is ascribed to be on $[\alpha_1]$ and thinks that it is unlikely his interlocutor would think otherwise, $[\alpha_2]$ and $[\beta_1]$ may be more eligible interpretations, as they would involve the speaker ascribing a degree of fragility that she would more plausibly take them to have.

Let's consider the interpretational values of each interpretation for utterance u_2 . Each content is scrutable for drinking cups. The γ -interpretations are less interesting in context, since it seems that my conversational goal in informing my housemates of a property of the cup is to prevent them from subjecting the cup to the sorts of stress which might break it. Since in ordinary household contexts speakers don't subject cups to extreme stresses, we don't seem to be interested in those stresses. All of the other interpretations seem interesting – it would certainly be relevant to our goals and interests to know that the cup breaks in response to cases of moderate stress or to moderate stress *and* miniscule stress. However, the α -interpretations are pretty implausible, since drinking cups *don't* usually break in even a moderate proportion of miniscule stress cases and I take my housemates to be attentive enough to be aware of this fact. The γ -interpretations are implausible for the opposite reason: of course the cup breaks in response to extreme stress, as almost all cups do, so why would my roommates take me, a generally cooperative conversational participant, to ascribe such a trivial property? Whether $[\beta_1]$ or $[\beta_2]$ is more plausible may again require a more careful explication of the situation: doubt about the likelihood of the cup breaking in the higher proportion of moderate stress cases might lead someone to take the $[\beta_1]$ to be the more eligible

interpretation, but if the interpreters already take for granted that the cup breaks in the moderate proportion, then $[\beta 2]$ would seem to be the more informative and plausible and thus eligible interpretation.

Let's consider the interpretational values of each interpretation for utterance u_3 . Each content is scrutable for aeronautic hull materials. All of the interpretations are pretty interesting in context, because it would be important for the engineer to be aware of any of those dispositions to break in response to some degree of stress. The α -interpretations and $[\beta 1]$ are implausible, however, because they appear to be obviously false. If the engineer is dubious of the $[\gamma 2]$ property since it doesn't appear to be a disposition that titanium hull has, then $[\gamma 1]$ and $[\beta 2]$ would seem to be the most plausible; if the engineer thinks it is obvious that the titanium hull has $[\beta 2]$ and $[\gamma 1]$, then $[\gamma 2]$ will be the more informative, and thus plausible, interpretation.

As we have seen, $[\alpha 2]$ and $[\beta 1]$ and $[\beta 2]$ and $[\gamma 1]$ seem to track each other to some degree in interpretational eligibility. This is because graded dispositions also generally have graded test conditions, and a higher threshold may be modeled by appealing to test conditions of a higher grade or to a higher proportion of R-events among the T-events.

c. Test- and response-variance

Let's consider two cases of test variance and one case of response variance. Our two pottery classroom utterances, u_1 and u_2 are a familiar case of *test-variance*: in u_1 , fragility tests correspond to *moderate physical stress* events, but in u_2 , fragility tests correspond to *heat stress* events. Let's consider these two interpretations of the two contexts:

$[\alpha]$: $\langle T_\alpha, R_\alpha, V_{\alpha,1} \rangle$ where T_α events involve a moderate physical stress and R_α events involve breaking in response to moderate physical stress

$[\beta]$: $\langle T_\beta, R_\beta, V_{\beta,1} \rangle$ where T_β events involve heat stress and R_β events involve breaking in response to heat stress

and both valuation functions assign ‘1’ to moderate or high proportions of R-events among the T-events and ‘0’ to less than moderate proportions.

