Context-dependence of dispositional ascriptions

Tomasz Wysocki (tomwysocki@causation.science), Louisa Reins & Michael R. Waldmann
Department of Psychology, University of Göttingen, Germany

Abstract

Dispositional concepts like *fragile* or *vulnerable* are crucial for prediction and action planning, yet no established cognitive theory explains their mental representation. We propose a *relevance theory of dispositions*, positing that the function of dispositional concepts is to encode an object's disposition-characteristic behavior across relevant situations. We formalize this theory as an idealized cognitive mechanism, in which dispositional strength is calculated as a sum of response intensities across situations weighted by relevance. We test this theory in three experiments examining judgments of vulnerability, fragility, flammability, and solubility. Their results show that a disposition is judged stronger when it would manifest in actual rather than merely possible situations (Experiment 1), in situations an agent intends to realize (Experiment 2), and in high-stakes situations (Experiment 3).

Introduction

The Queen offered Snow White the poisonous apple, for she knew very well what would happen if Snow White bit it. And indeed, what happened next happened because the apple was poisonous. Despite the role of such dispositional concepts in everyday life, no research has yet proposed a theory of disposition ascriptions. Here, we formulate such a theory.

The theory explains disposition ascriptions in terms of the average intensity of disposition-characteristic responses across relevant situations. It rests on the premise that dispositional terms are used to encode efficiently an object's disposition-characteristic behavior across relevant circumstances. The theory predicts an unprecedented degree of context-dependence in dispositional ascriptions. Here, we empirically test three of these predictions: that it matters more for dispositional ascriptions what happens in a situation that is actual vs. merely possible (Experiment 1), intended vs. immaterial (Experiment 2), and high-stakes vs. low-stakes (Experiment 3).

Counterfactual and proportion theories

"A thing is full of threats and promises," says Goodman (1954, p. 40) of dispositions. Dispositions are ubiquitous yet baffling, and what makes them so is their modal status. We ascribe them to objects all the time, though they are only potentially observable: Deadly katanas may never kill anyone, safely stored in a museum. We use dispositions in action planning, as witnessed by 'delicious donuts!' and 'slippery steps!' signs designed to influence our actions.

A disposition always comes with characteristic *responses* (e.g., cracking, breaking, shattering for fragility, dissolving for solubility) and *stimuli* (hitting, dropping, crushing; submerging in or exposing to water).

Philosophy has developed two families of theories of dispositions. *Counterfactual analyses* intend to explain categorical judgments and account for them in terms of counterfactuals. A statuette of salt is soluble because it would dissolve if submerged in conditions that don't interfere with the process of dissolving (Bird, 1998; Choi, 2008; Gundersen, 2002; Lewis, 1997). *Proportion theories* also intend to explain graded judgments: the strength of the disposition (e.g., how fragile an object is) is given by the proportion of possible situations where the object responds (e.g., breaks) to all situations where the object is present (Manley & Wasserman, 2007, 2008; Vetter, 2014, 2015). Both families in their actual form don't predict context-dependence in the sense used here: the proportion of situations where an object breaks or whether it satisfies the counterfactual "if it were hit in such-and-such circumstances, it would break" doesn't depend on which situations are actual, intended, or high-stakes.

Relevance theory of dispositions

The theory we propose is motivated by the cognitive role of disposition ascriptions, which we take to be: *dispositional concepts are used to encode efficiently an object's disposition-characteristic behavior across relevant situations*. There are a priori reasons to find this principle plausible, as one would expect language to facilitate communicating the threats and promises that things are full of. Consider two Starship cases:

[1] *Actual*. A starship is roaming through hostile space. If it were attacked by torpedoes, bombs, lasers, or phasers, it would shatter. That's why the captain is considering adjusting the course so that the starship takes its entire journey through a nearby nebula. The nebula gases dampen any explosion, which means if the starship were attacked inside the nebula by torpedoes, bombs, lasers, or phasers, it wouldn't even scratch or dent. However, the radiation from the nebula would make some crew members sick.
The captain weighs the pros and cons of entering the nebula and decides to take the original course that omits the nebula.

The other case is the same but for the captain's decision:

[2] *Merely possible*. The captain weighs the pros and cons of entering the nebula and decides to adjust the course and travel through the nebula.

We predict that people will find the starship more vulnerable

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

in [1] than in [2] because the actual situation is more relevant than the merely possible one, and only in [1] would the starship be destroyed if it were attacked in the actual situation.

