Semantics & Pragmatics Volume 6, Article 1: 1–42, 2013
http://dx.doi.org/10.3765/sp.6.1

Modals with a taste of the deontic*

Joshua Knobe
Yale University

Zoltán Gendler Szabó
Yale University

Submitted 2012-05-16 / Accepted 2012-07-16 / Revision received 2012-08-22 / Published 2013-03-29

Abstract The aim of this paper is to present an explanation for the impact of normative considerations on people's assessment of certain seemingly purely descriptive matters concerning freedom, causation, and intentionality. The explanation is based on two main claims. First, the relevant judgments are modal: the sentences evaluated are contextually equivalent to modal proxies. Second, the interpretation of predominantly circumstantial or teleological modals is subject to normative constraints which make certain possibilities salient at the expense of others.

Keywords: causation, freedom, intentional action, modality, normativity, pragmatics

When we assess an action as right or wrong we often want to know whether it was forced upon the agent, whether it had certain causal effects, and whether it was performed intentionally. It seems natural to think that these questions are prior to and independent of normative considerations. Recent experimental results have shown that in fact this is not so: when people consider such questions their answers often depend on whether the action has violated a norm.

Our aim here is to present a new account of these puzzling phenomena. It is tempting to see them as manifestations of diverse errors but we will argue that the data permit a unified and charitable explanation. At the heart of our account is a claim about modality. We argue that each of the judgments

* We are grateful for comments from Michael Bratman, Tad Brennan, Mark Crimmins, Josh Dever, Kai von Fintel, Tamar Gendler, Shelly Kagan, Angelika Kratzer, Jonathan Phillips, Rob Rupert, Jonathan Schaffer, Stewart Shapiro, Ted Sider, Jason Stanley, Brian Weatherson and Seth Yalcin, as well as to audiences at the Arché Research Centre, the University of Cologne, the University of Colorado, Cornell University, Stanford University and Yale University.

that show this surprising impact of normativity is modal and the modality in question, though primarily circumstantial or teleological, nonetheless shows a taste of the deontic.

1 The experiments

Before starting in with the explanation, it will be necessary to get a sense for the data. We begin with a brief summary of the three effects we want to focus on.

1.1 Freedom

Consider the distinction people draw between actions an agent chooses freely and actions that are forced upon her. The standard way of testing such matters is to have people read brief vignettes and then answer questions about the actions of the characters in these vignettes. We can then systematically vary specific factors within the vignettes and thereby determine the role they play in shaping judgments. Thus, if we want to know whether normative considerations are having any impact, we can construct a pair of vignettes that are identical except for the normative status of the agent's behavior and then check to see whether we find a corresponding difference in the responses to the scenario described.

In one recent study (Phillips & Knobe 2009), participants were randomly assigned to one of two conditions. Participants in the first condition were given the following vignette:

> 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 wife's expensive cargo overboard.
>
> Thinking quickly, the captain took her cargo and tossed it into the sea. While the expensive cargo sank to the bottom of the sea, the captain was able to survive the storm and returned home safely.

These participants were asked whether they agreed or disagreed with the following sentence:

ship/cargo The captain was forced to throw his wife's cargo overboard.

Participants in the other group received a vignette that was identical, except for changes designed to alter the normative status of the agent's behavior (changes shown in italics):

> 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 *wife* overboard.
>
> Thinking quickly, the captain took his *wife* and tossed her into the sea. While the captain's *wife* sank to the bottom of the sea, the captain was able to survive the storm and returned home safely.

These participants were asked whether they agreed or disagreed with this sentence:

ship/wife The captain was forced to throw his wife overboard.

The results showed a significant difference between conditions. Participants tended to disagree with *ship/wife* but to agree with *ship/cargo*. This difference provides at least some initial indication that normative considerations are playing a role in determining people's judgments about whether an agent has been forced to perform an action.

But, of course, a single experiment like this one can never provide decisive evidence. It will always turn out that the two vignettes differ in numerous respects — some of which have nothing to do with normative considerations *per se* — and one might always worry that one of these other differences is actually at the root of the observed effect. As long as one is relying just on a single experiment, it will be difficult to properly address this worry. The best approach, then, is to construct a number of different pairs of vignettes that share the same basic structure but that differ radically in their details (Phillips & Knobe 2009, Young & Phillips 2011). So, for example, in a separate study, participants received a vignette about a doctor who is ordered by the chief of surgery to prescribe a medicine that will either help a patient (in one condition) or harm a patient (in the other). The doctor is described as reluctantly agreeing to prescribe the medicine in both cases, but participants tend to say that he was forced in the help condition but not in the harm

condition (Phillips & Knobe 2009). As we accumulate more and more pairs of vignettes that show this same pattern of responses, it begins to seem increasingly implausible to search for a separate explanation for the effect on each pair. The more parsimonious explanation is that normative judgments actually are having an impact on their judgments as to whether or not the agent was forced to act.

1.2 Causation

If the effect described in the previous section arose only for this one expression, it would be natural to suppose that it was due to some idiosyncratic feature of the verb *force*. However, experimental results indicate that a similar effect arises for numerous other expressions. For example, one can find the same basic asymmetry in judgments about causation.

In one recent study of this effect (Knobe & Fraser 2008), all participants received 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.

The actions of the two agents are similar: both of them take a pen, and if either of them had not done so, the problem would not have arisen. But the two actions differ normatively: only one of the agents is violating a rule. The key question now is whether this normative difference has an impact on causal judgments.

To get at this question, participants were asked whether they agreed or disagreed with the following sentences:

pen/professor Professor Smith caused the problem.

pen/assistant The administrative assistant caused the problem.

Overall, participants tended to agree with *pen/professor* while disagreeing with *pen/assistant*. This result provides at least some initial evidence that normative considerations can influence causal judgments.

Still, if one looks only at the results of this specific experiment, a pressing worry remains. One might think that the judgments in this case are not truly being guided by normative considerations *per se* but rather by purely statistical facts about which actions are most typical. It is certainly more typical for employees to obey rules than to violate them, and the professor's action might therefore be seen as less typical than that of the administrative assistant. Thus, if one assumes that statistically atypical events tend to be picked out as causes, one could explain this result without any appeal to normativity.