Both interpretations are modeled by sets of events all of which T- and R-resemble the paradigmatic event in characterizing a test type falling under the sortal-independent category of *stress* (T_α , T_β , R_α , R_β) and a response falling under the sortal-independent category of *breaking* (R_α , R_β). Both interpretations take the speaker to ascribe properties which are scrutable for clay pots, are in general properties which we might be interested in knowing whether or not a certain clay pot has, and might plausibly be ascribed in some context. However, there is heightened conversational interest in u_1 in the disposition to break in response to moderate physical stress, since Sasha’s handling of M makes the possibility of it being moderately physically stressed very salient and Malia is visibly concerned with this possibility becoming an actuality. Thus, it is very plausible, given her conversational goal to prevent Sasha from destroying M , that she would inform her of the disposition to break in response to moderate physical stress. In u_2 , on the other hand, Mr. Potter is visibly concerned with the pots being adequately sturdy to go into the kiln; since his public conversational intention is inform Malia that her pot does not have a disposition to break in response to this highly salient stress, it is plausible that he would inform Malia that her pot does not have this disposition. Thus, $[\alpha]$ is the more plausible interpretation in the context of u_1 , given a pronounced interest in the $[\alpha]$ -associated test condition (in light of its heightened salience due to Sasha’s handling of M), and $[\beta]$ is the more plausible interpretation in the context of u_2 , given a pronounced interest in the $[\beta]$ -determined test (in light of its heightened salience due to Mr. Potter’s intentions in examining the students pots and the fact that all conversational parties know that M will soon be subjected to this test).

Similarly for the case of widget-gunk assembly lines: each assembly line manager will be clearly interested in the stress processes associated with his particular assembly line, and as such, it is most plausible to interpret each as ascribing the disposition to break in response to this particular stress.

Similarly also for the variance between cases in which a ‘fragility’-response requires *physical disintegration* versus *ceasing to function*. In the context of the teenage boys picking easily destructible cars to assault with baseball bats, there is a more pronounced interest in choosing cars which will physically disintegrate. Given that my concern for my laptop in a case in which my little nephew is running around recklessly has principally to do with the possibility of threats to its *functioning* rather than its physical integrity, the disposition to cease functioning in response to moderate physical stress is the more plausible property for me to attribute by saying ‘That laptop is fragile.’

d. Causal variance

Consider two ‘poisonous’-ascriptions in two different contexts. In the context of u_1 , a worried mother explains her decision to move the drain cleaner out from under the sink where her son can reach it by saying ‘Of course it shouldn’t be down there where he can reach it – the damn stuff is poisonous!’ In the context of u_2 , after reviewing toxicology results after the son’s tragic accident, a doctor says ‘It wasn’t a poison that your son ingested, but rather a corrosive – drain cleaner isn’t actually poisonous.’ Here are two candidate interpretations of u_1 and u_2 :

$[\alpha]$: $\langle T_\alpha, R_\alpha, V_{\alpha,1} \rangle$ where T_α includes events in which a substance is ingested, inhaled, or otherwise enters a human body, and R_α includes a broad class of events in which the substance causes death or harm via a number of different causal processes

$[\beta]$: $\langle T_\beta, R_\beta, V_{\beta,1} \rangle$ where $T_\beta = T_\alpha$ and R_β is a subset of R_α including only events in which the substance causes death via a familiar chemical mode of action at the cellular level

and both valuations assign ‘1’ to moderate or higher proportions of manifestation and ‘0’ to less than moderate proportions of manifestation.

Both interpretations bear the right T-resemblances and R-resemblances, since the test events of each fall under the broad, type-independent category of *entering a system* and the response events each

fall under the broad, type-independent category of *causing harm to the system*.

Both interpretations take the speaker to ascribe a property which is scrutable for substances; however, in the context of u_2 , the particular resources of the doctor (toxicology results) can be used to identify the property ascribed by $[\beta]$ but not $[\alpha]$. In the context of u_1 , there is an overwhelming interest in the broader class of responses corresponding to R_α . Thus, it is more plausible to take the doctor to be ascribing the $[\beta]$ property in u_2 and the mother to be ascribing the $[\alpha]$ -property in u_1 .

e. Sortal variance

Consider the following ‘fragility’ ascriptions:

u_1 : ‘That vase is fragile.’

u_2 : ‘Mike is quite a fragile guy,’ in a context in which it is presupposed by the conversational participants that Mike is physically hard to injure.

u_3 : ‘The Everglades are a fragile habitat in serious danger.’

u_4 : ‘Afghanistan is fragile right now – if we don’t act, it may fall into chaos.’

