The psychological representation of modality

Jonathan Phillips and Joshua Knobe

Abstract A series of recent studies have explored the impact of people's judgments regarding physical law, morality, and probability. Surprisingly, such studies indicate that these three apparently unrelated types of judgments often have precisely the same impact. We argue that these findings provide evidence for a more general hypothesis about the kind of cognition people use to think about possibilities. Specifically, we suggest that this aspect of people's cognition is best understood using an idea developed within work in the formal semantics tradition, namely the notion of modality. On the view we propose, people may have separate representations for physical, moral and probabilistic considerations, but they also integrate these various considerations into a unified representation of modality.

The past few decades have witnessed an explosion of research on the way that humans understand physics, morality, and probability. This research has explored the impact on people's cognition of regarding an event as physically impossible, morally wrong, or highly improbable (Griffiths et al. 2010, Knobe 2010, Marr 1982, Spelke 1990).

One remarkably underappreciated fact is that these factors often have the exact *same* impact. That is, there are a number of different respects in which a judgment that something violates physical law has the same impact as a judgment that it is morally wrong, which in turn has the same impact as a judgment that it is statistically improbable. As we go on to argue, the number of different phenomena in which this precise pattern can be found suggests that there is some deeper connection between these factors.

Our basic proposal is that each of these three factors is relevant to how people represent possibilities. Thus, we propose that there is a single underlying representation that is affected by all three factors and that this underlying representation plays an important role across a number of different psychological phenomena. We refer to this representation as the *psychological representation of modality*.

We begin by considering a number of these phenomena in turn and present the empirical evidence that suggests that physical, moral and probabilistic considerations have a similar impact in each. We then propose an account of the psychological representation of modality and show that it provides a unified account of these different effects.

1 Phenomena to be explained

The phenomena we take up in this paper come from four separate research programs. These phenomena have therefore been investigated by independent researchers who have employed distinct methodologies and relied on different theoretical frameworks in interpreting their findings. Yet despite these numerous differences, one notices a strikingly similarity in the observations made in each of these separate research programs: physical, moral and probabilistic considerations all seem to have highly similar effects on judgments in each area of research. While we discuss each of these research programs at length below, it will be helpful to begin by getting a rough sense of the similarity that we are interested in.

1.1 Development of thinking about possibilities

To adults, it seems obvious that certain things are possible (e.g., throwing one's hat into the air), but that others are impossible (e.g., transforming one's hat into a bottle of whiskey). Research in cognitive development has investigated how young children think about possibilities such as these, and how their understanding of them changes over the course of development.

Much like adults, young children have little trouble judging that events that require violations of the laws of physics cannot actually happen (Levy et al. 1995). Yet, their understanding of possibility also differs in remarkable ways from adults' understanding. Unlike adults, children often explicitly judge that *improbable* events are impossible (Shtulman 2009, Shtulman & Carey 2007). Moreover, young children (3- to 5-year-olds) also judge that*morally bad* events can't happen, are impossible, and even require magic to happen (Chernyak et al. 2013, Kalish 1998, Kushnir et al. 2015, Phillips & Bloom 2017). Thus, young children tend to regard events as impossible if they involve violations of the laws of physics, are morally bad, or are statistically improbable.

1.2 Freedom

A central topic in both ancient and contemporary philosophy has been the distinction between cases in which a person acts freely and cases in which a person is instead forced to act. This distinction is at the foundation of many philosophical debates ranging from coercion to free will to political liberty (Aquinas [1273]1952, Aristotle [340 BCE]2002, Berlin 1970).

While the theoretical discussion of these issues continues, experimental philosophers have sought to inform these various debates by conducting studies that investigate which factors are relevant to how people ordinarily make judgments.

Unsurprisingly these studies show that people are highly sensitive to what can physically occur. That is, they more judge that people are free to do something if they were *physically* capable of not doing it (see, e.g., Woolfolk et al. 2006). However, more recent research has also uncovered that morality may play a similar role: participants are more inclined to judge that people are free to do something when they regard it as *morally good* (Phillips & Knobe 2009, Young & Phillips 2011). In short, one finds that physical and moral considerations have similar effects on ordinary judgments of force and freedom.

1.3 Causal selection

Suppose that a forest fire was started by the simultaneous presence of oxygen, dry leaves, and a lit match in close proximity. Which of these was the cause of the forest fire? Given just the physics of the scenario, the fire clearly would not have occurred without all three elements. Yet the way in which we select the cause of the forest fire seems to go far beyond this. Typically, people would say that the match caused the forest fire, or they might (depending on the circumstances) be inclined to regard the dry leaves as the cause, but almost no one would ever consider the oxygen to be the cause of the forest fire. The problem of *causal selection* is the problem of explaining why people privilege certain causal factors over others in cases like these.

A large number of researchers working in philosophy, computer science, law, and cognitive psychology have argued over the best way of formally accounting for how causes are selected from amongst all of the things that contributed to an outcome (e.g., Halpern & Hitchcock 2015, Hart & Honoré 1985, Hitchcock & Knobe 2009, Woodward 2006). In these formal discussions, one factor that has often been noticed by those trying to resolve the problem of causal selection is that *improbable* events are often selected as the cause of the eventual outcome (see, e.g., Hart & Honoré 1985). Yet, probability is not the only factor which helps determine which causes people select. One relatively surprising factor is the moral status of the event. Specifically, people seem to pick out *morally bad* events as causes of eventual outcomes, even when the outcomes are neutral or good (Hitchcock & Knobe 2009). As a number of researchers have noted (e.g., Halpern & Hitchcock 2015), there is a notable similarity in these judgments between the effect of probabilistic considerations and the effect of moral considerations (Kominsky et al. 2015)).

1.4 Explicit counterfactual reasoning

Imagine you were delayed by a traffic jam on the way to the airport and that you missed your flight as a result. You might engage in a great deal of nonconscious processing about these events, but in addition, you would likely also engage in the explicit process of reasoning through what might have happened if things had gone differently. We refer to this conscious process as *explicit counterfactual reasoning*. Work in cognitive and social psychology, and more recently, neuroscience, has focused on understanding humans' general capacity to engage in this sort of explicit counterfactual reasoning (Byrne 2016, De Brigard et al. 2013). Within this field, one important research question has been *which* possibilities people simulate when they engage in explicit counterfactual reasoning (Kahneman & Miller 1986, Roese 1994). We may typically consider, for instance, what might have happened if there had not been a traffic jam, but we rarely consider what might have happened if the flight's departure had been delayed by a surprise ice storm.

One of the most robust findings from this research is that people tend to consider counterfactuals in which typical or *probable* events take place, and rarely consider counterfactuals in which improbable events occur. Even less frequently do they consider counterfactuals involving physical violations. Strikingly, in much the same way, people also tend to consider counterfactuals in which *morally good*, rather than morally bad things occur (McCloy & Byrne 2000b, N'gbala & Branscombe 1995). That is, which counterfactuals people tend to entertain is affected by physical, probabilistic and moral considerations.

1.5 Taking stock

Across four different phenomena, we find the same factors (physics, morality and probability) playing a role in judgments relevant to each of these markedly different questions. Moreover, we also find precisely the same sort of impact when the event in question involves a violation of physics, is morally bad, or is statistically improbable. Again and again, we see that the effect of something being morally bad is quite similar to the effect of something being improbable, but never find that it is similar to the effect of something being statistically probable or not involving a violation of physics. This striking similarity across these different kinds of cognition naturally suggests that there may be some more general, unified way of explaining the impact of these factors throughout the diverse phenomena where one finds them playing a role.

One approach to offering a unified account would be to argue that the impact of all of these factors reduces to the impact of some single factor. For example, one could argue that the impact of morality and physics can actually be explained in terms of differences in probability. On this view, there is no independent effect of morality in these various phenomena; it is simply that morally bad actions tend to be less probable. However, a number of empirical results make this sort of reductive approach look unlikely to succeed. In studies of causal selection, for example, one finds that even when it is extremely likely that an immoral action will be done (and extremely unlikely that a morally neutral action will be done instead), people still select the immoral action as the cause rather than morally neutral one (Roxborough & Cumby 2009). Similar patterns have been found for a number of other various phenomena as well, including the development of thinking about possibility (Phillips & Bloom 2017).

Let us therefore consider an alternative approach. Instead of trying to reduce some of these factors to others, we ask what all of the factors may have in common. Given the diversity of the three factors under discussion (physical possibility, morality and probability) and the diversity of the different kinds of cognition in which these factors play a role, any account which offers a unified explanation will necessarily involve a certain degree of abstraction. The question we now face is where such an account is to be found.

2 The linguistic representation of modality

To address this question, we begin by making a detour in a perhaps unexpected direction, namely, to the field of *formal semantics*. Research in formal semantics is concerned with questions about language. Specifically, formal semanticists are concerned with questions about the meanings of linguistic expressions, and often proceed by developing formal models that capture those meanings. It may seem at first that this is not a particularly plausible place to go looking for explanations for the phenomena under discussion here, but we will try to show that this initial impression is misleading. A theoretical framework that comes out of research in formal semantics actually gives us precisely the resources we need to begin making progress on these issues.

We proceed in two steps. First, we focus on questions about language. We look at the meanings of certain linguistic expressions and at a framework that formal semantics has developed to capture them. Then, with this framework in hand, we zoom back out. We return to the psychological phenomena introduced above and suggest that this framework gives us the tools we need to explain them.

Let us begin by looking at a striking fact about the linguistic expressions people use to talk about physical laws, morality and probability. Oddly enough, people sometimes use the very same expressions to talk about these three seemingly unrelated issues. As one example, consider the English expressions *can* and *can't*. These expressions can be used in claims about physics as in (1).

(1) Particles can't go faster than the speed of light.

But they can also be used in claims about morality as in (2).

(2) You can't keep treating her that way – look at how upset she is!

Moreover, they can be used in claims about probability as in (3).

(3) You can't complete an entire career in research without making a few mistakes.

We introduced this phenomenon by using the example of the English expressions 'can' and 'can't', but the points we have been making here actually apply to a whole class of different expressions. These expressions are called *modals*. In English, they include 'can', 'have to', and 'must', among others. Other languages have other modal systems (German has 'können', 'müssen' and 'sollen'; French has 'pouvoir' and 'falloir'; Russian has 'может' and 'надо'). A question now arises as to why language allows us to use the very same expressions in these three seemingly very different ways.

One possible view would be that modal expressions are simply ambiguous. For example, one might think that the English word 'can' has a number of distinct senses. It would have a physical sense, a moral sense, a probabilistic sense, and a few others as well, but we just happen to use the same term for all of them. On this picture, there would be no way to develop a unified account of the meaning of this word. We would have to develop a completely separate account for each of the separate senses. One problem that this approach has faced, however, comes from the cross-linguistic research on modal systems in different languages. Rather than finding that these different proposed senses have separate terms in other languages, one instead finds that the pattern observed in English also occurs in many other languages as well: the same modal terms can often be used to make physical, moral and probabilistic claims (for cross-linguistic work on modality, see, e.g., Matthewson 2016, Nauze 2008, Vander Klok 2012).