This is a legitimate concern, and a number of further studies have been conducted to address it. First, it has been shown that the effect arises even when participants are told that faculty members always take pens and that administrative assistants never do (Roxborough & Cumby 2009). In this revised version of the study, the behavior of the administrative assistant is clearly more statistically atypical, yet participants continue to pick out the faculty member as the cause. This latter result cannot be explained in terms of statistical typicality and seems more clearly to involve a role for normative considerations. (For further experiments and discussion, see Sytsma, Livengood & Rose 2011.) Second, follow-up studies have looked at cases in which different participants make different moral judgments about the very same scenario. For example, one study presented all participants with the story of a woman who decides to terminate her own pregnancy (Cushman, Knobe & Sinnott-Armstrong 2008). Even though all participants received the very same case, some were pro-choice while others were pro-life, and these different participants therefore made quite different moral judgments about the case they had received. The results showed that this difference in moral judgment led to a corresponding difference in causal judgment, with pro-life participants being more likely to see the woman's action as a cause of subsequent outcomes. Here again, the results would be difficult to explain in terms of statistical typicality alone and seem to point to an independent role for normativity.

Another possible worry would be that the original effect might be best explained in terms of conversational pragmatics (Driver 2008). Perhaps saying

that someone caused something bad implicates that the person is to be blamed for the outcome, and perhaps people are reluctant to say that the assistant caused the problem not because this is false but because they do not want to endorse the implicature. This latter explanation, too, is perfectly compatible with the results of the original study, but follow-up work provides evidence against it. Thus, suppose that we alter the story so that the outcome becomes something good. (For example: the administrative assistant follows the rule, the faculty member breaks the rule, and these two behaviors are together sufficient for a good outcome.) In this altered version, there is no sense at all in which an assertion that an agent caused the outcome can implicate that she is deserving of blame. (One can't implicate that someone deserves blame by saying: "She caused this wonderful outcome.") Nonetheless, the original pattern of results continues to emerge. Even when the outcome is good, people say that it was caused by the agent who violated the norm and not by the agent who obeyed the norm (Hitchcock & Knobe 2009).

In short, the pattern of experimental evidence across a range of recent studies suggests that causal judgments can be affected by normative considerations and that this effect is not due simply to implicatures concerning blame.

1.3 Intentional action

Finally, consider intentional action. One might initially suppose that the question as to whether or not an agent performed a behavior intentionally is an entirely descriptive question, simply a matter of what the agent's intentions are and how the agent acts. Yet, once again, a series of experimental studies indicate that normative considerations can play a role.

To see the basic effect at work in such cases, consider the following vignette:

> Jake desperately wants to win the rifle contest. He knows that he will only win the contest if he hits the bulls-eye. He raises the rifle, gets the bull's-eye in the sights, and presses the trigger.
>
> But Jake isn't very good at using his rifle. His hand slips on the barrel of the gun, and the shot goes wild ...
>
> Nonetheless, the bullet lands directly on the bull's-eye. Jake wins the contest.

Now, keeping this vignette in mind, ask yourself whether you agree with the sentence:

rifle/bull's-eye Jake hit the bull's-eye intentionally.

The experimental results indicate that most participants do not agree with this sentence (Knobe 2003). The issue here, presumably, is that the agent is succeeding entirely by luck; he doesn't really have any control over the outcome of his behavior.

But now suppose that we switch around the normative significance of the story. We can leave intact all of the facts about the process, while simply altering the target.

> Jake desperately wants to have more money. He knows that he will inherit a lot of money when his aunt dies. One day, he sees his aunt walking by the window. He raises his rifle, gets her in the sights, and presses the trigger.
>
> But Jake isn't very good at using his rifle. His hand slips on the barrel of the gun, and the shot goes wild . . .
>
> Nonetheless, the bullet hits her directly in the heart. She dies instantly.

The corresponding sentence is then:

rifle/aunt Jake hit his aunt intentionally.

Yet people tend to say that this latter sentence is true (Knobe 2003). Of course, one might initially suppose that the difference here arises simply because it is easier to hit an aunt than a bull's-eye (Guglielmo & Malle 2010), but more tightly controlled studies show that the effect continues to arise even when the difficulty of hitting the two targets is kept the same (Sousa & Holbrook 2010). The effect has also been shown to arise even when the agent is doing something that requires no skill of any kind, such as rolling some dice (Nadelhoffer 2004) or guessing a random number (Nadelhoffer 2005). In all of these cases, the experimental results indicate that people are more willing to consider the behavior intentional when it violates a norm than when it does not.

1.4 Two desiderata

Thus far, we have been reviewing three different effects from the experimental literature. Stating the results as neutrally as possible what we found is this. In each case, we consider a sentence containing a word that appears to be entirely descriptive, but we find that normative considerations influence how the word is used. The question now is how these various effects are to be explained.

One possible approach would be to seek separate explanations for each of the separate effects: one explanation for the effect on *force*, another for the effect on *cause*, a third for the effect on *intentionally*. But this sort of approach risks missing an insight into the pattern as a whole. It is reasonable to expect that the root of the unexpected impact of the normative is the same across the board. So, we place our bets on the opposite strategy: we believe the ideal explanation of these asymmetries should be a *unified* one. This is our first desideratum.

One way of providing a unified explanation would be to posit a performance error of some sort that impacts people's judgments in a number of different domains. That is, one could say that normative considerations actually are not relevant to any of these questions but that people are messing up in some way and allowing their judgments to be improperly influenced. Here again, we will be adopting a different approach: we will be trying to develop an explanation for the encroachment that is *charitable*. This is our second desideratum.

We expect that our second desideratum will raise more eyebrows than the first, so let us say a few words in its support. The first thing to note here is that it is not *obvious* that there has to be a performance error at work in these effects. Participants show many of the asymmetries even in 'within-subject' designs where they receive the two cases back-to-back and can clearly see that the only difference between them is a normative one (Knobe & Fraser 2008, Young, Cushman, et al. 2006). In short, it seems that one cannot dismiss these effects as merely the product of some minor slip-up that could be easily corrected on further reflection.

But, of course, that is not the true test of the performance error approach. What we really want to know is whether it is possible to develop a specific performance error hypothesis that can accurately predict and explain the data. A number of researchers have provided possible hypotheses (Adams & Steadman 2004, Alicke 2008, Nadelhoffer 2006, Nanay 2010). Each of these

hypotheses describes a specific error and explains how such an error could lead to the asymmetries observed.