The *relevance* of situations takes many forms, and we can't offer an exhaustive theory of what counts as relevant. What we do offer is a rough characterization: relevant are these situations in which the object's disposition-characteristic behavior may matter for actions; threats and promises are action-guiding. We propose the following rules of thumb.

First, *actual* situations, per the examples above, are *ceteris paribus* more relevant than merely possible ones, for we live and act in actual situations. Second, the situations of *intended use* are more relevant than unintended ones, as intentions matter for action-planning. Third, *stakes* matter, for what happens in high-stakes situations, where the object's response can come with dire consequences, is more relevant than what happens in low-stakes ones, where the response's consequences are of less concern. Fourth, situations where the *disposition-specific stimulus* happens are more relevant than the ones where it doesn't. The role of dispositions is to convey a threat or a promise, and these typically materialize only when the stimulus is present. Some of these considerations have already been discussed in the literature: actuality (Baron & Hershey, 1988; Lucas & Kemp, 2015; Quillien & Lucas, 2024) and stakes (Lichtenstein et al., 1978; Lieder et al., 2018; Sunstein & Zeckhauser, 2011) have been reported to play similar roles in other psychological phenomena.

With these considerations in mind, we developed the *relevance theory of dispositions*. The theory evaluates the extent to which a certain object has a certain disposition–the *strength* of the disposition. Take a family of situations that contain the object and factors causally relevant for whether the object responds, and take the disposition-characteristic responses. Every situation from the family of situations is represented by a deterministic causal model, which describes what happens in that situation. In each situation, one of the responses (including the non-response) happens, and each of these is characterized by a specific intensity: e.g., shattering is more intense than breaking, which is more intense than staying intact. Lastly, a situation comes with a weight representing its relevance. The strength of the disposition is the weighted sum of the response intensities over all situations, i.e., models.

Intensity ordering can be recovered from strength judgments: if two cases differ only by response intensity in exactly one situation, strength order reflects intensity order; and if the same response happens in all relevant situations, intensity reduces to strength. Some may think that for some dispositions, intensity admits of little variation; things may just dissolve or not dissolve, for instance, with no intermediate levels. However, our research on the interaction of intensities and probabilities suggests that people view the intensity of at least the four dispositions tested here as varying on a continuum, as, e.g., people judge tablets that dissolve in water only partially as less soluble than the ones that do so fully (Klimenta, 2025).

Formalized, the cognitive model reads as follows. Strength s of disposition D ascribed to object o in context C depends on the relevance ω of each model (situation) $\mathfrak{m}$ and the intensity ι of responses:

$$s = s_{o,D,C} = \sum_{\mathfrak{m} \in \mathrm{S}} \omega_{\mathfrak{m}} \cdot \iota(\vec{\sigma}), \tag{1}$$

where S is the family of situations, i.e., models, and $\vec{\sigma}$ is the sole solution to $\mathfrak{m}$, i.e., the specification of the events that happen in the situation. We assume that intensity and relevance are normalized, and relevance satisfies the Kolmogorov axioms.[1] We further posit that relevance is the *sampling propensity* (see, e.g., Icard 2016) with which the mind entertains different situations to arrive at the strength score. Therefore, relevance could be measured using methods for evaluating sampling propensity, although we don't attempt this in the present study.

In our formalism, variables are divided into response, stimulus, background variables, and the causal basis. A model always includes response variables, which can have multiple values denoting the many possible responses (cracking, breaking, shattering) and must have one value denoting the non-response. It also always includes causal basis variables, which represent the features of the object that cause the responses (e.g., the material, but not the color for fragility). The causal basis may also be a placeholder, as the person making the ascriptions often doesn't know what features of the object cause the response. A model typically includes stimulus variables, which can have multiple values denoting the many possible stimuli and must have one value denoting the lack of any stimulus. The values can denote both standard stimuli for the target disposition (e.g., dropping, hitting, smashing for fragility) as well as non-standard stimuli implied by the context. Finally, a model typically includes background variables—all variables (a person takes to be) causally relevant for the response that don't count as the basis or stimulus.

Dependencies between the variables are expressed with structural equations. Consider the model for case 1. It has four binary variables (Figure 1). Stimulus S: 1_S (0_S) if the ship is (not) attacked. Response R: 1_R (0_R) if the ship is (not) destroyed. Causal basis B: 1_B for the starship having the composition it does, 0_B for having a composition that would withstand the attack. Background K: 1_K (0_K) for taking the straight course (the course through the nebula). A variable is endogenous if its value is determined by the values of other variables; otherwise, it's exogenous. In the situation described by this scenario, the starship is destroyed only if it's attacked while having the composition it does and while taking the straight course, a dependence encoded by the structural equation (Figure 1):

$$R \leftarrow S \wedge B \wedge K, \tag{2}$$

R is endogenous as its value depends on S's, B's, and K's; the other variables are exogenous.