Contemporary research in formal semantics has therefore moved toward a very different alternative, with the field now strongly favoring a single unified theory that accounts for all of these uses. The basic framework underlying this work was first introduced in a series of influential papers by Kratzer 1977, 1981 and has since been developed in numerous ways, both theoretically and empirically (for a book-length review, see Portner 2009). It is now clearly the standard approach, and we will be building on it in the hypothesis we propose here.[1]

[1] It should be noted that there have been various challenges to this standard approach. It has been objected that the approach does not correctly capture the semantics of epistemic modals (Egan et al. 2005, Veltman et al. 1996, MacFarlane 2009, Yalcin 2007, 2015), that it does not capture the information-sensitivity of deontic modals (Cariani 2013, Kolodny & MacFarlane 2010), and that the correct semantics would rely on scales rather than quantifiers (Lassiter 2011). (For some replies to these objections, see (Björnsson & Finlay 2010, Dowell 2011, Khoo 2015, Klecha 2014, von Fintel 2012).

Although these challenges raise important issues in formal semantics, they will not be especially

The framework for modality is usually presented quite formally, but before getting into the formal details, we will try to present the basic ideas at a more intuitive level. In essence, the framework involves two keys ideas.

The first is that people are concerned not only with how things actually are, but also with other possibilities – other ways that things could have been. For example, the actual winner of the 2012 election was Obama, but we can imagine possibilities in which the winner was Romney or Gingrich or Cain. There are even possibilities in which highly far-fetched things occur. For example, there are possibilities in which we pass a constitutional amendment and then elect a gerbil as our President.

The second is that people do not treat all possibilities equally. In any given context, people are not concerned with all conceivable possibilities but only with the possibilities in a more restricted set. Thus, in an ordinary conversation about politics, people might be concerned with possibilities in which Romney becomes President but not with possibilities in which a gerbil becomes President. They simply ignore such possibilities entirely. We will refer to the set of possibilities people are concerned with in any given context as the *domain*.

With these two notions in the background, we can now give a rough account of what the modal expressions of our language mean. Basically, a sentence like (4-a) means something like (4-b).

(4) a. John can do that.
b. There is a possibility in the domain in which John does that.

And (5-a) means something like (5-b).

(5) a. John has to do that.
b. In all possibilities in the domain, John does that.

Similarly for each of the other modal expressions.

We can now offer a simple explanation for the fact that physical, moral and probabilistic considerations all impact people's use of modals. The idea is simply that *all of these considerations have an influence on which possibilities are included in the domain*. In other words, we don't need to suppose that modal expressions themselves have a number of separate meanings. Instead, we can offer a unified theory of what each modal expression means. It's just that the meaning of each modal expression is given in terms of a domain, and the domain is determined

relevant to the psychological questions under investigation here. In particular, even the researchers who have raised forceful objections to the standard view would agree with the specific aspects of that view that we draw on in what follows. In fact, the basic commitments of the current proposal can be seen in early work on modal frames and on deontic and alethic logics (Kripke 1963, von Wright 1953).

differently in different contexts.

For a simple example, take the case of physics. In many contexts, people simply ignore all possibilities that involve violations of physical laws. We can then make claims about physical law using a sentence like (1).

(1) Particles can't go faster than the speed of light.

This sentence says that, among the possibilities in the domain (i.e., the possibilities in which there are no physical violations), there is no possibility in which particles go faster than the speed of light.

The same basic approach can then be applied for sentences involving moral considerations. In many contexts, we simply ignore the possibilities in which people do morally bad things. We can then use a sentence like (2).

(2) You can't keep treating her that way – look at how upset she is!

This sentence says that, among all the possibilities in the domain (i.e., the possibilities in which you do not do actions that are morally bad), there are none in which you treat her that way. This same approach can also be applied to sentences involving probability, like (3).

(3) You can't complete an entire career in research without making a few mistakes.

In this case, the sentence says that among all of the possibilities in the domain, (i.e., the possibilities in which highly improbable things do not happen), there are none in which you complete an entire career in research without making a few mistakes.

This same basic approach can be straightforwardly extended to the many other factors (e.g., goals, conventional norms, prudential concerns) that also influence which possibilities are simply ignored and which are included in the domain. Thus, while we have a unified way of accounting for the meaning modal expressions, the results of applying this single framework will depend on which factors are relevant in a given context.

Thus far, we have been presenting this structure in a more informal, intuitive manner, but in the existing literature, it is usually presented more formally. Following Kratzer 1977, we can introduce a function, f, that maps each possibility onto a set of possibilities. Then suppose we are wondering whether a particular proposition has to be the case. We use the function f to map our actual situation onto a set of possibilities. That set then serves as the domain, and we check to see whether φ is the case in all of the possibilities in that domain. More formally:

$$[\![\textsf{have to } \varphi]\!]^{w,f}\text{=1 iff for all } v \in f(w),\ [\![\varphi]\!]^{v,f}\text{=1}$$

For example, suppose we are trying to determine whether, in a particular situation, it has to be the case that it is raining outside. First, we use the function f to go from the situation in question to a set of possibilities that serves as the domain. Then we ask whether it is the case that it is raining outside in all of the possibilities within this set. [2]

We now need to add just one further idea that will play a key role in the argument that follows. Despite everything we have said thus far, one might still think that there is some important sense in which the different uses of modal expressions are *separate*. That is, even if there is a common logical structure at work in cases where we are talking about moral considerations, probabilistic considerations, or considerations of some other type, one might think that any given modal will be evaluated based on just one type of consideration and no others. Thus, one might think that sentences like (2) will be evaluated purely with respect to moral considerations, sentences like (3) purely with respect to probabilistic considerations, and so forth.

Importantly, existing work suggest that many natural language modals do not actually work in this way. In many cases, a whole variety of different considerations can play a role in the evaluation of a single modal. In other words, even if we are looking just at a single modal, it may be that the domain of that modal is determined jointly by physical, moral and probabilistic considerations (Knobe & Szabó 2013).

For one example, take the sentence:

(6) To get an A in this class, you have to study for the final exam.

This sentence seems to say that all of the relevant possibilities in which you get an A

[2] This formalization makes a number of simplifying assumptions. In particular, it assumes a simple dichotomy whereby every possibility either falls inside or outside of the relevant set. To capture people's judgments, we need to additionally introduce an *order* on possibilities (Kratzer 1981). The suggestion then would be that we don't just have a dichotomy (with each possibility either in or out) but an ordering (with some possibilities ranked higher than others). On this more precise formalization, we introduce two different functions, a 'modal base' (usually called f) and an 'ordering source' (usually called g). The modal base maps each possibility onto a set of possibilities, while the ordering source maps each possibility onto an order on possibilities. 'Have to' can then be defined:

$$[\![\textsf{have to } \varphi]\!]^{w,f,g}\text{=1 iff for all } v \in f(w) \text{ such that there is no } v' \in f(w) \\ \text{such that } g(w)(v',v),\ [\![\varphi]\!]^{v,f,g}\text{=1}$$

In other words, we first use the function f to map the situation at hand onto a set of possibilities; then we use the function g to map that situation onto an order. Finally, we take the possibilities within the set that are most highly ranked in the order. These possibilities serve as our domain of quantification. Note that this formalization requires what is sometimes known as the 'limit assumption' (Lewis 1981). (We return to the issue at the heart of this more complex formalization in §3.1

in this class are possibilities in which you study for the final exam. But how does one determine which possibilities are the relevant ones and which we can simply ignore? Well, a whole host of different considerations seem to play a role here. Consider the possibility of getting an A by using telepathy to read the teacher's mind. We ignore this possibility because it involves physical violations. Or consider the possibility of getting an A by cheating on the exam. We ignore this possibility because it is morally bad. Or take the possibility that you don't study for the exam but correctly guess the answers to all of the questions on the final exam. We ignore this one because it is so wildly improbable. It is only because all of these possibilities fall outside the domain that a sentence like (6) can come out true.

Let us now sum up. Research in formal semantics has led to the development of a unified theory of modals. At the core of this theory is the idea of a certain sort of representation of possibilities. More specifically, the theory posits a representation of the set of possibilities we are concerned with in any given context. We will refer to this sort of representation in what follows as a representation of *modality*.

With that idea in hand, it then becomes possible to explain why physical, moral and probabilistic considerations (along with various other factors) all impact people's use of modals. The answer is that all of these considerations are relevant to the question of which possibilities we are concerned with. We therefore arrive at a surprising conclusion: physical, moral and probabilistic considerations are all relevant to questions of modality.

3 The psychological representation of modality

In the previous section, we focused specifically on issues in the study of language. We now want to argue that these linguistic issues are giving us a glimpse of a far broader fact about human cognition, a fact that holds even for aspects of cognition that have no particular connection to language.

The first thing to notice is that many of the issues we have been discussing in the study of language also arise in the study of non-linguistic cognition. As we emphasized in the previous section, we need a theory about possibilities to explain people's use of certain linguistic expressions ('can', 'have to'). However, it seems that we also need a theory about possibilities to explain aspects of cognition that do not involve these linguistic expressions (e.g., causal judgments). Thus, any adequate account of these aspects of cognition will have to involve some sort of framework for describing the way possibilities figure in people's reasoning.

We can now introduce the hypothesis that we will be elaborating and defending throughout the remainder of the paper. Our hypothesis is that people's representation of possibilities in cognition more broadly can be captured using the framework we reviewed in the previous section. More specifically, we propose to take two key claims that have already been defended within the existing literature in the study of language and apply them to the study of cognition.

i. People are capable of representing possibilities, but they do not treat all possibilities equally. Instead, they are concerned in any given context only with the possibilities in a more restricted domain.

ii. This domain is determined jointly by physical, moral and probabilistic considerations (and a whole lot else besides). In particular, a possibility tends not to be represented as in the domain to the extent that it is a physical violation, morally bad, or highly improbable.

In short, our claim is that people's capacity for thinking about possibilities is governed by a kind of representation that integrates a number of different types of considerations (physical, moral, probabilistic). We will refer to this representation as the *psychological representation of modality*.

Perhaps the best way to get a sense for the key idea here is to consider a concrete example. Imagine an agent whose car breaks down on the way to the airport. She is now trying to figure out what to do next, and she is considering a range of possible options. At least in principle, this agent might entertain the following five possibilities:

Hail a taxi

Call a friend

Steal someone else's car

Convince the airport to delay the flight

Levitate and fly to the airport

Looking at this list of possibilities, she might immediately judge that one of them is morally bad and another involves a physical violation. The question now is how these judgments impact her representation of the possibilities themselves.

One plausible view would be that people simply represent possibilities as having certain properties. These properties then come in various different types (physical, moral, probabilistic, etc.). On this view, our agent would associate the possibilities she was entertaining with properties of different types.

Hail a taxi

Call a friend

Steal someone else's car (*morally bad*)

Convince the airport to delay the flight (*statistically improbable*)

Levitate and fly to the airport (*physical violation)*

People's understanding of these three properties has already been explored in existing work, and that work gives us a good sense of how these representations might impact the agent's subsequent cognition (Griffiths et al. 2010, Spelke 1990, Sripada & Stich 2006). Thus, it is plausible that these relatively well-understood judgments are at play here too.