These hypotheses also make new predictions, which have been put to the test in further experiments. There have been studies using reaction time measures (Guglielmo & Malle 2011), patients with lesions to the specific brain regions (Young, Cushman, et al. 2006), participants with Asperger's syndrome (Zalla & Leboyer 2011), and numerous studies that simply use additional vignettes (Nichols & Ulatowski 2007, Sripada & Konrath 2011). In each case, the data have failed to vindicate the performance error hypotheses. Instead, studies have again and again shown that the predictions made by these hypotheses are not borne out (for a review, see Knobe 2010).

A possible response would be to say that although the existing performance error hypotheses happen not to be quite right, we simply have not yet picked out the right sort of error. This might ultimately prove correct, but our hunch is that it would be best at this point to begin looking elsewhere for a solution. Perhaps the reason we can't find the error in people's judgments in these cases is that there is no error to be found.

Charity towards the majority should not be combined with dismissing the minority. We will defend the view that normative considerations can and normally do impact our judgments about freedom, causation, and intentionality. But it is not part of our view that they should or inevitably will have such an impact. We sometimes find ourselves in special contexts where conversational participants seek to diminish the effect of this impact. When we make a conscious effort to assess a situation with full objectivity and impartiality, perhaps we can diminish or entirely neutralize the effect shown in the examples above. Perhaps this is what sometimes happens in scientific or philosophical inquiry and perhaps this is what should always happen in a jury trial. Moreover, some people might be especially prone to think in this way, and some of them might be inclined to stick with it even in everyday settings. A good explanation should leave room for this, and we will certainly try.

2 Modality as the common core

Our hypothesis is that the judgments under consideration are modal. Although there are important differences among the target sentences, we propose that in understanding each within the contexts of its respective vignette people exercise a certain modal assessment. It is this modal assessment, we

suggest, that explains the impact of normative considerations.

At this point, the straightforward way to proceed would be to present and justify a modal semantics for *force*, *cause*, and *intentionally*. With such a semantics at hand, we could say that the relevant judgments are modal simply because they are judgments of the truth or falsity of modal sentences. However, we believe that the connection between our target sentences and modality is less direct. These sentences are not modal; they just have a salient modal entailment which becomes equivalent to the original in the context established by the vignette. We call sentences that express these contextually equivalent modal entailments *modal proxies*.

By pairing the target sentences with their modal proxies we unify the explanatory tasks we face: instead of trying to explain how normative considerations impact sentences about freedom, causation, and intentionality, we can focus just on how such considerations impact seemingly non-normative modal claims. This brings out what we take to be the common core of the phenomena under consideration and it also reduces the dialectical burden on the semantic component of our explanation. Contextualism about freedom, causation, or intentionality are controversial doctrines. But when it comes to modality, contextualism is the established view.[1]

Accordingly, we proceed in two stages. First we associate each of the original target sentences with a corresponding modal proxy; then we offer a more general characterization of the relationship we have called 'contextual equivalence' that obtains between the proxies and the original sentences.

2.1 *Force*

The significance of modality is clearest in the case of *force*. Speakers who judge that the captain was forced to throw the cargo overboard are likely

1 Setting aside epistemic modality for the moment, where relativist accounts are a strong competitor (Egan, Hawthorne & Weatherson 2005, Stephenson 2007, but see Yalcin 2011). Portner 2009 is a standard survey of current views on the semantics of modals in natural languages; it mentions a wide variety of approaches, not one of which is invariantist. Most semantic approaches interpret modals via quantification over possibilities, and at the very least the domain of this quantification is supposed to be contextually determined. Many approaches to modal semantics employ an accessibility relation instead of explicit domain restriction, but these too accept that the relevant accessibility relation must be contextually provided. Semantic minimalists reject this sort of context-sensitivity; but they also part with mainstream semantics in rejecting that *ready*, *tall*, or *usually* are context-sensitive. We reject semantic minimalism; for arguments see Szabó 2006.

to justify their claim by asserting *ship/cargo*^{M}, while those who deny that the captain was forced to throw his wife overboard will be likely to deny *ship/wife*^{M}. (We will use the superscript 'M' to indicate the modal proxies of our target sentences.)

ship/cargo^{M}	Given the circumstances, the captain had to throw the cargo overboard.
ship/wife^{M}	Given the circumstances, the captain had to throw his wife overboard.

Indeed, even those who disagree with the dominant judgments about these cases would probably not object to casting the disagreement in these modal terms. It appears that the judgments made in this case have genuine modal correlates and that — at least in the context established by the vignettes — people feel comfortable moving back and forth between the original and the proxy.

To put this claim to the test, we conducted a new study.[2] Participants received the very same stories about the captain and the storm that had been used in earlier research, but instead of being given the original sentences about whether the agent was forced, they were given the modal proxies *ship/cargo*^{M} and *ship/wife*^{M}. The results showed that the asymmetry observed for the target sentences also arose for the modal proxies. Participants tended to agree with the claim that the captain had to throw the cargo overboard, but they tended to disagree with the claim that the captain had to throw his wife overboard.[3] (Moreover, the results showed that this difference between conditions in people's judgments about the modal proxies was entirely mediated by a difference in their normative judgments: participants only gave different judgments about the modal proxies to the extent that they gave correspondingly different judgments about what it would have been best on the whole for the captain to do.[4])

2 Participants were 42 people recruited through Amazon's Mechanical Turk. Each participant rated sentences on a scale from 1 ('disagree') to 7 ('agree').

3 Ratings for the cargo condition (M = 6.8, SD = .4) were significantly higher than those for the wife condition (M = 2.6, SD = 2.3), t (40) = 7.9, p < .01.

4 In addition to the target sentences and modal proxies, each participant was asked an explicitly moral question about which option would have been better on the whole for the captain to choose: throwing the cargo/wife overboard or not throwing the cargo/wife overboard. We used mediational analysis to examine the relationship between condition, moral judgment and modal judgment. Condition had a significant impact on moral judgment

2.2 *Cause*

Turning to judgments about causation, matters become a little bit more complex. A broad array of different researchers in both philosophy and psychology have suggested that causal judgments might in some way be connected to modal reasoning, but different researchers have offered interestingly different views about precisely how this connection works (Halpern & Pearl 2005, Hitchcock 2007, Lewis 1973, 2000, Woodward 2003). Our approach here will rely on the traditional claim that causes necessitate their effects. The traditional claim has a myriad of counterexamples: cases when the cause occurs but something intervenes and as a result the effect which normally would follow fails to occur. We suggest that while such interventions are acknowledged they are also typically ignored. When someone reads the vignette they assume that if possible interventions were important they would have been made salient, and since they had not been, they can be properly ignored. Thus, we suggest that when people agree with the claim that the professor caused the problem, they will also agree with the explicitly modal proxy *pen/professor*$^{\text{M}}$, and when they disagree with the claim that the administrative assistant caused the problem, they will also disagree with the explicitly modal proxy *pen/assistant*$^{\text{M}}$.

pen/professor^{M}	Given the actions of the professor, the problem had to occur.
pen/assistant^{M}	Given the actions of the administrative assistant, the problem had to occur.