[1] For simplicity, intensity ι is applied to the entire solution $\vec{\sigma}$, since a model in principle can have multiple response variables: a starship can be vulnerable in virtue of its hull breaking, life-support failing, and propulsion shutting down. Really, though, ι is a function of response variables $\vec{R}$ alone, $\iota(\vec{\sigma}) = \iota(\vec{\sigma}[\vec{R}])$.

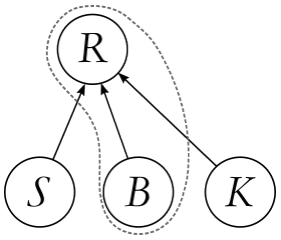

Figure 1: The causal model of the starship case.

Now, the scenario invites the reader to entertain four different combinations of exogenous values that entail four different situations. First, the starship is attacked (1_S) while taking the original course (1_K), which implies the ship is destroyed (1_R). We will denote this model by $\mathfrak{M}_{S=1\wedge K=1}$; its solution is the first row in Table 1. Second, the starship isn't attacked (0_S) while taking the original course (1_K), which implies that the ship stays undamaged (0_R): $\mathfrak{M}_{S=0\wedge K=1}$. Third, the starship is attacked (1_S) while taking the course through the nebula (0_K), which implies that the ship stays undamaged (0_R): $\mathfrak{M}_{S=1\wedge K=0}$. Finally, it isn't attacked (0_S) while taking the course through the nebula (0_K), and the ship stays undamaged (0_R): $\mathfrak{M}_{S=0\wedge K=0}$.

The next step is assigning *relevance* to the situations. We will apply the rules listed above. Being attacked comes with more weight because that is the standard stimulus for vulnerability: $\omega_{S=1\wedge K=k}>\omega_{S=0\wedge K=k}$ for k=0, 1. Meanwhile, taking the original course comes with more weight because it's the actual state: $\omega_{S=s\wedge K=1}>\omega_{S=s\wedge K=0}$ for s=0, 1. These reasons, of course, don't imply precise weights, and we will use values that express their ordinal rank (column $\omega^{\text{act}}_{\mathfrak{M}}$ below).

Overall, the family of situations, their weights, and the events unfolding in each are given in Table 1, where we set the intensity of 'being destroyed' to the maximal value 1 and of 'remaining intact' to the minimal value 0.

Table 1: Situations for the starship case

$\mathfrak{M}$	S	B	K	R	$\omega^{\text{act}}_{\mathfrak{M}}$	$\iota(\sigma)$	$\omega^{\text{poss}}_{\mathfrak{M}}$
$\mathfrak{M}_{S=1\wedge K=1}$	1	1	1	1	.4	1	.3
$\mathfrak{M}_{S=0\wedge K=1}$	0	1	1	0	.2	0	.1
$\mathfrak{M}_{S=1\wedge K=0}$	1	1	0	0	.3	0	.4
$\mathfrak{M}_{S=0\wedge K=0}$	0	1	0	0	.1	0	.2

This representation allows us to apply (1) to calculate how vulnerable the ship is: $s = \sum_{\mathfrak{m}} \omega_{\mathfrak{m}} \cdot \iota(\sigma) = .4 \cdot 1 = .4$. Now, consider the sister case [2], where the captain makes the opposite decision. Equation 2 stays the same, but the weights differ. Because the course through the nebula is actual, the solutions for $\mathfrak{M}_{S=s\wedge K=k}$ get the weights from solutions for $\mathfrak{M}_{S=s\wedge K=1-k}$, as per column $\omega^{\text{poss}}_{\mathfrak{M}}$. This representation, in turn, entails a lower strength: $s = \sum_{\mathfrak{m}} \omega_{\mathfrak{m}} \cdot \iota(\sigma) = .3 \cdot 1 = .3$.

In this project, we don't attempt to estimate the weights of individual situations entertained by the participants, but test ordinal strength differences between different conditions based on our assumptions about relevance. Also, notice that we expect people to differ with respect to how much pragmatic factors influence relevance. In particular, we expect that some people—like the authors of some metaphysical theories—won't be influenced by these factors almost at all. What matters for us is whether our manipulations change the overall distributions of participants' responses.