It is worth emphasizing how intuitive and straightforward this basic idea is. In many different instances, people independently use their capacities to make physical, statistical and moral judgments. For example, when deciding whether or not to punish another person, people rely on a specific capacity for moral cognition (Cushman 2008). When make decisions involving uncertainty, people rely on a specific capacity for statistical cognition (Halpern 2003); and when predicting the behavior of physical objects, people rely on a specific capacity for reasoning about physics (McCloskey et al. 1983). Thus, an obvious hypothesis is that people use these various capacities when reasoning about possibilities too. Some are regarded as violating physics; others as statistically unlikely; and still others as morally wrong. On this approach, there is nothing similar, for example, about regarding a possibility as morally bad and regarding it as a physical violation. These are just two completely separate properties, which one would expect to play completely independent roles.

What we are suggesting is that there is more to the story. People represent possibilities as having various types of properties, but they can also make use of a different sort of representation. They can represent a possibility as falling entirely outside the domain. Thus, our agent might represent her situation as follows:

Hail a taxi

Call a friend

~~Steal someone else's car~~

~~Convince the airport to delay the flight~~

~~Levitate and fly to the airport~~

On this latter view, there actually is something deeply similar about seeing a possibility as morally bad or as a physical violation, and both of these are similar in turn to seeing the possibility as highly improbable. Possibilities of all of these types are not represented as falling within the domain. Speaking informally, one might describe such possibilities as ‘irrelevant’, ‘not worth considering’, ‘not real possibilities at all’. As we will sometimes put it here, these possibilities are regarded as *ruled out*.

To treat these possibilities in this way, the agent need not have thought about them in terms of linguistic modality. In fact, she need not have thought about them in linguistic terms at all; she can simply have begun to consider what to do given her situation. This basic idea is the core of the current proposal for a *psychological* capacity for representing possibilities.

Still, although the representation we are discussing is not itself linguistic, we suggest that it displays the very same surprising qualities that have been uncovered in the linguistic representation of modality. Most importantly, the key claim will be that people do not have a separate representation for each different type of consideration (physical, moral, probabilistic). Rather, people have a single unified representation of the domain that is shaped jointly by all of these considerations.

3.1 The nature of modal representation

At the core of our account is a claim about the ways in which people’s psychological representation of modality is influenced by various different considerations. One obvious hypothesis would be that people take into account different considerations in different cases, depending on the goals they have in the situation at hand. Thus, there might be certain cases in which physical, moral, and the probabilistic considerations are all worth taking into account (as in our example of the agent whose car breaks down), and in those cases, people should be influenced by all three of these considerations. However, there might be other cases in which it seems clear that it would only make sense to take one of these considerations into account (e.g., focusing only on physical considerations). In such cases, one might think that people would focus exclusively on the one consideration they regard as genuinely important to the question they are trying to address, and that which possibilities are included in the domain would no longer be constrained by the various other kinds of considerations.

We reject this hypothesis. Instead, we suggest that people show a quite general tendency to construct a domain based on physical, moral and probabilistic considerations. They show a tendency to take all of these considerations into account when it makes sense to do so, but they also show a tendency to take all of these considerations into account when there is nothing to be gained by doing so. Accordingly, we predict that people will tend to exclude from the domain certain possibilities (e.g., those that are morally bad) even when they can easily see on reflection that there is no rational reason to do so.

This is not to say that there are no cases in which human beings form a modal representation that relies more specifically on just one type of consideration. Such cases certainly do exist. However, our suggestion is that these instances are best understood as an exception and a rarity. The capacity to represent the domain in a way that focuses just on one type of consideration is an achievement that requires a certain amount of effort, sophistication, and conscious reasoning. The default tendency of our capacity for modal representation is to blend together all of these different considerations.

The four phenomena we consider in depth provide evidence for this more radical aspect of our proposal. Upon reflection, it does not seem reasonable to take into account all three considerations (physical, moral, statistical) in all four of these phenomena. Yet, in each case, we do find all of these considerations playing a role, whether in the development of thinking about possibilities, in intuitions of whether somone acted freely or was forced, in judgments causal selection, or in explicit counterfactual reasoning.

The basic suggestion we are arguing for is that the psychological representation of modality should not be understood as simply the result of domain-general reasoning applied to possibilities, but rather as an independent process that typically operates relatively inflexibly and without conscious deliberation. The operation of this process then impacts judgments across a broad range of what might appear to be quite different aspects of people's subsequent cognition.

3.2 The central role of modal representations

What we want to show now is that this hypothesis can help to explain some of the puzzling phenomena discussed above. As we noted there, people make judgments about a number of apparently distinct properties (physical, probabilistic, moral), and these judgments impact a number of apparently distinct processes (development of thinking about possibilities, causal selection, assessments of freedom, explicit counterfactual reasoning). The question now is how to understand that impact. Why exactly do each of the judgments influence each of these processes?

The most obvious approach to explaining these effects would be to suggest that people have a number of completely separate representations (a physical representation, a moral representation, etc.) and that these different representations all happen to bring about the same response. We refer to this first sort of view as the *separate representations view*. On this view, we assume that people only represent the various specific properties, and explain the impact of these representations on various phenomena by positing a relatively complex web of relationships (see Figure 1).

Yet this picture immediately leaves us with a mystery. Again and again, we find that the same collection of different representations all influence a single process. One wants some kind of explanation for the fact that these purportedly different

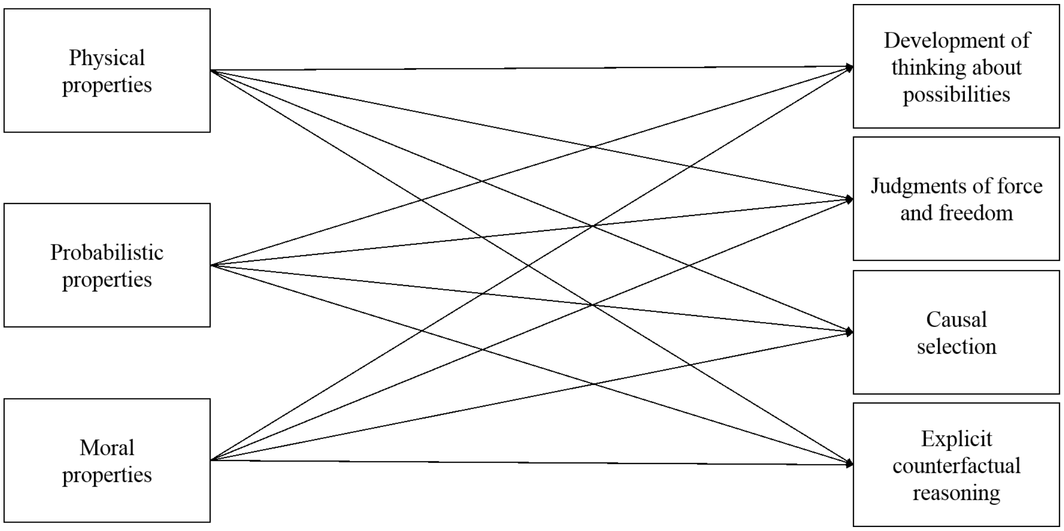
Physical properties
Probabilistic properties
Moral properties
Development of thinking about possibilities
Judgments of force and freedom
Causal selection
Explicit counterfactual reasoning

Figure 1: Schematic model of the impact of physical, probabilistic and moral properties on four different phenomena according to the separate representations view

representations so often travel together.

The present hypothesis opens up the possibility of explaining these effects in a different way. Perhaps these effects are driven by a single representation that is itself shaped by a number of different considerations (the representation of modality). We refer to this latter possibility as the *modal representation view*, and we argue that it provides the better explanation of the four phenomena under discussion here.

On this hypothesis, we don't need a web of distinct connections between the processes and the representations of the various separate properties. Instead, things become far simpler. At the core of the explanation is the claim that judgments about all three of the separate properties can impact representations of modality. All we need then is a connection between each of the processes and this unified representation of modality (see Figure 2). Given the way that people understand modality, it follows immediately that each of these processes will be influenced by judgments of each of the separate properties.

In fact, it is even simpler than that. It is not as though we have to go through each of the separate processes (explicit counterfactual reasoning, causal selection, etc.) and posit separate connections between each of them and the representation of modality. Rather, the key predictions follow immediately from the basic notion of what the psychological representation of modality is. In each case, the psychological representation of modality will play exactly the same role. Specifically, the claim

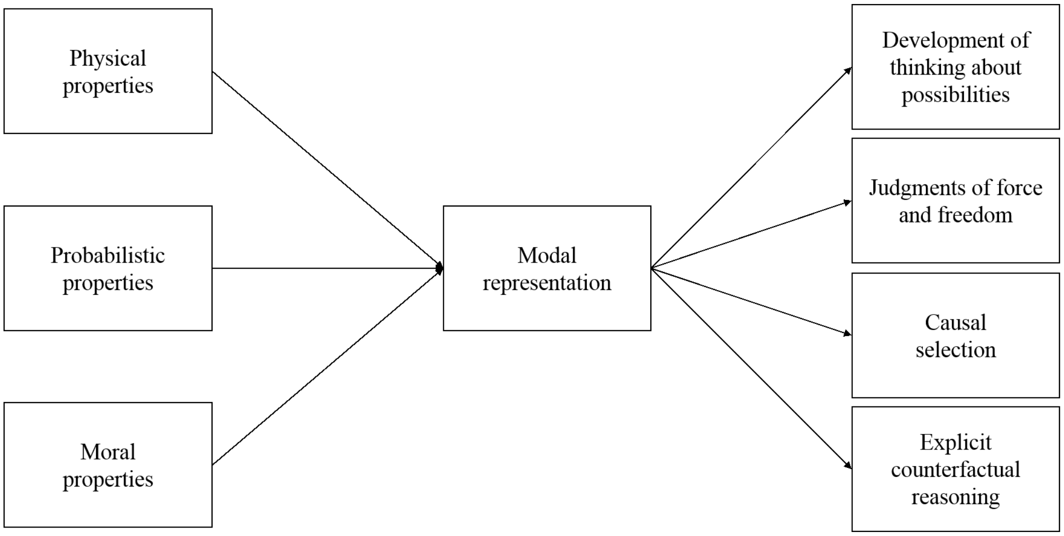
Physical properties
Probabilistic properties
Moral properties
Modal representation
Development of thinking about possibilities
Judgments of force and freedom
Causal selection
Explicit counterfactual reasoning

Figure 2: Schematic model of the impact of physical, probabilistic and moral properties on four different phenomena according to the modal representation view

will be that all of these processes operate only on possibilities that fall within the domain and ignore those possibilities that are ruled out.

4 Explaining the phenomena

Let us now return to the four puzzling phenomena with which we began. We will be taking up each of them, this time in far more detail, and arguing that each is better explained in terms of the modal representation view than in terms of the separate representations view.