To confirm this suggestion, we ran an additional experiment. Participants were given the story of the missing pens and then randomly assigned to evaluate a sentence about that vignette. Some participants were given one of the original causal sentences: *The professor/administrative assistant caused the problem.* Others were given one of the proposed modal proxies we identified.[5]

($ß = .75$, $p < .01$). When moral judgment was entered as a regressor, the impact of condition on modal judgment decreased from $ß = .78$, $p < .01$ to $ß = .35$, $p < .01$. A Sobel test showed that this reduction was significant, $Z = 21.4$, $p < .01$. In other words, the difference between conditions appears to be impacting people's modal judgments in part by impacting their moral judgments.

5 Participants were 80 people recruited through Amazon's Mechanical Turk. Each participant rated the sentence on a scale from 1 ('disagree') to 7 ('agree'). Data were analyzed using a 2 (sentence type: causal vs. modal) x 2 (agent: professor vs. assistant) ANOVA. There was a significant main effect of agent, $F(1, 76) = 43.6$, $p < .001$. There was no significant main effect of sentence type and no significant interaction.

Unsurprisingly, we replicated the original effect whereby people agree more with the causal sentence about the professor than they do with the causal sentence about the administrative assistant.[6] But we also found the same effect for the modal proxies: people showed a moderate level of agreement with the explicitly modal sentence about the professor but disagreed with the explicitly modal sentence about the administrative assistant.[7]

2.3 *Intentionally*

Turn now to the case of *intentionally*.[8] The original experiment showed an asymmetry in people's intentionality judgments, such that people were less inclined to say that the agent hit the target intentionally in the bull's-eye case than they were in the aunt case. We now want to suggest that this asymmetry in intentionality judgments arises because there is an asymmetry in people's intuitions about the role of *luck* across the two cases. Specifically, the claim is that people are more inclined to say that the agent hit the target through sheer luck in the bull's-eye case than they are in the aunt case.

To determine whether or not people do have such asymmetric intuitions about the role of luck, we conducted one final experiment.[9] Participants received either the bull's-eye vignette or the aunt vignette. All participants were then asked whether they agreed or disagreed with two statements: (1) *Jake hit the bull*s-eye/his aunt intentionally.' (2) *Jake hit the bull*s-eye/his aunt through sheer luck.' As one might expect, we replicated the original intentionality asymmetry, with participants being significantly less inclined to say that Jake acted intentionally in the bull's-eye case than in the aunt case.[10] The more important finding, however, was the one for the luck question. There, we found a significant effect whereby participants were more inclined to say that Jake hit his target by sheer luck in the bull's-eye case than in the

6 Mean rating for the causal statement about the professor: 5.4, mean rating for the causal statement about the assistant: 2.3, t (33) = 6.6, $p < .001$.

7 Mean rating for the modal statement about the professor: 4.5, mean rating for the modal statement about the assistant: 2.7, t (43) = 3.5, $p = .001$.

8 Throughout this section, our thinking has been deeply influenced by the work of Falkenstien (forthcoming).

9 Participants were 52 people recruited through Amazon's Mechanical Turk. Each participant rated the sentence on a scale from 1 ('disagree') to 7 ('agree'). The order of questions was counterbalanced.

10 Mean rating in the bull's-eye case: 6.6; mean rating in the aunt case: 3.3, t (50) = -7.1, $p < .001$.

aunt case.[11]

We can now draw on this finding to propose a modal correlate for intentionality judgments. First, along with many traditional philosophical analyses (e.g., Mele & Moser 1994), we assume that the claim that an act was performed intentionally entails that it was not performed by sheer luck. Second, we suggest that the judgment that an agent performed an action by sheer luck is modal.

The key idea here is that the claim that an action was performed 'by sheer luck' can be spelled out in terms of causation. Thus, when people say that the outcome in the bull's-eye case arose by sheer luck, they mean that it was caused by the various lucky events that occurred in the vignette (e.g., the hand slipping on the barrel of the gun) rather than by the agent's psychological states (e.g., his intention to hit the target). Suppose we now use the word *fluke* to pick out the lucky events that occurred in both vignettes. If we continue to rely on the idea that causes necessitate their effects, the modal proxies for our two sentences become:

rifle/bull's-eye^{M} It is not the case that given the fluke, Jake had to hit the bull's-eye.

rifle/aunt^{M} It is not the case that given the fluke, Jake had to hit his aunt.

These modal proxies differ from the others in that they make use of a stipulation that is hard to make precise. One could say that the fluke is the particular event that occurred when Jake's hand slipped on the barrel but the gun somehow stayed on target, but thinking about just what *that* event might be will quickly lead to puzzles of event individuation. The puzzles are made worse by the fact that the event was *random*. Finally, we had to phrase the modal proxies using the somewhat awkward explicit clausal negation in order to eliminate different readings. The joint effect of such factors may easily make the reflective wary about passing judgments about *rifle/bull's-eye*^{M} and *rifle/aunt*^{M}, which unfortunately renders them less amenable to direct experimental test.

11 Mean rating in the bull's-eye case: 5.9; mean rating in the aunt case: 4.4, t (50) = 2.9, p = .005.

2.4 The nature of modal proxies

Here is the summary of the target sentences and their proposed modal proxies:[12]

ship	The captain was forced to throw the (a) cargo/(b) his wife overboard.
ship^{M}	Given the circumstances, the captain had to throw (a) the cargo/(b) his wife overboard.
pen	(a) The professor/(b) The assistant caused the problem.
pen^{M}	Given the action of (a) the professor/(b) the assistant, the problem had to occur.
rifle	Jake hit (a) the target/(b) his aunt intentionally.
rifle^{M}	It is not the case that given the fluke, Jake had to hit (a) the target/(b) his aunt.