The most novel feature of the theory is the ubiquitous context-dependence of graded dispositional ascriptions. As it predicts many more contextual effects than its metaphysical counterparts, in what follows we test exactly that.

Experiment 1. Actuality

In Experiment 1, we examined the role of actual vs. merely possible situations. The first relevance rule of thumb states that actual situations are *ceteris paribus* more relevant than merely possible ones, for an object's behavior in the actual situation is *ceteris paribus* more consequential for our interests than its behavior in merely possible situations. The two starship cases illustrate just that: because the actual route of the starship in [1] makes successful attacks possible, it's more vulnerable than the starship in [2], which travels through the nebula that dampens explosions. Other theories don't predict this discrepancy (e.g., Manley and Wasserman 2008, p. 77).

Methods

Participants and Design. The experiment implemented a 2 (first case version: merely possible, actual) × 4 (disposition: fragile, flammable, vulnerable, soluble) factorial design. Participants were randomly assigned to two dispositions. For the first disposition, the case version was chosen randomly. For the second disposition, participants were assigned the other case version. We recruited 141 Prolific workers. Per the preregistration, this sample afforded 90% power of detecting a medium difference (d=.5) at α=.05 between judgments in *merely possible* vs. *actual* cases. Inclusion criteria in all experiments were that participants were native English speakers, had a approval rate of 95% or higher, and used a personal computer.

Materials. Here and in the following experiments, we studied four dispositions: VULNERABILITY, FRAGILITY, SOLUBILITY, and FLAMMABILITY. For each disposition, we created two test cases in which a standard stimulus would either cause a disposition-specific response only in the actual situation (*actual*) or only in a situation that is possible but not actual (*merely possible*). This pair for VULNERABILITY is given by cases [1] and [2] above.

Disposition strength was assessed with a single question (e.g., "How vulnerable is the starship?"), with responses recorded on a slider with labeled endpoints ("not vulnerable at all" to "very vulnerable") and converted to values from 0 to 100.

Results and Discussion

We analyzed strength scores using a linear mixed-effects model with case version, disposition, and their interaction as fixed effects. A random intercept for participants accounted for the

repeated measurements. The analysis revealed a significant main effect of case version ($F(1, 274)$=38.20, p<.001), confirming that the actuality of the situation where the response would happen matters for evaluating dispositional strength. We also observed a significant interaction between case version and disposition ($F(3, 274)$=2.68, p=.047), while the main effect of disposition wasn't significant ($F(3, 274)$=2.54, p=.057).

Planned contrasts confirmed that average strength was significantly higher for actual than for merely possible cases (M_{diff}=22.4, $t(138)$=6.18, p<.001, $CI_{95\%}$=[15.2, 29.5]). *Ceteris paribus*, people judge a disposition as stronger when its response would occur in the actual rather than merely possible situation. Simple effects tests showed significant actuality effects for flammability (M_{diff}=39.3, p<.001), fragility (M_{diff}=19.8, p=.007), and vulnerability (M_{diff}=18.3, p=.013). For solubility, the effect pointed in the same direction but didn't reach significance (M_{diff}=12.0, p=.101).[2] These distributions are illustrated in Figure 2.

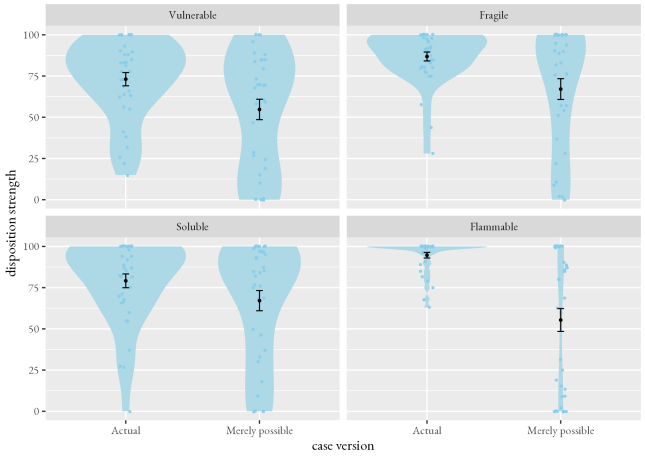

Figure 2: Disposition strength by actuality.
Here and henceforth, black circles and whiskers represent means and standard errors, and blue circles represent individual data points.

Overall, the results indicate that a disposition that would manifest in actual circumstances is perceived as stronger than one that would manifest only in merely possible ones. This sensitivity to the modal status of the stimulus is a unique prediction of relevance theory, distinguishing it from its proportion-based and counterfactual competitors.