In each case, we adopt the same basic explanatory approach. First, we argue that the phenomenon involves a cognitive process that operates on representations of possibilities. This first part of the explanation does not involve introducing anything new or original. Existing research on each of the phenomena has already suggested they involve cognitive processes that operate on possibilities, and we will be relying heavily on that research here. We then add just one key element: the psychological representation of modality. In other words, we claim that the cognitive processes in question do not treat all possibilities equally but instead ignore those possibilities that are ruled out. We argue that this claim gives us just what we need to explain each of the phenomena.

Note that the claim we are defending here is very modest in one sense and yet very ambitious in another. The claim is modest in that our hypothesis would play only a small role in a full account of each of the phenomena under discussion. For example, if one wanted to put together a complete account of the way people make causal judgments, the most important task would be to develop a theory of the specific cognitive processes involved in causal cognition (Gerstenberg & Tenenbaum in press, Gopnik et al. 2004, Halpern & Pearl 2005). The representation of modality adds just one further piece to the puzzle. The same is also true for the other phenomena we take up.

Yet, at the same time, there is a sense in which the claim is highly ambitious. We are suggesting that the modal representation view can prove helpful in understanding phenomena in a wide array of different fields, including developmental psychology, social psychology, causal modeling and experimental philosophy. In short, it may be that the representation of modality is just one piece in each puzzle, but it is remarkable to see how this very same piece shows up in a large number of seemingly unrelated puzzles.

4.1 The development of understanding possibility

The field of cognitive development is broadly interested in the emergence of the cognitive processes in childhood, and in the way that these processes change as children grow older. Research in this area has helped to provide a clear picture of the default assumptions involved in many different cognitive processes by investigating them before the emergence of the more sophisticated ways of reasoning one finds in adults. Recently, a growing number of studies in cognitive development have taken this approach to the development of our understanding of *possibility*. This research has now begun to paint a helpful picture of the simplest operation of the mechanisms involved in reasoning about possibilities, long before the changes that eventually give rise to the sophisticated ways adults reason about possibility.

Typically, when adults are asked to make judgments about possibility, they tend to judge that events are impossible only when they involve some sort of physical violation. In contrast, recent work in cognitive development has shown that young children's understanding of possibility substantially differs from that of adults. Young children judge that things are impossible not only when they involve physical violations, but also when they are highly improbable or morally bad. That is, while adults' judgments are primarily based only on physical considerations, one finds that physical, moral, and probabilistic considerations all play the same role in children's judgments.

4.1.1 The data

A series of studies suggest that children regard improbable events as impossible. Shtulman and Carey 2007 presented both adults and 4- to 8-year-old children with a number of different actions, some of which involved physical violations (e.g., eating lightning for dinner), some of which were highly statistically improbable (e.g., finding an alligator under one's bed), and some of which were completely ordinary (e.g., wearing a baseball cap). For each of these actions, children were asked whether a person could do that action in real life. Unsurprisingly, adults judged that the actions involving physical violations could not be done in real life, but that the ordinary and improbable actions could actually be done in real life. In contrast, younger children judged that both the *physical violations* and the *improbable* actions could not be done in real life. Subsequent research (Shtulman 2009) additionally found that young children judge that such improbable actions are actually 'impossible' across a number of different types of events (physical, psychological, biological).

Similarly, children appear to regard *morally bad* actions as impossible. Initial evidence for this comes from researchers who have directly asked children questions involving modals, e.g., whether an agent 'can' or 'could have' done a particular action (Chernyak et al. 2013, Levy et al. 1995, Kalish 1998). In some of the earlier research on this topic, participants were shown pictures of other children and asked whether the child depicted could perform a particular action (Kalish 1998). Some of the actions involved physical violations (turning into a fish), while other actions were instead morally bad or socially impermissible (taking a bath while wearing shoes). In their responses to these modal questions, children did not differentiate between these different types of violations, and reported that all of these actions could not be done. Subsequent studies then extended this line of research by demonstrating that 4- to 5-year-old children judge that events involving morally bad actions are as impossible as events involving physical violations. Similarly, young children judge that morally bad events would require magic to happen (Phillips & Bloom 2017).

Importantly, children's judgments that it would be impossible or require magic for something immoral to happen do not seem to be driven by the mere likelihood of those events occurring. Adults judged that every immoral event used in these studies (e.g., a boy lying to his dad about not feeling well because he doesn't want to have to go to school) was much more likely than the paired improbable event (e.g., the child's dad taking him to a park all day instead of to school). Despite this pattern in adults' judgments about the actual likelihood of these events, children more frequently judged immoral events to require magic than improbable events.

4.1.2 Accounting for the data

In sum, we find a developmental pattern such that young children regard events involving physical violations, morally bad actions or statistically improbable occurrences to be impossible. Then, as children age, they increasingly tend to judge that only events involving physical violations are actually impossible. The question now is how to make sense of young children's puzzling judgments and the changes that occur as they age.

First consider how one would need to explain these data on the separate representations view. This view holds that there are just three separate representations (physical, moral, probabilistic). It then has to explain why children use all three of these representations when making judgments of possibility or magic. On the face of it, there certainly seems to be something perplexing about fact that children are answering questions about magic by relying on their moral representations of the events. Accordingly, the separate representations view needs some way of making sense of this pattern.

While it is not completely clear how one ought to go about making sense of the pattern, one approach would be to argue that morality has a more indirect effect. Specifically, it could be that there is some causal link between children's moral representation and their representations of what is physically possible. It may be, for example, that children believe that morally bad actions involve some sort of violation of physics, and they are simply relying on their physical representations when answering questions about possibility. (Something similar would have to be said about their representations of probability.) On this approach, then, the puzzle that the separate representations view ends up facing is one of explaining why young children regard events involving actions that are morally bad (or improbable) as involving physical violations.

The trouble here is that there is currently no other research that would support the necessary connections between these independent representations (physical, moral, etc.). Moreover, even if one were able to find support for these connections, one would still need to explain why the special connections that were posited to exist between children's independent representations are subsequently eliminated over the course of development, leaving adults with only the comparatively simple and disconnected representations.

Setting this approach to one side, let's now consider how the modal representation view would explain these developmental patterns. On the modal representation view, it is not that children have an additional set of connections that adults lack. Rather, children's responses are simply reflecting the most basic functioning of the psychological representation of modality.

According to the modal representation view, possibilities that are violations of physics, morally bad, or statistically improbable tend to be excluded from the domain by default. In the absence of intervening factors, this is the pattern we should expect, and is the pattern observed in young children's judgments. Thus, to explain the pattern of children's responses, we need only assume that the children correctly understood that the question they were being asked was one that was, at heart, a question about modality.

When we considered the separate representations view, it seemed surprising that children should think it is impossible to find an alligator under one's bed, and a defender of this view would therefore have to invoke some special additional process to explain the patterns observed in children's judgments. Now, in switching to the modal representation view, we face the the opposite problem. It becomes extremely easy to explain children's judgments, but a puzzle arises as to why the *adults* give the responses they do. Given that it is extraordinarily improbable to find an alligator under one's bed (and that the possibility should therefore be regarded as ruled out), why do adults say that such things are possible? To explain this kind of response, a defender of the modal representation view will need to say something about the additional sophisticated capacity that adults have developed.

One likely explanation is that the ability that adults have developed is one that allows them to prevent certain factors from constraining the possibilities that are included in the domain. For example, they may have realized that the question they are being asked is specifically about which events involve physical violations, and accordingly prevented other constrains (moral, statistical, and so on) from playing much of a role. In other words, it may be that modal reasoning defaults to taking into account a variety of different considerations (physical, moral, probabilistic), but adults have developed a capacity for a more sophisticated kind of reasoning that allows them to deviate from this natural default.

Recent research actually provides a test of this hypothesis by comparing adults' judgments of possibility when they reflectively deliberate to their judgments of possibility when they are forced to respond extremely quickly (and thus are less able to engage in any kind of sophisticated or effortful reasoning). When adults are unable to engage in sophisticated reasoning, their judgments of what is possible begin to strongly resemble those of young children (Phillips & Cushman 2016).

4.2 Freedom

Philosophers have long been concerned with determining when one acts freely and when one is instead forced to act. Such questions are at the core of a number of debates in areas ranging from political philosophy to metaphysics (Aquinas [1273]1952, Aristotle [340 BCE]2002, Descartes [1641]1984, Hume [1748]2007, Locke [1690]1975). In these philosophical discussions, as well as in in ordinary intuitive judgments about freedom, one finds that physics and morality play a similar role in shaping whether or not one is forced to do a particular action. Just as physical considerations limit which actions an agent is perceived as being able to pursue, moral considerations also constrains which actions are seen as available to that agent.

4.2.1 The data

The clearest examples in which an agent lacks freedom are those in which an agent is literally *physically incapable* of doing anything else. Consider, for instance, Locke's [1690]1975 classic example of a prisoner locked in his cell. As Locke suggests, we can make sense of the judgment that the prisoner does not freely stay in his cell by appealing the fact that he is not physically capable of leaving. Unsurprisingly, empirical studies consistently show that people's freedom judgments are indeed impacted by their judgments of whether an action would require physical violations to be done (see, e.g., Woolfolk et al. 2006).

Intriguingly, the philosophical discussion has also suggested that whether an agent is forced to act may also depend in some way on *morality*. Aristotle [340 BCE]2002, for example, develops an account of free action in which moral considerations play an essential role. Recent research in experimental philosophy has suggested that Aristotle's suggestion may actually capture a central aspect of the way that people ordinarily make judgments about force and freedom. A number of recent studies have demonstrated an impact of moral judgments on freedom judgments (Chakroff & Young 2015, Young & Phillips 2011, Phillips & Knobe 2009). In one such study (Young & Phillips 2011), participants in one condition were assigned to read the following vignette (adapted from Aristotle *NE* 1110^{a}8-9):

> While sailing on the sea, a large storm came upon a captain and his ship. As the waves began to grow larger, the captain realized that his small vessel was too heavy and the ship would flood if he didn't make it lighter. The only way that the captain could keep the ship from capsizing was to throw his expensive cargo overboard. Thinking quickly, the captain ordered one of his sailors to throw the cargo overboard. While the cargo sank to the bottom of the sea, the captain was able to survive the storm and returned home safely.

After reading, participants were asked indicate whether they thought that the captain was forced to throw his cargo overboard. The other half of participants read a vignette that was identical except that 'cargo' was replaced by 'passengers', making the captain's action morally bad. Despite the similarity in situation the captain faced in the two vignettes, participants only agreed that the captain was forced to throw the cargo overboard, and strongly disagreed that the captain was forced to throw the passengers overboard (Young & Phillips 2011). Additional studies have demonstrated in a number of different vignettes that this change in participants' judgments is driven specifically by the change in the moral status of the agent's actions (Chakroff & Young 2015, Young & Phillips 2011, Phillips & Knobe 2009).

4.2.2 Accounting for the data

Overall, then, existing results indicate that people's intuitions about freedom are impacted both by physical considerations and by moral considerations. The question now is how to explain these two effects.

First, consider the separate representations view. On this view, people simply have a number of distinct representations (physical, moral, etc.), and a question arises as to why these different representations each impact people's freedom judgments. If one starts out trying to explain the impact of the physical representation, an obvious first step would be to propose a principle that goes something like this:

> For people to conclude that an agent performed an action freely, they have to think that the agent was *physically capable* of not performing the action.