(For definiteness, we have offered a specific modal proxy for each of the original sentences whose use is to be explained. However, the core idea of the proposal does not depend on these precise modal correlates being the right ones. If you think that another, slightly different modal proxy might be more accurate, please hold on until Section 4. Then you can check to see whether the explanation we propose there works for your favored modal proxy as well.)

Each of the modal proxies is entailed by its target sentence but the converse is clearly false. Imagine a case where the captain fails to notice the storm but decides to throw his wife's expensive cargo overboard simply out of spite. This is a case where *ship/cargo*^{M} is true (given the circumstances, the captain still has to throw the cargo aboard) but *ship/cargo* is false: the captain throws the cargo overboard for his own reasons, and so he is not forced to act. Now, take a case when there is just one pen left at the desk of the secretary, the professor sends a graduate student to pick it up, but before the graduate student would get there the administrative assistant takes the pen. Here *pen/professor*^{M} is true (given the professor's action — sending

12 These sentences differ slightly from the ones used in the original experimental studies. The changes are insubstantial; we made some simplifications to reduce verbiage.

the graduate student to pick up the pen — the problem at the desk had to occur) but *pen/professor* is false: due to interference, the professor's action played no causal role in taking the last pen from the secretary's desk. Finally, consider a case just like the third vignette, except that when Jake gets his aunt in the sights he mistakes her for his uncle, whom he also wants to kill. He pulls the trigger; the shot goes wild but nonetheless hits and kills his uncle. Now *rifle/aunt*^{M} is true (it is not the case that given the fluke Jake had to hit his aunt) but *rifle/aunt* is false: Jake meant to kill his uncle, so he certainly did not shoot his aunt intentionally.

Fortunately, what we need to get our explanation going is not mutual entailment in all contexts — it is merely mutual entailment in the particular context established by the vignettes. At the time when we are evaluating our target sentences we take a lot of information about the case for granted. Some of this information has been conveyed *explicitly*: it is entailed by the text of the vignette. But most of it is conveyed only *implicitly*: it is taken for granted by any normal reader of the text without being entailed. Consider the first vignette. It entails that there was a storm threatening a ship, that the captain of the ship threw his wife's cargo overboard, and that the cargo sank to the bottom of the sea. It does *not* entail that the ship did not sink. We are told that the captain did what he could to save the boat, that he was able to survive the storm, and that he returned home safely, but all this is fully compatible with the possibility that the ship went down despite the captain's action, that he survived the storm on open sea, and that he was eventually rescued by another ship. Still, this is not a scenario normal interpreters of the vignette would envision. They take it for granted that the ship was saved; had it not been the speaker should have put things differently. Now consider what information is semantically encoded in *ship/cargo* and *ship/cargo*^{M}. Arguably, the former entails that the captain's action was influenced by something beyond his control of which he was aware, while the latter does not. We suggest that the vignette implicitly conveys the proposition that the captain's action was influenced by something beyond his control of which he was aware.

Our hypothesis is that the truth-conditional differences between our target sentences and their modal proxies are *canceled out* by the information conveyed by the vignette in which they are evaluated. Let's say that two sentences are *contextually equivalent* relative to a story just in case assuming the information conveyed explicitly and implicitly by the story neither can be true if the other is false. Then the claim is that relative to their respective

vignettes, *ship* is contextually equivalent to *ship*^{M}, *pen* to *pen*^{M}, and *rifle* to *rifle*^{M}.

Our hypothesis predicts, for example, that by and large people's judgments about the truth or falsity of the target sentences in our experiments will line up with their judgments about their modal proxies. But the hypothesis does *not* make the stronger prediction that people will be inclined to judge that the target sentences are true just in case their modal proxies are. In making judgments of equivalence people abstract from the contextual information they have, and so the semantic differences between the target sentences and their proxies come to the fore. Also, we are *not* saying that when people think about the truth or falsity of the target sentences they must consider the particular English sentence we used here as a modal proxy. That would amount to making the implausible prediction that people who know enough English to understand *ship/cargo* or *rifle/aunt*, but don't know what the words *circumstances* or *action* mean will react differently to the vignettes than the rest of us. What matters is that if the relevant contextual equivalence holds and we if can explain judgments about the proxies, those explanations carry over to the target sentences as well.

One final clarification: when we say that the target sentences are contextually equivalent to modals, we are not saying that these sentences actually *are* modals. In fact, we explicitly deny that they are. Genuine modals have important properties that these sentences clearly do not share. For example, consider the modal: *Jake had to hit his aunt*. This modal can be used with a given-clause, such as: *Given that he aimed at her, Jake had to hit his aunt*. As we discuss further below, this given-clause serves to explicitly restrict the domain of possibilities that are treated as relevant in interpreting the modal. But now suppose one tries to do the same with our corresponding target sentence. One then ends up with: *Given that he aimed at her, Jake hit his aunt intentionally*. Since *intentionally* is not itself a modal, there is nothing for the given-clause to restrict, and the sentence therefore makes little sense. (One doesn't quite see what the given-clause is supposed to accomplish.)

The point here generalizes. Contextual equivalence is a fragile matter: the fact that a sentence has a modal proxy is a certain context does not imply that a sentence of a similar form will have a similar modal proxy in another context. While we do believe that the phenomenon uncovered in the experiments is robust, that there are many perfectly natural contexts where various sentences about freedom, causation, or intentionality are evaluated via modal proxies of the sort we have assigned to our target sentences,

there is no recipe here for establishing a fixed link between such sentences and restricted necessity modals. To claim otherwise would be to mistake what we offer for something much more ambitious — a full-fledged semantic analysis.[13]

Let's sum up where we are. We have a two-step plan to provide a unified and charitable explanation of the results of the experiments discussed in the section 1. The first step is to reduce the problem of accounting for the impact of normative considerations on judgments about freedom, causation, and intentionality to the impact of normative considerations on modal judgment. The first step is now done: we have identified modal proxies for each of our target sentences. What remains is the second step: to explain why normative considerations impact the interpretation of the modal proxies.

3 Modal flavors

Our approach to modals will follow the classic framework introduced by Kratzer (1977, 1981), although we will simplify it somewhat. The framework is flexible enough to permit the introduction of principles which will provide an explanation of the surprising ways in which normative considerations impact people's intuitions about modals. But before we can begin sketching this new possibility, it will be necessary to briefly review the semantic framework.

