1 Introduction

Insofar as they exist at all, both dispositions and powers are properties of objects. Moreover, they are both properties which concern how objects are disposed to behave. But there would be no point in writing an Element about dispositions *and* powers if the two kinds of property were obviously identical. Nor would there be much point in writing a single text about both if they were not closely related. In this introductory section we'll clarify what makes them distinct and, if powers theorists have things right, what relates them. Sections 2 and 3 then take a deeper dive into the issues surrounding our philosophical understanding of each kind of property.

Let's consider dispositions first. The topic of dispositions is an important one because they would seem to be pervasive in *science* (see, e.g., Ellis and Lierse 1994; Harre and Madden 1975);[1] and dispositions have been used to define concepts in a variety of different philosophical domains. Examples include dispositional definitions of *beliefs* (e.g., Ryle 1949; Schwitzgebel 2002), *values* (e.g., Brower 1993; Smith et al. 1989), and *knowledge* (e.g., Constantin 2018; Yalowitz 2000). As we intend to use the term, a *disposition* is a property of being disposed in some specific way. Among these we count so-called canonical (or 'overt') dispositions, which are denoted by phrases of the form 'the disposition to M (when S)', where 'M' and 'S' are references to a behaviour ('manifestation') and (optionally) some kind of influence of that behaviour ('stimulus'), respectively. Examples include the disposition to break when struck and the disposition to sing the Macarena.

We also count among dispositions so-called conventional (or 'covert') dispositions.[2] These include properties denoted by terms ending '-ility', '-ivity', '-icity', such as fragility, conductivity, and elasticity. Properties like these clearly have associated with them some manifestation and (arguably, in these cases) a stimulus. Consequently, we might say that, at least in paradigm cases, a fragile object (one which has the property of fragility) is disposed to break when struck, and an elastic object (one which has the property of elasticity) is disposed to return to a particular length after an applied load is released. We also include among conventional dispositions properties denoted by terms without the tell-tale suffix, such as courageousness, brittleness, and locquatiousness. Again, we can tell these are also dispositions by virtue of their inferential connection with a manifestation (and maybe a stimulus[3]).

[1] This is notable regardless of whether or not the pervasiveness of dispositions in science constitutes good reason to believe in powers (cf. Williams 2011).

[2] The terms 'canonical disposition' and 'conventional disposition' are originally from Choi (2003); see also Choi (2008). The terms 'overt' and 'covert' are from Bird (2007b).

[3] For detailed discussion of whether dispositions should be associated with stimulus and manifestation or just manifestation conditions see Vetter (2015, chs. 2 and 3).

Although we talk of dispositions as properties, we commit ourselves here only to a very 'light-touch' realism about them. In English, at least, we have the clauses '. . . is disposed to M', '. . . tends to M', etc., and also '. . . is fragile', '. . . is elastic'. We can also form nominalisations of these clauses: 'the disposition to M', 'the tendency to M', 'fragility', 'elasticity'. Under the fair assumption that sentences constructed with these nominalisations require a referent for the respective nominal, then all speakers commit themselves to the existence of dispositions when they honestly employ such terms. Nevertheless, we do not think the commitment forces any deep metaphysical position on speakers. In Bird's (2016, 2018) phrasing, we understand dispositions as 'predicatory' properties, to which we are 'ontologically uncommitted', imbuing them with 'no metaphysical baggage'. Predicatory properties are linguistic entities that feature in *truths* (propositions or sentences), whereas ontic entities are what make the truths true; dispositions can be understood in the former sense. That is, we want to remain uncommitted over whether dispositional properties are ontic, or whether talk of such properties is just a *façon de parler*, not to be taken literally but rather as a convenient shorthand when what is really intended by the expression is the corresponding dispositional predication of an object, where, in classical logic at least, predicates do not carry existential commitment (see, e.g., Quine 1948). In this way we hope that talk of dispositional (predicatory) properties will be fairly uncontroversial.