This principle does seem like a plausible one, but notice that it would not help at all in making sense of the role of moral considerations. Thus, to explain the role of moral considerations, we would have to introduce some completely separate sort of psychological mechanism.

Now suppose we turn to the modal representation view. We would then be assuming that people have a single representation (the representation of modality) and that this representation is influenced by both physical and moral considerations. So we could reformulate our principle in such a way that it includes an explicitly modal notion:

> For people to conclude that an agent performed an action freely, they have to think that it was *possible* for the agent not to have performed that action.

The key change is that we are now framing the principle in terms of which things people regard as 'possible'. Thus, the prediction is that people's freedom judgments will not be determined entirely by people's physical representations. Instead, they should be influenced by the whole variety of different considerations that impact people's psychological representation of modality.

This principle then allows us to provide a unified explanation for the two effects. People's representation of modality can be influenced by physical considerations or by moral considerations, but either way, this representation has the same basic impact on freedom judgments. People consider which sorts of possibilities fall within the domain and which are ruled out. To the extent that they find that all other possible actions were ruled out, they tend to conclude that she did not act freely.

Recent experimental results have provided support for this explanatory hypothesis. Focusing on the case of the ship captain, subsequent studies have shown manipulating the moral status of the action affects intuitions about alternative possibilities (Knobe & Szabó 2013) and that these intuitions about alternative possibilities mediate the impact of moral status on freedom judgments (Phillips et al. 2015). In short, existing data seem to support the claim that the impact of moral considerations on freedom judgments arises because of a broader fact about the impact of moral considerations on people's representation of modality.

4.3 Causal selection

Consider how many potential causes there are for each ordinary event that occurs in our everyday lives. How is it, for example, that we decide that the children running around the antique shop were the cause of the expensive vase breaking, not the precarious placement of the vase, or the fragile material the vase was made of, or the hard floor the vase fell onto, or the parents' decision to bring their children to the antique store, or the shopkeeper's choice of displaying that vase rather than some other vase, or...? The problem here is pervasive: for any given event, we could potentially trace the cause of it back to any one of the myriad contributing factors, and also to the factors that contributed to those factors' contributions, and so on and so forth, indefinitely. Yet, despite the inherent difficulty of the problem of causal selection, we often find such tasks surprisingly easy. So how are we making these decisions? While the process is unquestionably complicated, two factors that are widely known to play a role in people's selection of causes are probability and morality.

4.3.1 The data

To see the role that *probability* plays in causal selection, let us return to the simple example of the forest fire that is started in the presence of oxygen, dry leaves and a lit match. Why is it that we tend not to select the presence of oxygen as the cause of the forest fire, but do select the lit match? The key here is to notice how *improbable* it is that oxygen would have not been present in the forest where the fire is started. By contrast, it is highly *probable* that there would not have been a lit match in the forest. Between these three potential causes, then, we can make sense of why we select the match as a cause by appealing to the difference in the probability of these events.

Similarly, one can see probability playing a role in classic attribution studies (Kelley 1967, 1973, Frieze & Weiner 1971, McArthur 1972). In one study, a set of participants were told that a man named John laughed at a comedian, and in addition were further information that:

> Almost everyone who hears the comedian laughs at him.

or that:

> Hardly anyone who hears the comedian laughs at him.

In both cases, participants are asked to determine what caused John to laugh. Was it something about the comedian, something about John himself, or something about both of them? Participants tended to only select John as a cause in the second case, when hardly anyone who hears the comedian laughs (McArthur 1972). The pattern here appears to be notably similar to the one observed in the case of the forest fire. In the first case, it is highly *improbable* that John would not have a trait that made him laugh at the comedian, whereas in the second case, it is highly probable that he would not have such a trait. Here again, we see probability judgments impacting causal judgments.

Much more recently, *morality* has also been shown to play a similar role in how people select the causes of an event (Alicke 2000, 1992, Alicke et al. 2011, Knobe & Fraser 2008, Hitchcock & Knobe 2009, Roxborough & Cumby 2009, Kominsky et al. 2015). For example, in one study (Knobe & Fraser 2008), participants were presented with the following vignette:

> The receptionist in the philosophy department keeps her desk stocked with pens. The administrative assistants are allowed to take pens, but faculty members are supposed to buy their own. The administrative assistants typically do take the pens. Unfortunately, so do the faculty members. The receptionist repeatedly e-mailed them reminders that only administrators are allowed to take the pens.
>
> On Monday morning, one of the administrative assistants encounters Professor Smith walking past the receptionist's desk. Both take pens. Later, that day, the receptionist needs to take an important message... but she has a problem. There are no pens left on her desk.

After reading the vignette, participants were asked to rate their agreement both with a statement that said Professor Smith caused the problem and with a statement that said the administrative assistant caused the problem. While the problem would not have arisen if either Professor Smith or the administrative assistant had not taken a pen, participants selected that only Professor Smith was a cause of the problem. Subsequent studies have also demonstrated that it is specifically the moral badness of Professor Smith's action (not merely the probability of the action) that leads participants to select him as the cause of the problem. In one study, for example, participants were told that the professor always takes the pens despite not being allowed to, while the administrative assistants never do (Roxborough & Cumby 2009). Even in this modified case, participants still selected Professor Smith as the cause of the problem despite the fact that it was extremely probable that he would take the pen, suggesting that the effect of immorality can occur independently of the effects of probability.

Thus, while we find that the effect of morality and probability are distinct, we also notice that the effect of an event being improbable and the effect of an event being immoral are quite similar. In fact, recent studies have looked at the exact patterns of these two effects and have provided evidence for their remarkable similarity, even down to the precise conditions under which such effects arise (Kominsky et al. 2015). The remaining question is why these two factors affect causal selection in precisely the same way.

4.3.2 Accounting for the data

One way of accounting for these results would be to assume, in line with the separate representations view, that probability and morality have completely separate effects on causal selection. This has been one traditional approach, with a number of researchers developing accounts specifically meant to handle the impact of probability (Kelley 1967, 1973), and others developing accounts specifically meant to explain the impact of morality (Alicke et al. 2011, Samland & Waldmann 2016, Sytsma et al. 2012). While this approach has demonstrated that there are many examples which are consistent with these specific accounts, such an approach does not offer a unified picture on which to understand both the impact of probability and morality. Moreover, these individual approaches simply have no way of explaining the fact that probability and morality affect causal selection in precisely the same way, since these accounts were specifically developed to explain *only* the impact of probability or *only* the impact of morality.

In contrast to this approach (and in line with the modal representation view), a number of researchers have developed frameworks that account for the effect of morality and probability in a unified way (Bello 2014, Blanchard & Schaffer 2017, Halpern & Hitchcock 2015, Icard et al. 2017, Knobe & Szabó 2013). While different accounts differ in the formalism they use to capture these effects, there are two central features that they share. First, they all emphasize the critical importance of considering possibilities that differ from what actually happened (e.g., possibilities in which there is not a lit match, and thus no forest fire). Second, they all include some specific account of how particular possibilities are the ones that are selected to play a role in causal selection. To see how this works, we can abstract away from the particulars of any one formal account, and apply the basic suggestions these researchers have proposed to the framework we've been employing all along.

At the heart of all of these accounts is the idea that causal reasoning in some way involves thinking about alternative possibilities. This basic view has been spelled out in quite different ways within different theoretical frameworks (Lewis 1973, Lombrozo 2010, Pearl 2000, Schaffer 2005), but the differences between those frameworks will not concern us here. Instead, we will be relying on a core claim that is shared by all of the frameworks: namely, that the judgment that factor *x* was the cause of an outcome in some way involves thinking about possibilities in which factor *x* differs in some way.

Returning to the example of the forest fire, we now want to explain why the match, but not oxygen is selected as a cause. The basic suggestion is that the causal judgment (7-a) involves representing possibilities picked out by (7-b), while (8-a) involves representing possibilities picked out by (8-b).

(7) a. The lit match was the cause of the forest fire.
b. Possibilities in which the match was not lit.

(8) a. The oxygen was the cause of the forest fire.
b. Possibilities in which there was no oxygen

The key difference between possibilities (7-b) and (8-b) is that while it that while it was highly probable that there could have *not* been a lit match in forest, it is highly improbable that there could have not been oxygen in the forest. Accordingly, possibilities like (7-b) should be included in the domain, while possibilities like (8-b) should not be. It is because of this difference in how people represent these possibilities that they see the lit match, but not the oxygen, as a cause of the forest fire.

In precisely the same way, we can also make sense of why the professor is selected as a cause of the problem that arose. In this case, the basic suggestion is that the causal judgment (9-a) involves representing possibilities picked out by (9-b), while (10-a) involves representing possibilities picked out by (10-b).

(9) a. The professor caused the problem.
b. Possibilities in which the professor did not take a pen.

(10) a. The administrative assistant caused the problem.
b. Possibilities in which the administrative assistant did not take a pen.

In this case, the key difference between (9-b) and (10-b) is that while it would have been morally good for the professor to have *not* taken a pen, it would not have correspondingly been morally good for the administrative assistant to have not taken a pen. Thus, possibilities like (9-b) should be included in the domain, whereas possibilities like (10-b) need not be. Once again, this difference in how we represent possibilities leads to a corresponding difference in causal judgment: people judge that the professor, but not the administrative assistant, caused the problem.

Recent studies have also provided direct evidence for the role of the relevance of these alternative possibilities in participants' causal judgments (Phillips et al. 2015). Participants were asked to complete a continuous measure of the relevance of the alternative possibility in which the professor (or the administrative assistant) did *not* take a pen. Participants found the possibility that the professor didn't take a pen to be much relevant than the possibility that the administrative assistant did not take a pen, and more importantly, these judgments of the relevance of alternative possibilities mediated the effect of morality on causal selection.

In one sense, the proposal we have made in this section is very similar to previously offered accounts of the effect of morality and probability on causal selection (Bello 2014, Blanchard & Schaffer 2017, Halpern & Hitchcock 2015, Icard et al. 2017, Knobe & Szabó 2013). Just like this previous research, we've argued that these effects are best explained by arguing (1) that casual selection involves some way of representing possibilities, and (2) that morality and probability affect which possibilities play a role in causal selection (in our framework, which possibilities are in the domain, and which are ruled out). Yet, in another sense, our suggestion goes beyond previous research. Unlike previous accounts, the current proposal suggests that these effects should not be thought of as something to be explained by any theory that is specific to *causation*. Rather, the effects of morality and probability on causal selection should be understood as arising from much more general features of the way that we represent possibilities.

The basic proposal that we've been making throughout is that the psychological representation of modality works in a relatively fixed way that is insensitive to the particular task that people are engaged in. In general, the psychological representation of modality tends not to include possibilities in the domain when they are morally bad or statistically improbable or violations of physics, and so we should expect these considerations to play a role even in cases where, upon reflection, many of these considerations may seem completely irrelevant to the specific question at hand. Applying this to the case at hand, then, the modal representation view suggests that there is nothing particular to causation that we need to explain why it is impacted by factors like morality and probability. All we need is the relatively uncontroversial suggestion that causal cognition involves the representation of possibilities, and thus relies on the psychological representation of modality.