3.1 From ambiguity to contextualism

One of the first things one notices about modal verbs like *have (to)* is that they are used in slightly different ways in different sentences.[14] For example, in normal contexts *have (to)* permits at least a circumstantial, a teleological, and a deontic reading:

13 We hold open the possibility that the target sentences may have a *conjunctive* analysis — one of their conjuncts being a modal proxy and the other a non-modal sentence. For such a proposal in the case of *know*, see Schaffer & Szabó forthcoming.

14 In English, as in other languages, modal expressions fall in a variety of syntactic categories. Following the literature, we are assuming that modal verbs (*have (to)*, *need (to)*, *ought (to)*, *dare (to)*, etc.), modal auxiliaries (*must*, *can*, *might*, *should*, etc.) and modal adverbs (*possibly*, *necessarily*, *probably*, *maybe*, etc.) are all interpreted as expressions taking clausal scope. This assumption may well be incorrect but not in a way (we hope) that would matter for our purposes.

(1) a. I had to sneeze during the talk.
b. Given my circumstances, I had to sneeze during the talk.

(2) a. You have to turn left at the next intersection.
b. Given your goals, you have to turn left at the next intersection.

(3) a. You have to stop seeing that woman.
b. Given your moral obligations, you have to stop seeing that woman.

The standard, theory-neutral term for these different uses is *flavor*. One obvious approach to flavors would be to suggest that there is an ambiguity in *have (to)*: it has a circumstantial meaning a teleological meaning, a deontic meaning, and so on, for each of the different flavors we might uncover.[15]

Yet, though this approach may at first seem plausible, Kratzer shows that it cannot be the right one. Focus just on the alleged deontic meaning of *have (to)* in (3a). It seems that this sentence could still be used with subtly different meanings in different contexts. In one context, it might be a specifically moral appeal. But in another context, it could be a prudential one, paraphraseable as *Given your best interests, you have to stop seeing that woman.* Or it could be an altruistic claim, as in *Given her best interests, you have to stop seeing that woman.* Alternatively, the speaker may withhold judgment on what morality demands and on what is in anyone's best interest, voicing simply what company policy requires. At the bottom of this slope lies a theory postulating an indefinite number of deontic meanings for *have (to).* Let's not go there.

The obvious alternative is to say that all deontic modals share a single meaning and the differences one finds among their different uses are due to the influence of conversational context. The function of *given*-clauses would be simply to make information relative to which they are to be understood more explicit.

But once we come this far it isn't easy to stop. Instead of positing separate meanings of *have (to)* to explain the differences between different flavors of modality, Kratzer suggests that these differences, too, are to be explained in terms of context-dependence. For example, perhaps the only reason we think *have (to)* is circumstantial in (1a) and teleological in (2a) is that we expect

15 Besides these three, it is customary to talk about epistemic, bouletic, doxastic, and stereotypical flavors as well as of ability modals. Philosophers who believe in distinctive logical, conceptual, physical, or metaphysical modalities — and believe in addition that these are expressible in natural language — would add further items to the standard list.

them to be used in different contexts. The expectation can be cancelled, in which case the interpretation shifts. In contexts appropriately primed, (1a) will express a teleological necessity and (2a) a circumstantial one:

(1) [My friend three seats to my left fell asleep during the talk and I sneezed to wake him up and spare him of further embarrassment. Afterwards, I say:]

a. I had to sneeze during the talk.

b'. Given my goals, I had to sneeze during the talk.

(2) [You are driving and I am holding a gun against your head. I say:]

a. You have to turn left at the next intersection.

b'. Given your circumstances, you have to turn left at the next intersection.

Kratzer's arguments here are highly persuasive, and it is now widely accepted that the differences between circumstantial, teleological and deontic modals should be understood not in terms of an ambiguity but rather in terms of a contextually given parameter. Kratzer herself takes this approach even farther — she suggests that the very same semantics to interpret epistemic modals as well:

(4) a. There has to be another copy of this book in the library.

b. Given what I know, there has to be another copy of this book in the library.

This last claim has been controversial. Some researchers have argued that while it might be possible to give a single unified theory of all of the non-epistemic modals (usually grouped together as *root modals*), epistemic modals are deeply different in important respects (Yalcin 2007, Gillies 2010).[16] We leave this controversy to one side and simply focus on root modals.

Our assumption, then, will be that when *have (to)* is interpreted as a root modal the expression has a single meaning. The question now is what that meaning might be.

16 Kratzer 1991: 650 allows that syntactic differences between epistemic and root modals may correlate with differences in their argument structure.

3.2 Modality and quantificational domains

To begin with, it may be helpful to introduce an analogy. Consider the quantifier *all* and its use in sentences like *All the beer is in the refrigerator*. Suppose this sentence is used by the host of a party when she sees one of her guests standing with an empty beer bottle looking around anxiously. Plausibly, in this context the sentence does not express the proposition that all of the beer in the entire universe is in the refrigerator. If we wanted a good paraphrase — a sentence that expresses more or less what the original sentence does without relying on the context as heavily as the original sentence does — we could say *All the beer that is such-and-such (i.e. in the room, around here, easily obtainable, designated for the party, etc.) is in the refrigerator*. The sentence quantifies over beer within a limited *domain* — not over all the beer there is, only over all the beer that is such-and-such.[17]

What are domains? It is natural to think that they are sets: the sets of those items we quantify over. But this cannot be right, for sets contain their members necessarily but a sentence evaluated under different circumstances would quantify over different items. Let's suppose the host of the party in the above example used the sentence *All the beer is in the refrigerator* to quantify over bottles of beer in his apartment and assume that all those bottles are indeed in the refrigerator. Then the sentence of the host expresses a truth. Now, consider a counterfactual situation where all those bottles are still in the fridge but there is an additional beer bottle placed perspicuously on the kitchen table. Intuitively, the sentence still expresses the very same proposition but now that proposition is false. We can accommodate this observation if we represent domains by functions from circumstances of evaluation to sets.[18]

If modals are quantifiers over possibilities, we should expect that they too quantify over a contextually restricted domains. Those who judge that in the scenario described by our first vignette the sentence *The captain had*

17 This view is standard but not uncontested. For a defense, see Stanley & Szabó 2000. Many philosophers (e.g., Bach 1994) believe that *All the beer is in the refrigerator* does not semantically express a proposition; some (e.g., Cappelen & Lepore 2005) contend that it expresses a minimal proposition whose truth-conditions we cannot spell out in non-disquotational fashion. These theorists agree that domains enter interpretation only at the level of ascertaining what the speaker uttering this sentence asserted in the context. If they are right, our story about domain restriction for modals should also be presented at that level. We do not believe that this would require substantive changes.