As we have seen, dispositions are necessarily tied up in some way with a specific manifestation and (according to some) a stimulus for that manifestation. More specifically, we take it that the necessary connection is a conceptual one, in the sense that mastery of dispositional terms demands knowledge of the way in which dispositions are tied up with their specific manifestation (and stimulus). At the very least, we take this to include knowledge of what the associated manifestation of the disposition is, but we think it also requires a more general comprehension of how and when manifestation is likely to occur. Section 2 is all about trying to spell out what this dispositional behaviour is and whether it can be specified in such a way as to provide an analysis of dispositions, i.e., conceptually revealing necessary and sufficient conditions for when a disposition is possessed by an object.

Powers are also necessarily tied up with the possibility for manifestation (perhaps given a stimulus), but in a different way than dispositions. Powers, as we employ the notion, are hypothesised properties that metaphysically explain the manifestation behaviour witnessed in disposed objects. Hence, powers are posited as the (ontic) truthmakers for true disposition *predications*. We follow others in referring to powers as a kind of 'causal basis' for dispositions, to be contrasted with other potential causal bases, like regularities among categorical

properties. The point of saying this is not to suggest that the possession of powers causes the possession of dispositions, but rather that dispositional behaviour is (broadly speaking) causal behaviour, and, moreover, causal behaviour that warrants some metaphysical grounds. Powers' connection with manifestation behaviour is, therefore, a necessity to be established either through extended metaphysical or a posteriori reasoning.

There is a long-standing complaint with dispositions that their explanatory worth is limited. We can explain why the patient got sleepy by saying that the medicine they took is soporific. But it would be of little explanatory import to explain why the medicine makes those who ingest it feel drowsy by reference to it having the property of soporificity. After all, soporificity just is that dispositional property conceptually explicated in terms of the manifestation of drowsiness (after ingestion).[4] This suggests that the reason why ingesting the medicine induces drowsiness must make reference to something else. For example, we can say that the medicine is an opioid, meaning that it has certain empirically discoverable properties which are causally responsible for drowsiness. For some philosophers, these further empirically discoverable properties which individuate opioids may well be powers. If they are, they would be features of opioids which necessitate and explain the characteristics definitive of the dispositional behaviour witnessed by those who take them (note, that is not necessarily to say they necessitate drowsiness). Moreover, not only would these powers explain that behaviour, they would also explain why opioids are soporific, i.e., possess the disposition of soporificity.

It's the fact that powers are supposed to do this deeper explanatory work that makes them different in kind from dispositions. Unlike dispositions, powers (if they exist at all) are '*ontic'* properties. They are the sort of thing that populates the world, are metaphysically committing, and, for that reason, controversial. Section 3 discusses some of the central reasons philosophers have thought the commitment worth making alongside their explanation of dispositional behaviour, as well as some of the nuances behind how it is exactly that powers are necessarily and explanatorily related with dispositional behaviour.

The example of opioids and soporificity reveals a functionalist way of understanding the relationship between powers (if there are any) and dispositions which we think can be illuminating. According to this idea, dispositions are properties picked out by a causal role and powers (if they exist) are the realisers of those roles. So, for example, soporificity would be picked out by the causal process resulting in drowsiness. And if the reason opioids make one drowsy when

[4] For more on the discussion of dispositions' explanatory value, see Mumford (1998, 133–141).

ingested is that they possess certain powers necessarily and are explanatorily connected with the causal process of bringing about drowsiness, then these are the powers which realise the causal role implicated in opioids' soporificity.