4.4 Explicit counterfactual reasoning

In various ways, all of the phenomena we have been discussing appear to involve a capacity for considering alternative possibilities. We now want to take up the phenomenon in which that capacity is most clearly and conspicuously manifest. Specifically, we will be discussing the conscious, controlled process of *explicit counterfactual reasoning*. Research has shown that explicit counterfactual reasoning impacts numerous aspects of people's lives, from psychological well-being to victim-blaming (Gilovich & Medvec 1995, Markman & Miller 2006, Epstude & Roese 2008, Branscombe et al. 2003), and it has therefore been investigated in great detail within the existing literature.

One of the central questions in the literature on explicit counterfactual reasoning concerns which counterfactual possibilities people tend to consider. Standardly, the tasks used to study this question involve presenting participants with a series of connected events that result in a given outcome and then asking participants to generate counterfactual possibilities in which the outcome would not have occurred. To get a sense for this, consider an example offered originally by Kahneman and Tversky 1982. After leaving his office, Mr. Jones decided to drive home by a scenic route, which he rarely gets to take. On the drive, he went going through an intersection after the light changed to green and was struck and killed by a teenage driver who was on drugs. Participants were then told that during the days following the accident, the Jones family 'often thought and often said, "If only..."'. Participants were then instructed to provide one or more completions of this thought.

Researchers have made a great deal of progress in exploring the factors that influence explicit counterfactual reasoning (for reviews, see Byrne 2016; Epstude & Roese 2008; Roese 1997. It appears that people show a strong tendency not to engage in such reasoning about possibilities that involve physical violations, improbable events, or morally bad actions.

4.4.1 The data

People tend to largely converge on the events that they elect to change in constructing such counterfactual possibilities. Part of this convergence arises because there is a general tendency to change events that are *statistically improbable* by replacing them with events that are more *probable*, and thereby prevent the outcome from occurring. In the case of Mr. Jones, for example, people tend to complete the Jones family's thoughts, 'If only Mr. Jones had taken his ordinary route instead of the scenic route home' (Kahneman & Tversky 1982). In contrast, people rarely undo an outcome by replacing a statistically probable event with one that is highly improbable (Wells et al. 1987). Similarly, people almost never undo an outcome by replacing an event that does not involve a physical violations with one that does, e.g., ‘If only Mr. Jones’s car had levitated moments before the accident...’ (Roese 1997).

Subsequent work on this question has also revealed that people exhibit a tendency to undo an outcome by replacing *morally bad* events with events that are *morally neutral* or *morally good*, but rarely undo an outcome by replacing morally neutral or good events with ones that are morally bad (McCloy & Byrne 2000a, N’gbala & Branscombe 1995). In one study, for example, participants read about a father named Joe who was late to pick up his son from school either because he stopped to help someone who was injured or because he was negligent and wanted talk to his friends instead of picking up his son. In both cases, when his father didn’t arrive, Joe’s son accepted a ride home from a neighbor, and was killed in a car accident on the drive home. While it is always true that Joe’s son would not have died if Joe had arrived on time, participants were strongly inclined to consider what would have happened if Joe hadn’t acted negligently, but not what would have happened if Joe hadn’t stopped to help the injured person (N’gbala & Branscombe 1995).

4.4.2 Accounting for the data

The modal representation view provides a straightforward explanation of this pattern of data. The key idea is just that there is a tendency for people only to engage in explicit counterfactual reasoning with regard to the possibilities that are represented as being in the domain. Possibilities that are either improbable or morally bad are ruled out, and people tend not to engage in explicit counterfactual reasoning about them. Thus, the modal representation view provides a unified explanation that applies to both of these effects.

As we noted at the outset, it is important to acknowledge that the modal representation view makes only a modest contribution to the larger problem of understanding counterfactual reasoning. First, we are proposing that people show a tendency not to engage in counterfactual thinking regarding possibilities that are ruled out, but this is only a general tendency, not a hard-and-fast rule. For example, consider possibilities involving improbable occurrences such as a legion of miniature hogs parachuting from the sky during a wedding. The theory we’ve been developing says that these possibilities are regarded as ruled out, and we therefore predict a general tendency for people not to engage in explicit counterfactual reasoning about them. But now suppose that we conducted a study in which participants were explicitly instructed: ‘Please write a paragraph about what would have happened if a legion of miniature hogs had parachuted from the sky during a wedding.’ Though participants would presumably continue to regard this possibility as completely ruled out, they could easily proceed to engage in explicit counterfactual reasoning about it.

Second, and more importantly, there will typically be an enormous number of different possibilities within the domain, but people can only engage in explicit counterfactual reasoning about a tiny fraction of them. Hence, one needs some explanation for the systematic effects whereby people tend to engage in explicit counterfactual reasoning about some of these possibilities and not others. To give just one example, research has consistently found that people are more likely to engage in explicit counterfactual reasoning about factors that an agent can *control* (e.g., Davis et al. 1995; Girotto et al. 1991). These sorts of effects are not themselves explained by the modal representation view. Presumably, they are to be explained in terms of the complex array of other psychological processes that have already been explored within the existing literature on explicit counterfactual reasoning (for reviews, see Byrne 2016, Epstude & Roese 2008, Roese 1997).

Still, although the present hypothesis cannot explain all the effects observed in people's explicit counterfactual reasoning, it does help to address a puzzle that has arisen within the existing literature. Specifically, it helps us to explain the similarity one finds between the patterns observed in explicit counterfactual reasoning and the patterns observed in other aspects of cognition. The explanation, we suggest, is that a number of different aspects of cognition are influenced by people's representation of modality.

As an example, consider the relationship between explicit counterfactual reasoning and causal selection. Many of the effects observed for explicit counterfactual reasoning can also be found for causal selection, and it has therefore been suggested that people's counterfactual reasoning influences their causal selection (Hilton 1990, Kahneman & Tversky 1982, Roese & Olson 1994, Wells & Gavanski 1989). However, there are also certain effects that are observed in explicit counterfactual reasoning but not in causal selection. In particular, explicit counterfactual reasoning appears to be influenced by controllability in a way that causal selection judgments are not (Mandel & Lehman 1996). Thus, we seem to be left with a mystery. Why is it that the pattern in people's explicit counterfactual reasoning is in some ways similar to the pattern in causal selection but also in some ways quite different?

The modal representation view provides a straightforward way of resolving this mystery. The claim is that the similarities are not due to a direct connection between the pattern in people's explicit counterfactual reasoning and the pattern in causal selection but rather to the influence of a third variable. Explicit counterfactual reasoning and causal selection are each influenced by a number of different factors, but they are also both influenced by the psychological representation of modality.

4.5 Natural language modals

We now want to briefly return to the phenomenon from which we originally departed. We began by looking at research that aimed to capture the meanings of certain linguistic expressions (‘can’, ‘must’, ‘have to’, etc.). At this point, it may be helpful to return to that linguistic research and examine its relation to the account we have been developing here.

Recall that people’s use of the relevant linguistic expressions can be influenced by a whole host of different considerations, including both physical considerations and moral considerations. We illustrated the roles of physics and morality using sentences like (1) and (2).

(1) Particles can’t go faster than the speed of light.

(2) You can’t keep treating her that way – look at how upset she is!

Work in formal semantics has led to the development of a formal model that makes it possible to offer a unified explanation of the influence of these various sorts of considerations.

We have been drawing on insights from that model throughout the present inquiry. Specifically, we took the model developed within formal semantics and used it as the basis for a hypothesis about a particular sort of psychological representation. We then argued that this hypothesis could help to explain a number of otherwise puzzling phenomena.

A question now arises as to whether the psychological representation we have been discussing also governs people’s use of the relevant linguistic expressions. A full-scale investigation of this question would go beyond the scope of the present paper. Still, we want to suggest that there is at least some preliminary reason to think that this approach is a viable one.

At the core of the models we reviewed from formal semantics is the idea of a domain of possibilities. Presumably, if people use these expressions in the way described by the models, they have some representation of this domain. But what exactly is the role of that representation in their cognition more broadly? At least initially, one might suppose that it serves only to govern their use of these specific linguistic expressions and doesn’t play any role in other aspects of people’s behavior.

However, the present inquiry opens up the possibility of a very different approach to understanding these phenomena. We have argued that people have a representation of a domain of possibilities that influences numerous aspects of their behavior. An obvious suggestion, then, would be that it is this very same representation that governs people’s use of modal expressions (‘can’, ‘must’, ‘have to’, etc.). In essence, the proposal is quite straightforward: the psychological representation governing people’s use of modal expressions simply is the psychological representation of modality.

Assuming that this suggestion turns out to be correct, the role of linguistic facts in our argument as a whole will be quite circumscribed. In arguing for our hypothesis, we drew heavily on insights from linguistics, but the linguistic facts do not occupy any special position in the hypothesis itself. Rather, the claim is that research in linguistics is pointing to something very general about how human beings understand possibilities. Once we understand this more general fact about human cognition, we can articulate a view in which the linguistic phenomena are explained in terms of precisely the same representation that explains all of the others (see Figure 3).

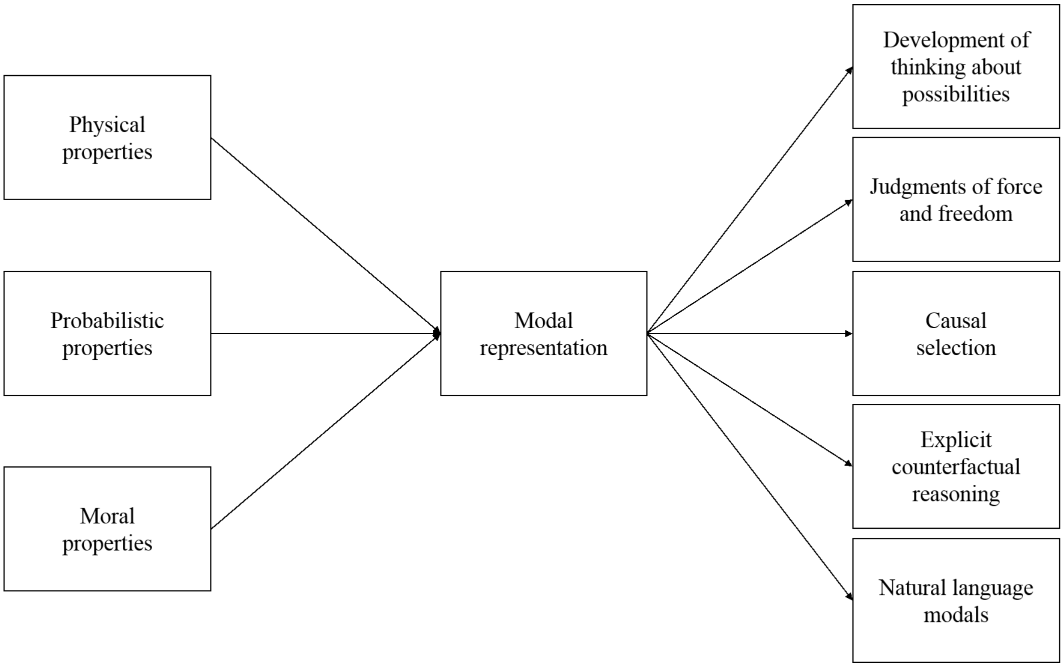

Figure 3: Schematic model of the relationship between the modal representation view and the linguistic expression of modality, among others

5 The cognitive science of modality

Our focus throughout this paper has been on certain specific phenomena that have already received extensive attention in the existing literature (causal selection, judgments of freedom, etc.). However, if the present account turns out to be on the right track, it points toward a promising new topic that has yet to be explored in depth. Independent of anything about the study of these specific phenomena, future research in cognitive science could take up questions about the psychological representation of modality as issues to be investigated in their own right.