18 For such a proposal, see Soames 1986: 356, Stanley & Szabó 2000: 252.

to throw the cargo overboard is true do not believe that there are absolutely no possibilities of any kind in which the captain refuses to throw the cargo overboard; they presumably realize that such stubbornness would not go against the laws of logic or even the laws of nature. A reasonable paraphrase for the sentence would be *In all possibilities that are such-and-such, the captain throws his wife overboard*. Here too, context has to select a domain for the sentence to quantify over — not the domain of all the possibilities that there are, only a domain of possibilities that are such-and-such.

There are good reasons to think that domain restriction for modals is more complex than domain restriction for quantificational determiners. Beer in a liquor store miles away is not in the domain because it is *irrelevant* in the context of helping out a party guest. But possibilities where the captain does not throw the cargo overboard are not at all irrelevant in the context of our first vignette. Such possibilities are directly invoked when it is said that the only way that the captain could keep the ship from capsizing was to throw his wife's expensive cargo overboard. It would be quite implausible to assume that upon reading about what would happen if the captain failed to throw away the cargo people simply discard this possibility. Intuitively, the necessity modal in *The captain had to throw the cargo overboard* does not quantify over all the relevant possibilities, only over those that are in some sense *best* among the relevant ones.

Context has to provide two pieces of information for the interpretation of modals: one to settle which possibilities are relevant, and another to settle which of the relevant possibilities are best. The value of the first parameter is a domain, representable as a function that assigns to any possibility a set of possibilities that are relevant in that possibility. The value of the second is a ranking, representable as a function that assigns to each possibility a partial order (reflexive, anti-symmetric, transitive binary relation) on possibilities. These jointly determine a domain, representable as a function from possibilities to sets of possibilities that are the best among the relevant ones in the possibility. Let's call this the *inner domain*, distinguishing it from the *outer domain*, which is just the value of the first parameter. The idea is that modals quantify over inner domains.

Assuming possibilities are worlds (an optional but standard assumption) we can identify domains with functions from possible worlds to sets of possible worlds and rankings with functions from possible worlds to partial orders on possible worlds. If we take modals to be operators (another optional but fairly common view), the semantic clause for *have to* can be

written as follows.[19]

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

Given-clauses function as explicit restrictors on the outer domain: possibilities incompatible with what is said to be given are all irrelevant in assessing a modal following the *given*-clause. The simplest way to achieve this effect is to say that the semantic value of *given* α (where α could be a noun phrase or a free relative) is also a domain, i.e. a function from possible worlds to sets of possible worlds:

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

Thus, for example, *Given the storm, the captain had to throw his wife overboard* is true in world w relative to outer domain $f(w)$ and ranking $g(w)$ just in case *The captain throws his wife overboard* is true in every possible world that is best according to $g(w)$ among the worlds in $f(w)$ where the storm occurs.

The semantic clauses given here diverge from Kratzer's in two ways. First, on Kratzer's view, the two contextual parameters provide the same sort of items: functions from possible worlds to sets of sets of possible worlds. She calls these *conversational backgrounds*. If we identify sets of possible worlds with propositions, these functions can be seen as assigning *premise sets* and the truth-conditions of necessity modals can be stated in terms of entailment. It is well-known that that the semantics for counterfactuals and modals provided in Kratzer's premise semantics is equivalent to a version of ordering semantics.[20] Second, Kratzer rejects the *limit assumption*, according to which an inner domain always assigns a non-empty set to any possible world. In cases when it does not, our semantic clause yields vacuous truth-conditions, which in turn yield uncomfortable predictions. Thus, Kratzer

19 For the record, we believe neither of these assumptions. We prefer to think of possibilities more along the lines of situations and we think modals are genuine quantifiers binding situation variables at the level of logical form. But these are some of the many semantic assumptions we do not wish to defend here — they have nothing to do with the topic of this paper.

20 See Lewis 1981. Setting aside complications arising from rejecting the limit assumption, Kratzer's semantics is equivalent to the ordering semantic introduced in Pollock 1976. Lewis has also argued that the differences between Kratzer's premise semantics on the one hand, and Lewis's or Stalnaker's versions of ordering semantics on the other are relatively minor.

uses a more complex clause for modals which permits infinite chains of relevant worlds that that are better and better in terms of the ranking. It remains controversial whether the limit assumption holds. Moreover, there are mechanisms less drastic than wholesale abandonment of our semantic clauses to accommodate apparent exceptions to the limit assumption.[21]

While there might be good reasons to prefer premise semantics to ordering semantics, and there certainly are good reasons to be wary of the limit assumption, we feel comfortable that the simpler semantics we outlined will do fine for our purposes. The ranking we appeal to is definable in any sensible account of the semantics of counterfactuals and degree modals. Kratzer relies on it too; it is just the she extracts it from a conversational background she calls *ordering source*. And violations of the limit assumption arise only over infinite domains of relevant possible worlds, and we have no reason to think that in evaluating the target sentences in our experiments infinite domains of possibilities are ever contemplated.[22]

3.3 Impurity

Thus far, we have been reviewing some of the basic elements of (a variant of) the classic Kratzer framework. We now want to suggest that this framework gives us just the resources we need to explain the surprisingly pervasive impact of normative considerations on modal intuitions.

If modals were ambiguous, it would be natural to posit a list of possible flavors and to assume that any given modal had to fit neatly into one of them. Some modals would be purely circumstantial, some purely teleological, some purely deontic, but no single modal could include a mix of these different flavors. (It would make no sense, for example, to suppose that a given modal was best interpreted as being mostly teleological but also a little bit deontic.) We will refer to this view about the relationship between different flavors of modality as the assumption of *modal purity*.

As soon as one gives up the idea that modals are ambiguous and shifts in-

21 For arguments in favor of the limit assumption, see Stalnaker 1980. For a non-semantic mechanism to handle apparent violations of the limit assumption, see Swanson 2011.