Experiment 2. Intention

In this experiment, we examined whether intentions influence disposition ascriptions. Consider two versions of the case for FRAGILE (see the preregistration for the other cases):

[3] *Response*. John is assembling a large helicopter model and needs two pieces, a turbine for the engine and a battery. He looks up a turbine that was recommended by other modeling enthusiasts. He finds an independent consumer report, which specifies the results of tests on the turbine installed in three engines most commonly used:
1. In AirBox912, the turbine will not break provided that the engine does not work continuously for more than 14 hours.
2. In Windmaker962, the turbine will break after 1 hour of continuous use, as it cannot withstand the pressure from the engine.
3. In Otto973, the turbine will not break for any duration of use.
John double-checks what engine he is working with. It is Windmaker962. He also checks available batteries he could install in his helicopter. They all turn out to be too expensive, and with a heavy heart he decides to abandon the project because of that.

The *non-response* case differs only in the second sentence of the last paragraph.

[4] *Non-response*. John double-checks what engine he is working with. It is Otto973.

In *response*, the turbine would break in the conditions in which John intended to use it; in *non-response*, it wouldn't. However, the turbine's physical properties don't differ between the cases. Relevance theory predicts that strength scores will be higher in the first case because the mentioned situation fits the agent's intention, which makes it *ceteris paribus* more relevant than in the other case, where there is a mismatch. The competing theories predict no such effect. On counterfactual theories, the turbine is fragile (or non-fragile) in both cases, since the same counterfactuals hold in each. On proportion theories, the pieces are equally fragile because the proportions of possible worlds in which the pieces break are equal. Notice that since John abandons the project for independent reasons, the difference in evaluations isn't explained by the turbine likely breaking (or not breaking) in the future, once the plan is implemented, i.e., the effect of intention can't be explained by the effect of actuality.

Methods

Participants and Design. The experiment followed a 2 (case version: response, non-response) × 4 (disposition) design. Each participant read two cases, each with a different disposition and in a different version. Disposition pairs, order within pairs, and case versions were counterbalanced. We recruited 146 Prolific workers who satisfied the usual criteria. Per the preregistration, this sample size afforded 90% power of detecting a medium difference (d=.5) at α=.05 between judgments in *response* vs. *non-response* cases.

Materials. For each disposition, two corresponding test cases were created. In the *response* cases, the object would exhibit the response in the intended-use situation; in the *non-response* cases, it wouldn't. To ensure that the effect isn't due to differences in desirability, the stories differed on whether the agent wanted the object to exhibit the response. For *flammable* and *soluble*, the agent was looking for an object that would exhibit the response in the intended situation. For *fragile* and *vulnerable*, the agent was looking for an object that wouldn't exhibit the response in the intended situation (as in [3] and [4] above—John needed a turbine that wouldn't break in his would-be helicopter). Participants assessed the object using a slider (0–100).

[2] We don't want to place much weight on the insignificant result for solubility, as the experiments weren't powered to detect differences for individual dispositions, but only for the main effect of case version.

Results and Discussion

We analyzed strength ratings using an LME model with case version (*response* vs. *non-response*), disposition, and their interaction as fixed effects, including a random intercept for participants. The analysis revealed a significant main effect for case version ($F(1, 140.75)$=69.00, p<.001) and a significant interaction ($F(3, 267.25)$=3.45, p=.017), while the main effect for disposition wasn't significant ($F(3, 241.06)$=1.90, p=.130).

Planned contrasts on the estimated marginal means confirmed the hypothesis. Dispositions were judged significantly stronger in response than in non-response cases (M_{diff}=27.8, $CI_{95\%}$=[21.2, 34.4], $t(141)$=8.31, p<.001). Simple effects tests showed significant differences for all dispositions: vulnerability (M_{diff}=18.5, p=.005), fragility (M_{diff}=38.1, p<.001), solubility (M_{diff}=40.0, p<.001), and flammability (M_{diff}=14.6, p=.02). Figure 3 illustrates score distributions.