There is much still to debate over the details of this functional understanding. One important question is whether or not dispositions are to be identified (a posteriori) with their realisers, so that soporificity *just is* a power which necessitates sleepiness upon ingestion. Such a view would nicely explain why dispositions seem so hard to conceptually analyse because it would render dispositions themselves real, occurrent, constituents of our ontology and so not apt to be analysed *away*. It would also confer upon dispositions the very explanatory causal role of their realisers without succumbing to problems of overdetermination. On the other hand, it seems to us that there may be some prospect of analysing dispositions, even if this analysis is not reductive in the sense that it cannot do without appeal to certain modal notions (see Section 2). Moreover, we sympathise with Prior et al. (1982) who argue that dispositions can (at least in principle) be realised by more than one causal occupant (e.g., it's not just the properties of opioids that can make us sleepy!). This would seem to demand that dispositions be instead identified with the second-order property of having some or other causal base which can perform a particular causal role.

Despite the importance of these kinds of issue, we leave their further discussion to another occasion (though see Hawthorne and Manley 2005; Mumford 1998; Prior 1985; Tugby 2022a, sec. 3.6 and 3.7). From hereon we keep discussion of dispositions and powers (Section 3) fairly distinct. This is reflected in how we have divided up writing this Element: one of us (Toby Friend) drafted this introduction and the discussion on dispositions (Section 2) while the other (Samuel Kimpton-Nye) drafted the discussion on powers (Section 3).[5]

2 Dispositions

2.1 Introduction

This section seeks a plausible analysis for dispositions. This is provided, we take it, by the provision of a schematic bi-conditional in which the left-hand side (the 'analysandum') is replaced by the attribution of a disposition and the right-hand side (the 'analysans') is replaced by non-trivial conditions true of exactly those things that satisfy the attribution, and knowledge of which would suffice for mastery of the dispositional concept. We follow tradition in aiming for a *single, unified* form of analysans rather than a plurality correspondent with different kinds of dispositional property.

[5] Inevitably, our individual philosophical preferences are not totally aligned and this is reflected in the emphasis placed for and against various views discussed in the longer sections.

The history of analysis of dispositions has its source in Carnap (1936). Carnap wanted to show how dispositions (more correctly, *dispositional expressions*) could be interpreted free of what he saw as unempirical terminology, such as modal operators and subjunctives. Ideally, a disposition like *solubility* was to be analysed by a relationship between a test condition (*stimulus*) and a resultant behaviour (*manifestation*) expressed using the minimal resources of first-order quantified logic. As we'll remark on further in the next subsection, philosophers (including Carnap) were quick to point out the difficulties of doing so and nowadays it is generally acknowledged that analyses of dispositions – insofar as they are possible – must employ some modal terms (see Bird 2012; McKitrick 2018, ch. 2; Mumford 1998, ch. 3; Schrenk 2017, ch. 2 for more on the history). The question, then, is *which modal terms*, and *in what way* must they be employed*?*

We begin by describing three platitudes about dispositions that justify the infamous 'Simple Conditional Analysis' (SCA) as our initial foil. We then introduce the heuristic of structural equations modelling. In the three subsections which follow (2.2, 2.3, and 2.4), we then use this heuristic to present problem cases for the preceding strategy for analysis before presenting another strategy in response. We go through various iterations (nine problem cases and eight strategies) until subsection 2.5, in which we sketch our own strategy for analysis that makes explicit reference to structural equations.

2.1.1 Three Platitudes

We begin by voicing some platitudes and points of consensus about dispositions.

First platitude: a disposition towards M is a 'directedness towards' some kind of paradigmatic behaviour conceptually associated with the disposition (Martin 2007; Molnar 2003; Tugby 2013). The term 'manifestation' is widely used for this behaviour, even though the behaviour might not be directly observable to the senses. The crucial idea behind 'directedness' is that being disposed towards a manifestation M should not entail actually doing M. An object disposed to behave in way M may not ever behave in way M.[6]

Second platitude: a disposition's manifestation can be triggered under specific conditions. For example, the sonority of a bell is manifested under striking with a hard object; malleability is manifested under pressure, etc. The term 'stimulus' is widely used to reference these conditions. Many, however, prefer the term 'manifestation partner', which is more conducive to understanding manifestation as an effect both of the disposition and whatever further conditions are