Each of these utterances seems to ascribe a different property: u_1 ascribes the disposition to break in response to physical stress, u_2 ascribes the disposition to break down emotionally in response to emotional stress, u_3 ascribes the disposition to lose equilibrium in response to environmental stresses, and u_4 ascribes the disposition to collapse politically in response to socio-political stresses. Each of these different contents corresponds to a property which is scrutable, interesting, and plausible to attribute for the sortal category of the object modified – vases are the sort of thing which can be physically fragile and for which we care about physical fragility, persons are the sort of thing which can be emotionally fragile and for which we care about emotional fragility, ecosystems are the sort of thing which can be environmentally fragile and for which we care about environmental fragility, and nation-states are the sort of thing which can be politically fragile, and for which we care about political fragility. On the other hand, these contents are generally inscrutable, uninteresting, and implausible to attribute to things of the different sortal categories: vases

don't have emotional properties, ecosystems don't have the relevant sociopolitical properties, persons don't have the relevant environmental properties, and so on. Each content determines a T-event class which is characterized by a test condition falling under the broad type-independent category of *stress* and an R-event class which is characterized by a response condition falling under the broad type-independent category of *breaking*. Yet in the specific context of ascribing 'fragility' to things of the different sortal types, the contents which interpretationally 'fit' *the sortal* as well as the larger context will be the most eligible.

In each of these cases in which the model is applied, the use of paradigmatic resemblance and general interpretational constraints seems to give a good explanation of why the specific contents which are intuitively more or less eligible.

19. Applications to lexical competence, vagueness, and conceptual unity

I argued in §4e that the move from the (SCA) to the complex conditional accounts lost in theoretical utility what it gained in prospects for extensional adequacy. I suggested that we didn't have to take this tradeoff: we should be able to give an account which is both extensionally adequate and adequately characterizes the context-sensitivity of dispositional predicates while still retaining plausible applications to lexical competence, vagueness, and conceptual unity. I suggested that the account I'd provide in the third part of the paper did just this – in this section I'll try to pay off this promissory note, describing applications to lexical competence, vagueness, and conceptual unity.

a. Lexical competence

What sort of knowledge and capacities does it take to be a competent user of a dispositional predicate? There was a very simple answer suggested by the (SCA) – it requires knowledge of disposition-associated test and response conditions as well as a general capacity for counterfactual reasoning. That is the entire conceptual apparatus

deployed in the analyses of the (SCA) – two rough-and-ready conditions and the Lewis conditional.

I have a lot of sympathy with this simple account. Its biggest strength is that the knowledge required for lexical competence is knowledge which competent users very plausibly have. On the other hand, it isn't clear what to say about cases – like 'brave' – in which there doesn't seem to be rough-and-ready associated test and response conditions.

The (QCA) inherits this simple account of lexical competence, with the added feature that lexical competence requires the ability to distinguish normal from abnormal background conditions.

The (PCA) and (QCA) would seem to require knowledge of massive conjunctions and disjunctions (respectively) of precise conditions. These precise conditions are necessary to weed out the mask, mimic, and fink counterexamples. I find it at least *prima facie* implausible that speakers possess this knowledge, although as Sanford Shieh has pointed out to me, it can't be that the *complexity* alone of the conjunctions and disjunctions that make it implausible, since we seem to have implicit knowledge of all sorts of unbelievably complex theories (e.g. grammar).