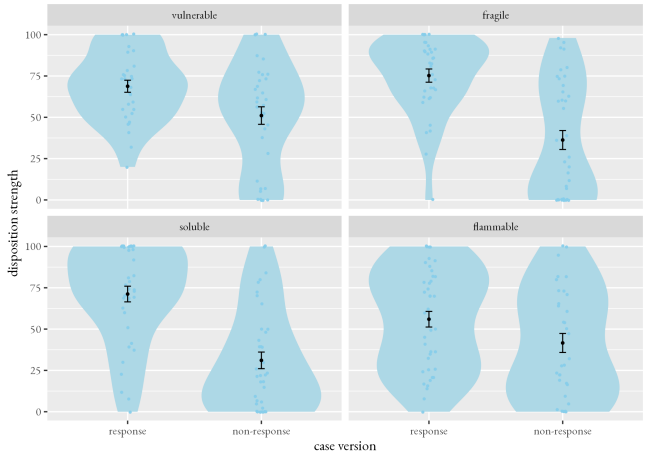

Figure 3: Strength by intention

These results indicate that, *ceteris paribus*, a disposition that would manifest in intended circumstances is perceived as stronger than one manifesting in unintended ones. For instance, an engine part that would break in a machine the user intends to build is judged more fragile than an identical part that would break in a machine the user has no interest in constructing. This influence of intention on strength is a prediction of relevance theory that the metaphysical accounts can't accommodate.

Experiment 3. Stakes

In Experiment 3, we tested whether disposition ascriptions are sensitive to stakes. We also changed another methodological aspect. In previous experiments, people judged a single object, and one may worry that the effects observed would disappear when two objects were judged together. In this experiment, we presented people with cases where two objects shared the causal basis, but their disposition-specific behavior had different downstream consequences. This way we can address the worry that, having realized that the objects have the same causal basis, people would be reluctant to evaluate them differently. Consider:

> [5] In the laboratory, there are two sealed test tubes of the same type. As they are made of fortified glass, if they fell on the floor, they wouldn't break. However, if someone accidentally stepped on them, they would break.
> The tube marked red contains a substance whose fumes are extremely poisonous but odorless. So, if the tube broke or even cracked, the researchers working in the room would start suffocating and could possibly die. The tube marked green contains a substance whose fumes smell like vanilla. So, if this tube broke or even cracked, the researchers working in the room would be pleasantly surprised.

Relevance theory predicts that the disposition would be stronger in an object whose response would be more consequential than that of the other object: the red tube should be judged as more fragile than the green tube. Metaphysical theories predict no such difference.

Methods

Participants and Design. As a pilot, this study wasn't preregistered.[3] We used a within-subjects design with the factor *stakes* (high vs. low). 96 Prolific workers assessed both objects on a single screen using sliders (0–100). Question order was counterbalanced. After the main task, participants answered whether the objects shared the causal basis (e.g., "Were the two tubes of the same shape and made of the same material?").

Materials. For each disposition, we developed one scenario describing two intrinsic duplicates. The manifestation of the disposition in one object was highly consequential (e.g., poisonous fumes), whereas it was trivial in the other object (e.g., vanilla smell).

Results and Discussion

The statistical analysis was restricted to participants who passed the causal-basis manipulation check (N=83). We used an LME model predicting strength from stakes, disposition, and their interaction, with a random intercept for participant. The analysis revealed a significant main effect of Stakes ($F(1, 79)$=15.21, p<.001). There was also a main effect of Disposition ($F(3, 79)$=4.68, p=.005), but the interaction wasn't significant ($F(3, 79)$=1.94, p=.130).

Planned contrasts confirmed that average strength was significantly higher for high-stakes than for low-stakes objects (M=65.9 vs. M=52.6; M_{diff}=13.3, $t(79)$=3.91, p<.001). Simple effects showed significant differences for fragility (M_{diff}=15.5, p=.013) and vulnerability (M_{diff}=25.5, p<.001). The effects for flammability (M_{diff}=8.2, p=.261) and solubility (M_{diff}=3.8, p=.602) followed the predicted direction but didn't reach significance. For the distributions, see Figure 4.

The results indicate that even when two objects are explicitly recognized as having the same causal basis, the one whose manifestation is more consequential is judged to have a stronger

[3] Wysocki et al. (2026) report on the final, preregistered study; its results are similar and support the same conclusions.

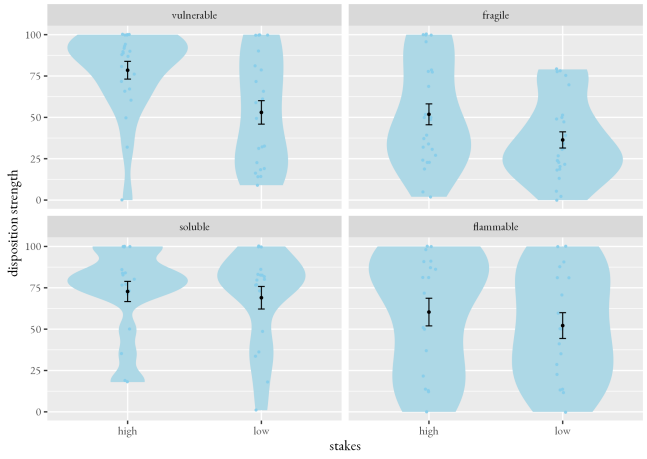

Figure 4: Strength for juxtaposed objects by stakes.

disposition. For instance, the tube holding poison is judged as more fragile than the one holding perfume. The alternative accounts fail to capture this effect.