[6] However, this aspect of the platitude has been brought into question (Friend 2021).

needed (Heil 2010; Martin 2007; Mumford and Anjum 2011). Moreover, calling one property the 'stimulus' and the other the 'disposition' can often seem arbitrary, e.g., with the manifestation of heat exchange between a hot and a cold object. Nevertheless, 'manifestation partner' may go too far the other way, prohibiting any understanding of what triggers a manifestation being other than another property. Triggers could potentially be totality facts, absences, or background conditions.[7] Importantly, we won't assume that every disposition has a specific stimulus (see subsection 2.3.3, *Problem #6*). Some dispositions can be stimulated under a variety of conditions (e.g., breakability), other dispositions appear to have no stimulus requirements at all (e.g., loquaciousness). A successful analysis should accommodate all these sorts of disposition.

Third platitude: the directedness of disposed objects towards manifestation is modal. We've remarked on the widely appreciated failure of excising modality from dispositional expressions' interpretation. The problem is that the dependence of manifestation on a stimulus seems obviously *conditional*. Yet the (non-modal) material conditional that manifestation occurs *if* the object is stimulated is satisfied by anything that is never stimulated (Carnap 1936). So, if dispositions were analysed by this conditional, an iron pot which never undergoes any applied stress would be falsely deemed just as fragile as a delicate porcelain vase. Instead, as many have remarked, what is needed is to link the stimulus and manifestation by a *counterfactual conditional*: manifestation *would occur were* the object stimulated. Counterfactual conditionals don't give rise to the same issue, since even if an object is never stimulated it typically *could be*.

Prior (1985, 5) called the connection between counterfactual conditionals and dispositions 'pre-theoretic common ground'; earlier, Quine (1974, 9) admitted (alongside Ryle 1949 and Storer 1951) that 'there is no denying that in its bumbling way, this intensional conditional somehow conveys the force of the dispositional idiom'; and later, Mumford (1998, 87) concurred with Martin (1994) that dispositions must be connected 'somehow' with conditionals. These remarks suggest that we could take the employment of counterfactual conditional analysans as a further platitude of dispositions. But there are prominent dissenting voices. Some eschew any attempt to characterise the modality of dispositionality in non–sui generis terms (Anjum and Mumford 2018; Martin 1994; Mumford and Anjum 2011). But even assuming these philosophers are wrong, others maintain that the dispositions' modal character is one of the *possibility* of manifestation rather than

[7] The term 'stimulus' also suggests that the associated conditions for manifestation must be *causes* of it. We think this is often a fair assumption (Handfield 2010; McKitrick 2010). Nevertheless, some dispositions may not be so easily thought of in this way, such as those in quantum mechanical contexts or Lagrangian mechanics (Katzav 2004; Nolan 2015; Smart and Thébault 2015). Space precludes us from engaging further with these cases.

manifestation's dependence on stimulus (Aimar 2019; Vetter 2015). We will explore these alternatives in subsections 2.3 and 2.4. Nevertheless, whichever option we pursue, the modality of dispositions' directedness is no longer in doubt.

With these platitudes in mind, we can now give a more precise description of what an analysis of dispositions must involve: it must spell out the 'directedness' platitude in terms of the modal implication of any disposed object's manifestation, including saying how, if at all, stimuli are involved. To that end we commence our search as many others have previously, by scrutinising the 'Simple Conditional Analysis' (where `$\Box\!\!\rightarrow$' is the counterfactual conditional).

> SCA. For all x, Dx if and only if Sx $\Box\!\!\rightarrow$ Mx.

Our go-to example, again unoriginal, is the analysis of *fragility*. Fragility is a disposition directed towards breaking, which we assume for the time being has the stimulus of (relatively low) applied stress. In the form of SCA, the analysis of fragility looks as follows:

> FRAGILE. For all x, x is fragile if and only if were x to undergo (relatively low) stress x would break.