Here are a few reasons to be skeptical of the sort of knowledge required by the (PCA) and (QCA). I have argued (and the adherents of the complex conditional accounts seem to agree) that finks and masks are excluded from disposition tests and that competent speakers can intuitively identify and exclude these cases. However, I find it implausible that this exclusion is 'piecemeal' in the sense that the exclusion of a certain fink scenario f_1 is done on the basis of its violating one or some of the conjuncts or disjuncts, while exclusion of fink f_2 requires appealing to different conjuncts or disjuncts, exclusion of fink f_3 requires appealing to yet again different conjuncts or disjuncts, and so on. It seems to me that we should prefer to explain fink and mask exclusion by appealing to *general* constraints, and I have done just that in my account – fink and mask exclusion happens because of *causal constraints* which are governed by general *interpretational* rules. Additionally, it seems that speakers

could be (relatively) competent with dispositional predicates without possessing the knowledge of the big conjunctions and disjunctions – for instance, if they have knowledge of rough-and-ready associated conditions and impose the general causal constraints I’ve suggested are responsible for fink- and mask-exclusion. Finally, the (DCA) in particular requires knowledge that dispositional predicates correspond to large *disjunctions*, which involves a particular implausibility. Categorization by disjunction is grossly cognitively inefficient, since there need be no underlying unity among the disjuncts. I’ll return to this worry in connection with the underlying unity of disposition concepts, but it’s worth noting here that it seems at least somewhat implausible that our competence with dispositional predicates arises not because of any underlying unity of disposition concepts, but rather by brute assimilation of disjunctions.

On the other hand, my analysis seems to suggest a picture of lexical competence similar to that of the (SCA): competence consists in knowledge of associated sortal-independent test and response categories and a general competence for semantic interpretation. Within the model, paradigmatic events of manifestation are used to create these type-independent categories on the fly. In reality, competent speakers may develop these categories over time and associate them with dispositional predicates. Different type-independent categories may develop in different ways: for instance, in the case of ‘fragile,’ lexical concepts associated with ‘stressed’ and ‘breaks’ may play a significant role. In any event, the knowledge required for competence is knowledge that natural language speakers plausibly have, and the proposed mechanisms for correctly using and interpreting dispositional predicates are cognitively very efficient – lexical constraints imposed by the test and response categories, and general interpretational principles.

Since neither the habitual account nor the proportional account analyze dispositional predicates (only explicit disposition ascriptions), they are silent on the topic of lexical competence.

b. Vagueness

It is relatively uncontroversial that dispositional predicates are vague terms. Just as it would be ideal for an analysis to provide a mechanism for the context-sensitivity of dispositional predicates by identifying parallel context-sensitivity in the analysans, it would also be ideal for an analysis to identify sources of vagueness in the analysans.

Again the (SCA) has a relatively plausible explanation of the vagueness of dispositional predicates: it follows from the vagueness of the associated conditions and the subjunctive conditional. For instance, it certainly seems plausible that the vagueness of ‘fragile’ tracks the vagueness of ‘stressed’ and ‘breaks.’

But again we see that the (PCA) and (DCA) do not seem to offer plausible explanations, in this case because they use *precise* rather than vague conditions. Perhaps some vagueness in these accounts can be chalked up to the vagueness of the Lewis conditionals, but it seems quite implausible to suppose that this vagueness could account for the vagueness of dispositional predicates.

On my proposed analysis, there are two sources of the vagueness of dispositional predicates: the vagueness of associated type-independent test and response categories, and the vagueness of what is demanded by interpretational constraints. This seems to provide at least a *prima facie* advantage over the (PCA) and (DCA).

c. Conceptual unity

It seems natural to think that dispositions – fragility, solubility, poisonousness, and so on – form unified conceptual categories. On the (SCA), this conceptual unity is very simple – all fragile objects, for instance, are such that if they were stressed, they would break. On the (PCA), there is something very complicated in common – if fragile object were subjected to some very complex conjunction of precise conditions, they would give the response (also presumably specified by a conjunction of precise conditions). This is not particularly plausible. It doesn’t account for the fact that two objects might both be fragile but break in *different* fragility-tests. Nor does it seem to account for cross-sortal unity –

certainly a fragile nation-state doesn't give the specific fragility-response to a very precise fragility-test that a fragile vase does.