General Discussion

We contrasted relevance theory with its metaphysical counterparts, on which dispositions reduce to the modal profiles of objects: lists of counterfactuals true of them or worlds where they exhibit disposition-specific responses. However, the three experiments show that this metaphysical picture descriptively fails. People privilege some situations over others when evaluating a disposition. Specifically, we showed that actuality, intentions, and stakes—all modes of relevance—influence disposition evaluations. These findings can be explained by relevance theory (Equation 1), which formalizes dispositional strength as a weighted sum of intensities across situations.

Now, one could object to such a stark contrast: metaphysical theories could be extended to accommodate our results. Context-dependence could enter these theories via notions of possible-world ordering, anchoring, and the like. It is indeed possible to treat our results as providing empirical material to shape these extensions. However, either such extensions are formally isomorphic to relevance theory, or they aren't. If they are, there's no need for them: our theory does the job. If they aren't, there should be empirical predictions that can distinguish between these extensions and our theory, which can only be tested once the former have been sufficiently developed. Moreover, it's far from obvious what such extensions should look like. For instance, many philosophers in the disposition literature acknowledge that the standard Lewis-Stalnaker similarity ordering wouldn't work for dispositions (Hájek, 2020; Manley & Wasserman, 2007, 2008; Vetter, 2014, 2015); adding, say, intention or stakes considerations wouldn't remedy old problems. There's also empirical evidence against the treatment of dispositions proposed within anchor semantics (Wysocki et al., 2026).

Most importantly, regardless of how competing theories could be extended, the prevailing spirit within the metaphysics of dispositions is that context should matter as little as possible. Many metaphysicians postulate dispositions to be intrinsic (Lewis, 1997; Molnar, 1999), i.e., fully determined by their causal bases, so that two intrinsic copies share all their dispositions. In light of McKitrick's (2003) examples of extrinsic dispositions, metaphysicians have divided dispositions into intrinsic, postulated as paradigmatic, and extrinsic, fully grounded in the causal basis of the object and its environment (Vetter, 2014, 2015). This perspective leaves no room for, e.g., stake- and intention-dependence, since intentions and stakes don't affect the physical makeup of the object and its environment.

Another salient objection pulls in the opposite direction: our results speak to the pragmatics of disposition ascriptions but not necessarily to their semantics—dispositional concepts may still be represented in a context-free way. While this is in principle possible, *first*, Ockham wouldn't have it that we possess two such mechanisms where one would do. And the benefits of context-insensitive representations, in addition to a context-sensitive mechanism, aren't apparent. In particular, relevance theory postulates a plausible function for disposition terms—encoding objects' possible behavior of pragmatic consequence—and that function entails context-dependence for the sheer reason that being of consequence is context-dependent. Meanwhile, the context-free approach is backed by no such story. *Second*, if the mind harbored such representations, one would expect them to manifest in cases of explicit comparison, as these cases make it salient that two objects are intrinsic copies. However, Experiment 3 shows context effects even in such cases.

Lastly, we are happy to admit that there's a lot of variation between the adjectives that relevance theory currently doesn't explain. There likely are, we concede, adjective-specific effects, and further research ought to investigate what they are and why they arise. For instance, it might turn out that the lower sensitivity of SOLUBILITY to our manipulations arises from the fact that it's harder to imagine outcomes intermediate between an object dissolving or not dissolving at all, notwithstanding our results that tablets that dissolve partially are considered less soluble than ones that dissolve fully. It also might turn out that these differences are related to gradable adjectives having either open or closed scales (Kennedy & McNally, 2005; Klein, 1980; Rotstein & Winter, 2004), with the latter being associated with less context-dependence than the former.

There's more influence of context on disposition ascriptions than is dreamt of in armchair philosophy. Bottles containing poison are more fragile than their copies containing perfume. Pieces intended as machine parts that would burn if mounted are more flammable than their identical copies intended for no such use. This all wouldn't be so surprising once you come to believe that the function of disposition ascriptions is to encode effectively the object's behavior across situations that matter (for the speaker, the listener, or anyone affected). Relevance theory springs from this very belief.