SCA (or properly speaking, its instances) is plagued by well-known counterexamples. We will go through the nine we have identified in the literature and discuss various strategies for avoiding them. However, before diving into the problem cases we introduce a heuristic for representation that we will make considerable use of this throughout this Element.

2.1.2 A Tool for Representation

Our demonstration of many of the problem cases will be atypical in that we make use of structural equations modelling. This might make us seem guilty of using a sledgehammer to crack a nut, since the formalism is rather more involved than anything employed in the current debate. Nevertheless, we ask the reader to persevere for two reasons. First, we think that once understood, structural equations models provide one of the clearest resources for identifying the nuances of each problem case. Second, structural equations will be of central importance to developing our novel strategy for analysis provided in subsection 2.5, and a familiarity with them will greatly facilitate its introduction.

A structural equation has the form $B(x) \Leftarrow f(A_1(x), \ldots, A_n(x))$ and expresses an asymmetric relationship of *numerical counterfactual dependency* of the left-hand variable property B of an object x on some function f of right-hand variable properties $A_1, \ldots, A_n$ of x (for conciseness we often omit the object variable). A structural equation therefore encodes lots of counterfactual information about

how x would behave. Indeed, structural equations provide a counterfactual for every combination of values assigned to the right-hand variables (within some permitted range): if it *were* that $A_1(x) = a_1, \ldots, A_n(x) = a_n$ (i.e., each variable on the right-hand side were to take some specific distribution of determinate values) then *it would be that* $B(x) = f(a_1, \ldots, a_n) = b$.

Structural equations are bread and butter for contemporary causal analysis in science and philosophy of causation (Pearl 2000; Hitchcock 2001; Woodward 2003; Halpern and Pearl 2005). Causal results and hypotheses are typically expressed in terms of structural equations *models* (SEMs), an ordered pair $(\boldsymbol{V}, \boldsymbol{E})$ of variable set $\boldsymbol{V}$ and structural equations set $\boldsymbol{E}$ such that every variable in $\boldsymbol{V}$ is either on the left-hand side of at most one equation in $\boldsymbol{E}$ (and so dependent on other variables in $\boldsymbol{V}$) or else is 'exogenous' (having its value determined by factors outside of the model). Associated with any SEM is a causal graph where the variables in $\boldsymbol{V}$ are nodes and directed edges (arrows) lead from one variable A to another B just in case A features in the right-hand side of a structural equation in which B is the left-hand variable.

Our aim is to use SEMs to describe the causal relationships relevant to dispositional behaviour. These models will therefore include 'stimulus variables', 'manifestation variables', and 'disposition variables', which take values ranging over whether (and to what degree) an object is stimulated, manifests, and has some disposition, respectively.

Let's consider the SEM involving a fragile object x described in Table 1.

Table 1 displays the model's variables ($FR(x), ST(x)$ and $BR(x)$) and how to interpret their possible values. For example, if the variable $FR(x) = 1$ this indicates that x is fragile, and if $FR(x) = 0$, x is *not* fragile. Table 1 also displays whether the variables have a structural equation or not. In this model, only one variable is not exogenous $(BR(x))$ and so is the only variable with a structural equation $(BR(x) \Leftarrow FR(x) \times ST(x))$. The model's equation tells us that the values of $BR(x)$ are determined by $FR(x) \times ST(x)$. Moreover, it tells us that the determination is robust under counterfactual variations of the right-hand variable's values: for any combination of values for $FR(x)$ and $ST(x)$, if those variables were to take those values, $BR(x)$ would take the value given by the equation.

Since the one structural equation for the present model reveals $BR(x)$'s dependence on $FR(x)$ and $ST(x)$, the causal graph for this model will have two directed edges indicating causal influence of $FR(x)$ and $ST(x)$ on $BR(x)$, as displayed in Figure 1.