The (DCA) suffers the worst fate on the conceptual unity of dispositions: since dispositions are analyzed via a *disjunction* of conditionals, satisfaction of any one of which suffices for being disposed, there just *isn't* any unity among a disposition category. An object which satisfies the (DCA) in making one conditional true and another which makes a different conditional true have nothing in common in the analysis. This seems like a very unfortunate result, given the intuitive unity of our disposition concepts.

On my proposed analysis, the unity of disposition concepts is characterized by the unified type-independent test and response categories. All fragile objects would give the sort of response falling under the broad R-category to a disposition test falling under the broad T-category. This amounts essentially to the same basis for conceptual unity as the (SCA), except that in some cases the unified T- and R-categories may not be *lexical* conceptual categories.

Works Cited

Bird, Alexander. 'Dispositions and Antidotes.' *The Philosophical Quarterly* Vol. 48, No. 191: 227-234. 1998.

Bonevac, Daniel, Dever, Joshua, and Sosa, David. 'The Conditional Fallacy.' *The Philosophical Review* 115 (3): 273-316. 2006.

— 'The Counterexample Fallacy.' *Mind*, in press. Citations to draft at: webpace.utexas.edu/deverj/personal/papers/counterexample.pdf

Cappelen, Herman, and Lepore, Ernie. In insensitive Semantics: A Defense of Semantic Minimalism and Speech Act Pluralism. Blackwell, Oxford: 2005.

Carnap, Rudolph. 'Testability and Meaning.' *Philosophy of Science*, Vol. 3, No. 4 (October 1936), pp. 419-471.

Choi, Sungho. 'Dispositional Properties and Counterfactual Conditionals.' *Mind* Vol. 117 No. 468: 795-841. 2008.

Davidson, Donald. 'On the Very Idea of a Conceptual Scheme.' The Essential Davidson. Oxford University Press, Oxford: 2006. Originally printed in *Proceedings and Addresses of the American Philosophical Association*, 47: 5-20.

Fara, Michael. Dispositions and their Ascriptions. Princeton Ph.D. Dissertation, 2001.

— 'Dispositions and Habituals.' *Noûs* 39 (1): 43-82. 2005.

Gundersen, Lars Bo. 'In Defense of the Conditional Analysis of Dispositions.' *Synthese* 130: 389-411. 2002.

- Johnston, Mark. 'How to Speak of the Colors.' *Philosophical Studies* 68: 221-263. 1992.
- Lewis, David. Counterfactuals. Blackwell, Oxford: 1973.
- 'Counterfactual Dependence and Time's Arrow.' *Noûs* 13 (4): 455-476. 1979.
 - 'Finkish Dispositions.' *The Philosophical Quarterly* Vol. 47, No. 187: 143-158. 1997.
- Malzkorn, Wolfgang. 'Realism, Functionalism, and the Conditional Analysis of Dispositions.' *The Philosophical Quarterly*. Vol. 50, No. 201: 452-469. 2000.
- Manley, David and Wasserman, Ryan. 'A Gradable Approach to Dispositions.' *The Philosophical Quarterly* Vol. 57: 68-75. 2007.
- 'On Linking Dispositions and Conditionals.' *Mind* Vol. 117, No. 465: 59-84. 2008.
- Martin, C.B. 'Dispositions and Conditionals.' *The Philosophical Quarterly* Vol. 44, No. 174: 1-8. 1994.
- Mellor, D.H. 'The Semantics and Ontology of Dispositions.' *Mind* Vol. 109 No. 436: 757-780. 2000.
- Mumford, Stephen. Dispositions. OUP, Oxford: 1998.
- Szabó, Zoltán Gendler. 'Sensitivity Training.' *Mind & Language* Vol. 21, Issue 1: 31-38. 2006.