To illustrate the potential for such research, we briefly consider two questions about the psychological representation of modality that would be worthy of further investigation.

1. One obvious question is whether this modal representation is dichotomous, such that any given possibility is represented as either inside or outside of the domain, or whether the representation is actually more graded. Consider again an agent who is trying to get to the airport but whose car has broken down. As we have argued, there is an important sense in which people represent the possibility of convincing the airport to delay the flight as falling outside of the domain. At the same time, though, it may seem that people represent the possibility of levitating and flying to the airport as even farther outside the domain, such that this possibility can be understood as being somehow *more* impossible. A key question is how we ought to account for this difference.

One hypothesis is that the psychological representation of modality directly represents this kind of gradability, with some possibilities being represented as falling more within the domain than others. Approaches along these lines have been explored within existing work in formal semantics (e.g., Lassiter 2011), and it is certainly plausible that such an approach will prove helpful here as well. However, one downside to this hypothesis is that it would seem to require an extremely complex representation of modality. For each possibility, even those that seem highly far-fetched and not worth considering, people would have to represent the degree to which that possibility fell outside the domain.

A promising alternative approach would be to invoke the notion of *probabilistic sampling* (e.g., Icard 2016, Vul et al. 2014. The core idea behind this approach would be that there is a probabilistic process that determines which possibilities are represented as falling within the domain on any given occasion. Some possibilities have a high probability of being represented in the domain, others a much lower probability. Then, on any given occasion, people sample certain possibilities from this distribution. Existing work within this approach has led to the development of models according to which the probability distribution from which possibilities are sampled can be shaped by physical, moral and probabilistic considerations (Icard et al. 2017).

On this sampling approach, each possibility would be represented dichotomously as either inside or outside the domain. The appearance of gradability would then arise because some possibilities would have a higher probability of being represented inside the domain than others. Thus, in a case where you might be drawn intuitively to say that some possibilities fall even farther outside the domain than others, a more accurate description would be that some possibilities have an even lower probability of being represented as inside the domain than others.

2. A question arises about the relationship between the representations of the various separate considerations (physical, moral, probabilistic) and the representation of modality itself. We have suggested that the representation of modality is influenced by all of these different considerations, but how should we understand the psychological process underlying this influence?

One possible view would be that the process is to be understood in terms of a series of distinct stages. First, people arrive at representations of each of the separate considerations – a physical representation, a moral representation, a probabilistic representation. Then, in a subsequent stage, people integrate these various representations in forming a unified representation of modality.

This view is certainly a plausible one, and it may ultimately turn out to be correct, but all the same, there is at least some reason to think that the relationship might not be so simple. To take one example, determining whether or not a particular action was morally wrong often requires representing other actions that could have been taken (a modal representation). Moreover, there seem to be moral concepts that specifically encode modal information (e.g., the concept *reckless*), suggesting that there may be moral representations that simply cannot exist completely independently of all modal representations.[3] Thus, there may be some reason to question a picture on which people first compute representations of each of the separate considerations and only then compute a representation of modality.

Future research should aim to address this issue by directly exploring the relationship between the representations of these various considerations (physical, moral, statistical) and the representation of modality itself. One hypothesis would be that people actually do have representations of these separate considerations that are completely independent of modality. For example, in the case of morality, people clearly have a complex representation of moral value that depends on modality, but the existing research suggests that people also have a simpler representation of moral value that does not depend on modality (see, e. g., Cushman 2013, on model-free reinforcement learning of moral value). Thus, one approach to addressing this issue would be to suggest that people go through a series of stages in which they first compute some simple moral representation, which then influences the representation of modality, which then serves in turn as a basis for a more complex moral representation.

A more radical approach, however, would be to try to address the issue by reconceptualizing the representations of the various considerations (physical, moral, probabilistic). As we have seen throughout the present paper, it becomes possible to explain a number of surprising phenomena if we posit a single unified representation of modality that is shaped by all of these considerations. The problem we are raising here arises from the fact that our account still retains the assumption that people have a representation of moral considerations that is completely independent of this unified representation. Thus an alternative approach would be to try to resolve this

[3] We thank an anonymous referee at *Mind & Language* for this example.

problem by moving even farther in the direction we have been exploring here. On this alternative, one would reject the assumption that people have a representation of moral considerations that is completely independent of modality. In its place, one would develop an account according to which people's representation of moral considerations was in some way fundamentally intertwined with the representation of modality from the very beginning.

The questions we have just been considering, about the representation of the domain and the considerations that constrain it, are only two of the many questions that arise when one begins looking at the psychological representation of modality as a phenomenon worth exploring in its own right. The answers to these questions will be informative at multiple levels. Most directly, these answers will give us some insight into the psychological representation of modality itself. Just as importantly though, because of the central role that this representation plays throughout cognition, these answers will illuminate each of the many different phenomena that rely on the psychological representation of modality.

Acknowledgements: We would like to gratefully acknowledge Angelika Kratzer, Fiery Cushman, Tobias Gerstenberg, Paul Boom, Jonathan Schaffer, Brent Strickland, Charles Kalish, Paul Harris, Laurie Santos, John Turri, Liane Young, Dave Rand, Justin Martin, Adam Morris and members of the Moral Psychology Research Group for helpful comments and discussion. JP would also like to thank the SHAME writing group.

References

Alicke, M. D. 2000. Culpable control and the psychology of blame. *Psychological bulletin* 126. 556–574. http://dx.doi.org/10.1037/0033-2909.126.4.556.

Alicke, M. D., David Rose & Dori Bloom. 2011. Causation, norm violation, and culpable control. *Journal of Philosophy* 108(12). 670–696.

Alicke, Mark D. 1992. Culpable causation. *Journal of personality and social psychology* 63(3). 368.

Aquinas, Thomas. [1273]1952. *The summa theologica of saint thomas aquinas*. Encyclopedia Britannica.

Aristotle. [340 BCE]2002. *Nicomachean ethics*. Oxford University Press.

Bello, Paul F. 2014. Mechanizing modal psychology. In *Ethics in science, technology and engineering, 2014 ieee international symposium on*, 1–8. IEEE.

Berlin, Isaiah. 1970. *Four essays on liberty*, vol. 969. Oxford University Press New York.

Björnsson, Gunnar & Stephen Finlay. 2010. Metaethical contextualism defended. *Ethics* 121(1). 7–36.

Blanchard, Thomas & Jonathan Schaffer. 2017. Cause without default. In Helen Beebee, Christopher Hitchcock & Huw Price (eds.), *Making a difference*, Oxford University Press.

Branscombe, Nyla R, Michael J A Wohl, Susan Owen, Julie A Allison & N Ahogni. 2003. Counterfactual thinking, blame assignment, and well-being in rape victims. *Basic and Applied Social Psychology* 25(4). 265–273. http://dx.doi.org/10.1207/S15324834BASP2504.

Byrne, Ruth MJ. 2016. Counterfactual thought: From conditional reasoning to moral judgment. *Annual review of psychology* 67(1).

Cariani, Fabrizio. 2013. 'ought' and resolution semantics. *Noûs* 47(3). 534–558.

Chakroff, Alek & Liane Young. 2015. Harmful situations, impure people: An attribution asymmetry across moral domains. *Cognition* 136. 30–37.

Chernyak, Nadia, Tamar Kushnir, Katherine M Sullivan & Qi Wang. 2013. A comparison of american and nepalese children's concepts of freedom of choice and social constraint. *Cognitive science* 37(7). 1343–1355.

Cushman, Fiery. 2008. Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition* 108(2). 353–380.

Cushman, Fiery. 2013. Action, outcome, and value: a dual-system framework for morality. *Personality and social psychology review : an official journal of the Society for Personality and Social Psychology, Inc* 17(3). 273–92. http://dx.doi.org/10.1177/1088868313495594. http://www.ncbi.nlm.nih.gov/pubmed/23861355.

Davis, C. G., D. R. Lehman, C. B. Wortman, R. C. Silver & S. C. Thompson. 1995. The undoing of traumatic life events. http://dx.doi.org/10.1177/0146167295212002.

De Brigard, Felipe, Donna R Addis, Jaclyn H Ford, Daniel L Schacter & Kelly S Giovanello. 2013. Remembering what could have happened: Neural correlates of episodic counterfactual thinking. *Neuropsychologia* 51(12). 2401–2414.

Descartes, René. [1641]1984. *The philosophical writings of Descartes*, vol. II. Cambridge University Press. http://dx.doi.org/10.1016/0191-6599(88)90158-1.

Dowell, Janice J. L. 2011. A flexible contextualist account of epistemic modals. *Philosophers' Imprint* 11(14). 1–25.

Egan, Andy, John Hawthorne & Brian Weatherson. 2005. Epistemic modals in context. In G. Preyer & G. Peter (eds.), *Contextualism in philosophy*, 131–170. Oxford University Press.

Epstude, Kai & Neal J. Roese. 2008. Mental Simulation and Causal Attribution: When Simulating an Event Does Not Affect Fault Assignment. *Personality and Social Psychology Review* 12(2). 168–92. http://dx.doi.org/10.1177/1088868308316091. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2408534{&}tool=pmcentrez{&}rendertype=abstract.

von Fintel, Kai. 2012. The best we can (expect to) get? challenges to the classic semantics for deontic modals. In *Central meeting of the american philosophical association, february*, vol. 17, .

Frieze, Irene & Bernard Weiner. 1971. Cue utilization and attributional judgments for success and failure. *Journal of Personality* 39(4). 591–605. http://dx.doi.org/10.1111/j.1467-6494.1971.tb00065.x.

Gerstenberg, Tobias & Joshua B. Tenenbaum. in press. Intuitive theories. In Michael Waldmannn (ed.), *Oxford handbook of causal reasoning*, Oxford University Press.

Gilovich, T & V H Medvec. 1995. The experience of regret: what, when, and why. *Psychological review* 102. 379–395. http://dx.doi.org/10.1037/0033-295X.102.2.379.

Girotto, Vittorio, Paolo Legrenzi & Antonio Rizzo. 1991. Event controllability in counterfactual thinking. http://dx.doi.org/10.1016/0001-6918(91)90007-M.

Gopnik, Alison, Clark Glymour, David M Sobel, Laura E Schulz, Tamar Kushnir & David Danks. 2004. A theory of causal learning in children: causal maps and bayes nets. *Psychological review* 111(1). 3.