If we take this SEM to characterise causal relationships any object x whatsoever can be involved in, then it predicts FRAGILE. For it entails that if x were fragile $(FR(x) = 1)$ and were stimulated $(ST(x) = 1)$ then it would break $(BR(x) = FR(x) \times ST(x) = 1)$, whereas if x were not fragile

Table 1 Details for a causal model for fragility

Variables			
Symbol	**Possible values**	**Interpretation**	**Structural equations**
FR	1	*x* is fragile	(Exogenous)
	0	*x* is not fragile	
ST	1	*x* undergoes stress	(Exogenous)
	0	*x* does not undergo stress	
BR	1	*x* breaks	$BR \Leftarrow FR \times ST$
	0	*x* does not break	

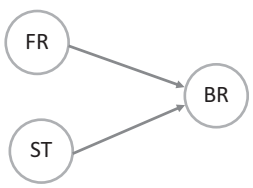

Figure 1 Causal graph for a simple causal model for fragility

($FR(x) = 0$) and were stimulated ($ST(x) = 1$) then it would not break ($BR(x) = FR(x) \times ST(x) = 0$). Assuming possible worlds are strongly centred on the actual world (so that the actual world is the closest possible world in which any true proposition is true) then FRAGILE is entailed. However, the SEM predicts further counterfactuals not directly relevant to the truth of FRAGILE. For instance, it predicts that if something were fragile ($FR(x) = 1$) but did not undergo stress ($ST(x) = 0$) then it wouldn't break ($BR(x) = FR(x) \times ST(x) = 0$). Although this kind of counterfactual is not part of traditional dispositions' analyses, we take it to be plausible in many contexts. In subsection 2.5 we will argue that, in fact, it should be a feature of fragility's analysis.

Given how widely dispositions are thought to be causally implicating we find it surprising that SEMs are yet to be invoked in discussion of their analysis. As we aim to show, SEMs provide a fertile heuristic for displaying many of dispositions' causal features. But it's important not to get carried away. Structural equations models do not add any metaphysical assumptions about the relationships among those included variables than are already implied by the counterfactuals encoded in their equations. Instead, SEMs' value comes from the fact they encode far more counterfactual information than is available from any single conditional. It is this feature which makes them instrumental for describing problem cases, and also for providing the basis of a new strategy for analysis (see Section 2.5).

2.2 One-Conditional Analyses

Here we look at four problems which have brought SCA into question, and at strategies for response which keep within the constraints of a one-conditional analysis, i.e., analyses that employ a single counterfactual conditional in the analysans.

2.2.1 Problem #1: Masks

The disposition of an appropriately stimulated object is 'masked' when it fails to manifest due to interference.[8] A classic example is when a fragile vase fails to break when subject to stress because it is packed in bubble wrap (Johnston 1992). Another example is when an antidote is taken to counteract the disposition of ingested poison to cause harm (Bird 1998). This kind of causal interference is easily captured in an SEM, as described in Table 2. The associated causal graph is displayed in Figure 2.

The counterfactuals entailed by this SEM are more complex than instances of SCA allow. A typical SCA for *poisonousness* might be the following:

> For all x, x is poisonous if and only if were x ingested harm would occur.

Table 2 Details for a causal model for masking poisonousness

Variables			
Symbol	**Possible values**	**Interpretation**	**Structural equations**
P	1	x is poisonous	(Exogenous)
	0	x is not poisonous	
I	1	x is ingested by agent	(Exogenous)
	0	x is not ingested by agent	
A	1	x is accompanied by antidote	(Exogenous)
	0	x is not accompanied by antidote	
H	1	Agent comes to harm	$H \Leftarrow (P \times I) \times (1 - A)$
	0	Agent does not come to harm	

[8] We do not here distinguish masks from 'antidotes' (or `interferers'), where the latter but not the former act after the stimulus has taken place (cf. Paoletti 2021).