Acknowledgements

This work was supported by a Reinhart Koselleck project (WA 621/25-1), funded by the Deutsche Forschungsgemeinschaft (DFG). We would like to thank the reviewers for very insightful comments, which motivated us to develop our arguments further.

References

Baron, J., & Hershey, J. C. (1988). Outcome bias in decision evaluation. *Journal of Personality and Social Psychology*, *54*(4), 569–579. https://doi.org/10.1037/0022-3514.54.4.569

Bird, A. (1998). Dispositions and antidotes. *Philosophical Quarterly*, *48*(191), 227–234. https://doi.org/10.1111/1467-9213.00098

Choi, S. (2008). Dispositional properties and counterfactual conditionals. *Mind*, *117*(468), 795–841. https://doi.org/10.1093/mind/fzn054

Goodman, N. (1954). *Fact, fiction & forecast*. University of London.

Gundersen, L. (2002). In defence of the conditional account of dispositions. *Synthese*, *130*(3), 389–411. https://doi.org/10.1023/a:1014845625688

Hájek, A. (2020). Minkish dispositions. *Synthese*, *197*(11), 4795–4811. https://doi.org/10.1007/s11229-015-1011-y

Icard, T. F. (2016). Subjective probability as sampling propensity. *Review of Philosophy and Psychology*, *7*(4), 863–903.

Kennedy, C., & McNally, L. (2005). Scale structure, degree modification, and the semantics of gradable predicates. *Language*, *81*(2), 345–381.

Klein, E. (1980). A semantics for positive and comparative adjectives. *Linguistics and Philosophy*, *4*(1), 1–45.

Klimenta, A. (2025). *Interaction between probability and intensity in disposition ascriptions* [Thesis defended at the University of Göttingen; Matrikelnummer: 21963202].

Lewis, D. K. (1997). Finkish dispositions. *Philosophical Quarterly*, *47*(187), 143–158. https://doi.org/10.1111/1467-9213.00052

Lichtenstein, S., Slovic, P., Fischhoff, B., Layman, M., & Combs, B. (1978). Judged frequency of lethal events. *Journal of Experimental Psychology: Human Learning and Memory*, *4*(6), 551–578. https://doi.org/10.1037/0278-7393.4.6.551

Lieder, F., Griffiths, T. L., & Hsu, M. (2018). Overrepresentation of extreme events in decision making reflects rational use of cognitive resources. *Psychological Review*, *125*(1), 1–32. https://doi.org/10.1037/rev0000075

Lucas, C. G., & Kemp, C. (2015). An improved probabilistic account of counterfactual reasoning. *Psychological Review*, *122*(4), 703–734. https://doi.org/10.1037/a0039610

Manley, D., & Wasserman, R. (2007). A gradable approach to dispositions. *Philosophical Quarterly*, *57*(226), 68–75. https://doi.org/10.1111/j.1467-9213.2007.469.x

Manley, D., & Wasserman, R. (2008). On linking dispositions and conditionals. *Mind*, *117*(465), 59–84. https://doi.org/10.1093/mind/fzn003

McKitrick, J. (2003). A case for extrinsic dispositions. *Australasian Journal of Philosophy*, *81*(2), 155–174. https://doi.org/10.1080/713659629

Molnar, G. (1999). Are dispositions reducible? *Philosophical Quarterly*, *49*(194), 1–17. https://doi.org/10.1111/1467-9213.00125

Quillien, T., & Lucas, C. G. (2024). Counterfactuals and the logic of causal selection. *Psychological Review*, *131*(5), 1208–1234. https://doi.org/10.1037/rev0000428

Rotstein, C., & Winter, Y. (2004). Total adjectives vs. partial adjectives: Scale structure and higher-order modification. *Natural Language Semantics*, *12*(3), 259–288.

Sunstein, C. R., & Zeckhauser, R. (2011). Overreaction to fearsome risks. *Environmental and Resource Economics*, *48*(3), 435–449. https://doi.org/10.1007/s10640-010-9449-3

Vetter, B. (2014). Dispositions without conditionals. *Mind*, *123*(489), 129–156. https://doi.org/10.1093/mind/fzu032

Vetter, B. (2015). *Potentiality: From dispositions to modality*. Oxford University Press.

Wysocki, T., Reins, L., & Waldmann, M. R. (2026). *Relevance theory of dispositions* [Manuscript submitted for publication].