Griffiths, Thomas L, Nick Chater, Charles Kemp, Amy Perfors & Joshua B Tenenbaum. 2010. Probabilistic models of cognition: Exploring representations and inductive biases. *Trends in cognitive sciences* 14(8). 357–364.

Halpern, Joseph Y. 2003. *Reasoning about uncertainty*. MIT Press.

Halpern, Joseph Y & Christopher Hitchcock. 2015. Graded causation and defaults. *The British Journal for the Philosophy of Science* 66(2). 413–457.

Halpern, Joseph Y & Judea Pearl. 2005. Causes and expanations: A structural-model approach. part i: Causes. *The British journal for the philosophy of science* 56(4). 843–887.

Hart, Herbert Lionel Adolphus & Tony Honoré. 1985. *Causation in the law*. Oxford University Press.

Hilton, Denis J. 1990. Conversational processes and causal explanation. http://dx.doi.org/10.1037/0033-2909.107.1.65.

Hitchcock, Christopher & Joshua Knobe. 2009. Cause and norm. *The Journal of Philosophy* 106(11). 587–612.

Hume, David. [1748]2007. *An enquiry concerning human understanding*. Broadview Press.

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

Icard, Thomas, Jonathan Kominsky & Joshua Knobe. 2017. Normality and actual causal strength. *Cognition* 161. 80–93.

Kahneman, Daniel & Dale T Miller. 1986. Norm theory: Comparing reality to its alternatives. *Psychological review* 93(2). 136.

Kahneman, Daniel & Amos Tversky. 1982. The simulation heuristic. In Daniel Kahneman, Paul Slovic & Amos Tversky (eds.), *Judgment under uncertainty: Heuristics and biases*, vol. 185, 1124–1131. Cambridge University Press. http://dx.doi.org/10.1093/oxfordhb/9780195376746.013.0038.

Kalish, Charles. 1998. Reasons and causes: Children's understanding of conformity to social rules and physical laws. *Child development* 69(3). 706–720.

Kelley, H. H. 1967. Attribution theory in social psychology. In *Nebraska symposium on motivation*, vol. 15, 192–238. Springer Science.

Kelley, H. H. 1973. The processes of causal attribution. *American Psychologist* 28. 107–128. http://dx.doi.org/10.1037/h0034225.

Khoo, Justin. 2015. Modal disagreements. *Inquiry* 58(5). 1–24.

Klecha, Peter. 2014. *Bridging the divide: Scalarity and modality*: University of Chicago, Department of Linguistics dissertation.

Knobe, Joshua. 2010. Person as scientist, person as moralist. *Behavioral and Brain Sciences* 33(04). 315–329.

Knobe, Joshua & Ben Fraser. 2008. Causal judgment and moral judgment: Two experiments. *Moral psychology* 2. 441–8.

Knobe, Joshua & Zoltán Gendler Szabó. 2013. Modals with a taste of the deontic. *Semantics and Pragmatics* 6(1). 1–42.

Kolodny, Niko & John MacFarlane. 2010. Ifs and oughts. *Journal of Philosophy* 107(3). 115–143.

Kominsky, Jonathan F, Jonathan Phillips, Tobias Gerstenberg, David Lagnado & Joshua Knobe. 2015. Causal superseding. *Cognition* 137. 196–209.

Kratzer, Angelika. 1977. What 'must' and 'can' must and can mean. *Linguistics and Philosophy* 1(3). 337–355.

Kratzer, Angelika. 1981. The notional category of modality. In H.J. Eikmeyer & H. Reiser (eds.), *Words, worlds, and contexts*, 38–74. de Gruyter.

Kripke, Saul A. 1963. Semantical considerations on modal logic. *Acta Philosophica Fennica* 16(1963). 83–94.

Kushnir, Tamar, Alison Gopnik, Nadia Chernyak, Elizabeth Seiver & Henry M Wellman. 2015. Developing intuitions about free will between ages four and six. *Cognition* 138. 79–101.

Lassiter, Daniel. 2011. *Measurement and modality: The scalar basis of modal semantics*: New York University dissertation.

Levy, Gary D, Marianne G Taylor & Susan A Gelman. 1995. Traditional and evaluative aspects of flexibility in gender roles, social conventions, moral rules, and physical laws. *Child development* 515–531.

Lewis, David. 1973. Causation. *Journal of Philosophy* 70(17). 556–567.

Lewis, David. 1981. Ordering semantics and premise semantics for counterfactuals. *Journal of philosophical logic* 10(2). 217–234.

Locke, John. [1690]1975. *An essay concerning human understanding*, vol. 3. Oxford University Press. http://dx.doi.org/10.2307/2175691.

Lombrozo, Tania. 2010. Causal-explanatory pluralism: How intentions, functions, and mechanisms influence causal ascriptions. *Cognitive Psychology* 61(4). 303–332.

MacFarlane, John. 2009. Epistemic modals are assessment-sensitive. In Andy Egan & B. Weatherson (eds.), *Epistemic modality*, Oxford University Press.

Mandel, David R & Darrin R Lehman. 1996. Counterfactual thinking and ascriptions of cause and preventability. *Journal of personality and social psychology* 71(3). 450.

Markman, Keith D. & Audrey K. Miller. 2006. Depression, control, and counterfactual thinking: Functional for whom? *Journal of Social and Clinical Psychology* 25(2). 210–227. http://dx.doi.org/10.1521/jscp.2006.25.2.210.

Marr, David. 1982. *Vision: A computational investigation into the human representation and processing of visual information*. New York, NY, USA: Henry Holt and Co., Inc.

Matthewson, Lisa. 2016. Modality. In Maria Aloni & Paul Dekker (eds.), *The cambridge handbook of formal semantics*, Oxford University Press.

McArthur, Leslie A. 1972. The how and what of why: Some determinants and consequences of causal attribution. http://dx.doi.org/10.1037/h0032602.

McCloskey, Michael, Allyson Washburn & Linda Felch. 1983. Intuitive physics: The straight-down belief and its origin. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 9(4). 636.

McCloy, Rachel & Ruth Byrne. 2000a. Counterfactual thinking about controllable events. *Memory & Cognition* 28(6). 1071–8. http://dx.doi.org/10.3758/BF03209355.

McCloy, Rachel & Ruth MJ Byrne. 2000b. Counterfactual thinking about controllable events. *Memory & Cognition* 28(6). 1071–1078.

Nauze, Fabrice Dominique. 2008. *Modality in typological perspective*: Institute for Logic, Language and Computation, Universiteit van Amsterdam dissertation.

N'gbala, Ahogni & Nyla R Branscombe. 1995. Mental simulation and causal attribution: When simulating an event does not affect fault assignment. *Journal of Experimental Social Psychology* 31(2). 139–162.

Pearl, Judea. 2000. *Causality: Models, reasoning, and inference*. Cambridge University Press.

Phillips, Jonathan & Paul Bloom. 2017. Do children believe immoral events are impossible? Under revision.

Phillips, Jonathan & Fiery A Cushman. 2016. Multiple Systems for Modal Cognition. In *Proceedings of the 38th annual conference of the cognitive science society*, 3007. Austin, TX. https://mindmodeling.org/cogsci2016/papers/0653/index.html.

Phillips, Jonathan & Joshua Knobe. 2009. Moral judgments and intuitions about freedom. *Psychological Inquiry* 20(1). 30–36. http://dx.doi.org/10.1080/10478400902744279.

Phillips, Jonathan, Jamie Luguri & Joshua Knobe. 2015. Unifying morality's influence on non-moral judgments: The relevance of alternative possibilities. *Cognition* 145. 30–42.

Portner, Paul. 2009. *Modality*. OUP Oxford.

Roese, Neal J. 1994. The functional basis of counterfactual thinking. *Journal of personality and Social Psychology* 66(5). 805.

Roese, Neal J. 1997. Counterfactual thinking. *Psychological bulletin* 121. 133–148. http://dx.doi.org/10.1037//0033-2909.121.1.133.

Roese, Neal J. & Olson. 1994. *What might have been: The social psychology of counterfactual thinking*. Psychology Press.

Roxborough, Craig & Jill Cumby. 2009. Folk psychological concepts: Causation 1. *Philosophical Psychology* 22(2). 205–213.

Samland, Jana & Michael R Waldmann. 2016. How prescriptive norms influence causal inferences. *Cognition* 156. 164–176.

Schaffer, Jonathan. 2005. Contrastive causation. *Philosophical Review* 114(3). 327–358.

Shtulman, Andrew. 2009. The development of possibility judgment within and across domains. *Cognitive Development* 24(3). 293–309.

Shtulman, Andrew & Susan Carey. 2007. Improbable or impossible? how children reason about the possibility of extraordinary events. *Child Development* 78(3). 1015–1032.

Spelke, Elizabeth S. 1990. Principles of object perception. *Cognitive science* 14(1). 29–56.

Sripada, Chandra & Stephen Stich. 2006. A framework for the psychology of norms. In Peter Carruthers, Stephen Laurence & Stephen P. Stich (eds.), *The innate mind, volume 2: Culture and cognition*, Oxford University Press.

Sytsma, Justin, Jonathan Livengood & David Rose. 2012. Two types of typicality: Rethinking the role of statistical typicality in ordinary causal attributions. *Studies in History and Philosophy of Science Part C :Studies in History and Philosophy of Biological and Biomedical Sciences* 43(4). 814–820. http://dx.doi.org/10.1016/j.shpsc.2012.05.009.

Vander Klok, Jozina. 2012. *Tense, aspect, and modal markers in paciran javanese*: McGill University dissertation.

Veltman, Frank, Jeroen Groenendijk & Martin Stokhof. 1996. This might be it. In Dag Westerstahl & Jeremy Seligman (eds.), *Language, logic, and computation: the 1994 moraga proceedings*, 255–70. CSLI.

Vul, Edward, Noah Goodman, Thomas L Griffiths & Joshua B Tenenbaum. 2014. One and done? optimal decisions from very few samples. *Cognitive science* 38(4). 599–637.

Wells, Gary L. & Igor Gavanski. 1989. Mental simulation of causality. http://dx.doi.org/10.1037/0022-3514.56.2.161.

Wells, Gary L., Brian R. Taylor & John W. Turtle. 1987. The undoing of scenarios. *Journal of Personality and Social Psychology* 53. 421–430. http://dx.doi.org/10.1037/0022-3514.53.3.421.

Woodward, James. 2006. Sensitive and insensitive causation. *The Philosophical Review* 115(1). 1–50.

Woolfolk, Robert L, John M Doris & John M Darley. 2006. Identification, situational constraint, and social cognition: Studies in the attribution of moral responsibility. *Cognition* 100(2). 283–301.

von Wright, George H. 1953. An essay in modal logic. *Philosophical Quarterly* 3(12). 287–287.

Yalcin, Seth. 2007. Epistemic modals. *Mind* 116(464). 983–1026.

Yalcin, Seth. 2015. Epistemic modality de re. *Ergo, an Open Access Journal of Philosophy* 2(19). 475–527.

Young, Liane & Jonathan Phillips. 2011. The paradox of moral focus. *Cognition* 119(2). 166–178.