22 A typical example when the limit assumption might be violated is *If I had been over seven feet tall, I would still be less than ten feet tall.* For any world w where I am over seven feet tall there is another w' just like w where I am closer to my actual height but still over seven feet tall. Arguably, w' is a relevant possibility that is better than w. It might be claimed with some plausibility that proper understanding of this sentence requires that, in some sense, we consider an infinite set of possible worlds.

stead to a theory based on context, the assumption of modal purity begins to look suspect. The most natural way of thinking about conversational contexts would be to assume that they embody a mixture of different information. In a given context, we might be primarily concerned with the circumstances but also somewhat concerned with not acting immorally and achieving certain goals. In other words, when one shifts from a theory based on ambiguity to a theory based on context, it is only natural to suppose that there can be *impure modals*.

Of course, even if one adopts all aspects of the framework outlined above, it would still be possible to hold onto the idea of pure flavors by insisting that both the outer domain and the ranking of a modal are fixed by contextual information of a single type. So, for example, let f be a purely bouletic domain iff $f(w)$ is a set of all worlds compatible with the satisfaction of someone's desires in w, let f be a purely stereotypical domain iff $f(w)$ is a set of all worlds compatible with certain expectations in w, etc. Similarly, let g be a purely teleological ranking iff $g(w)(v, v')$ holds just in case for certain goals if one of them is achieved in v' it is also achieved it in v, let g be a purely deontic ranking iff $g(w)(v, v')$ holds just in case for certain rules if one of them is followed in v' it is also followed it in v, etc. The assumption of modal purity is that contexts always assign a pure outer domain and a pure ranking to any occurrence of a modal.

Perhaps the strongest motivation for believing in pure flavors of this type is the existence of lexical restrictions on what sort of meaning modal expressions can have. (*Might* tends to be epistemic, *ought* tends to be deontic, and so forth.[23]) This fact suggests that modals truly do come in distinct flavors. Given the sort of contextualist framework we assume, one way of making sense of lexical restrictions would be to assume that certain modal expressions are licensed only when both the outer domain and the ranking are pure and of the appropriate type.

It is important to note, however, that one does not need to assume the existence of pure flavors to understand lexical restrictions. All one needs is the much weaker assumption that the outer domain and the ranking are determined on the basis of information that is *predominantly* of a single type. Thus, when the ranking is determined primarily by teleological considerations, it makes sense to say that the resulting modal is a 'teleological modal.' But this does not mean that the ranking is purely teleological and that no

23 These generalizations do have exceptions; cf. *You might try to chill the gazpacho before you serve it* and *Your passport ought to be in the drawer of the desk in the library*

other considerations play any role at all. It only means that the dominant flavor is teleological — there may also be a slight hint of the deontic there as well.[24]

To illustrate the basic point here, consider the seemingly straightforward sentence (7).

(7) To get to Harlem, you have to take the A train.

This is a paradigmatic example of a teleological modal, and it seems reasonable to suggest that the outer domain is circumstantial (we ignore worlds where the New York subway system is different) and the ranking is teleological (we rank worlds according to the efficiency with which the addressee achieves her goal of getting to Harlem). Yet, even in a case like this one, deontic considerations can play a certain role.

For example, suppose it turned out that you could get to Harlem very quickly by boarding a different train, taking the conductor hostage at gunpoint, and demanding that the train be rerouted to a different track. There seems to be some way in which this option is ruled out, which is why the sentence can come out true. Now, one might suggest that we ignore this possibility because we assume that this way of getting to Harlem wouldn't fit the agent's larger, unstated goals. But ask yourself, would your answer change at all if it turned out that the agent was a hardened criminal who has no concern whatsoever about taking hostages? If you think that the sentence would remain true even in this case, you think the inner domain here is shaped in this case by deontic considerations over and above circumstantial and teleological ones.

Once one begins thinking along these lines, it becomes easy to envision a wide variety of impure modals. There might be modals that are predominantly circumstantial but also have a taste of the bouletic, or modals that are predominantly teleological but have a taste of the stereotypical. As a number of researchers have already noted, though certain modals are predominantly deontic, they also come with a taste of the epistemic (Kolodny & MacFarlane 2010, Charlow 2011, Cariani 2011). These possibilities raise a

24 Yalcin (2007) notices that epistemic modals sometimes carry factual information: sentences like *Cheerios may reduce the risk of heart disease* and *Late Antarctic spring might be caused by ozone depletion* carry factual information and they strike us as expressing the results of actual research. He also suggests that that these are used to communicate knowledge of some relevant experts, or perhaps merely possible knowledge of some relevant group of people. The alternative is to say that these are impure epistemic modals whose ranking is fixed by predominantly but not exclusively by information about what is known.

number of interesting questions, but we will be focusing here on just one type of impurity.

Specifically, our claim is that deontic considerations play a role in the interpretation of root modals. Even when one turns to modals that would on the whole be best classified as circumstantial or teleological, one does not find that the inner domain is determined *exclusively* by circumstantial or teleological considerations. There is also a subtle taste of the deontic.

4 The economy of hope

We now have in place the two main ingredients we need to offer an explanation of the experimental data with which we began. First, there is the claim that the judgments in our experiments are modal. Second, there is the idea that even if the modality at work in the interpretation of these expressions is not predominantly circumstantial or teleological, there can still be a certain role for deontic considerations in their interpretation. Our aim now is to put together these two ingredients and use them to explain each of the effects we introduced above.

We will propose natural principles that constrain the inner domain of the modal proxies of our target sentences. In our semantics, the inner domain is determined via the outer domain and the ranking, so if the inner domain is constrained in some way, that constraint must come about by a constraint on one of these two factors. However, we will remain neutral with regard to certain details of implementation. In particular, we will not commit ourselves whether people adhere to the principle responsible for the taste of the deontic in our modal proxies by considering certain *prima facie* irrelevant possibilities relevant or by ranking certain *prima facie* less prominent possibilities as best. We don't even exclude the option that interpretations may not be entirely uniform in this regard.

We aim to explain the majority judgments. Some of the minority judgments presumably arise due to noise, but others may come about because certain speakers are willing to abandon the key principle we postulate. This counts as linguistic deviance but not the kind that should regarded as a mark of incompetence. We suspect that differences among the experimental subjects are similar to mundane cases of disagreement involving quantifier domains. ('— There is no beer. — That's not true, the store across the street is still open! — I didn't say there is no beer in the store; I just said there is no beer here in the apartment. — You said no such thing; what you said is

that there is no beer, *period*. — Oh, you are sooo annoying!') We may have theoretical reasons to adjudicate such disagreements one way or another, but even if we do we are unlikely to insist that one party is in serious error.

4.1 Freedom

Let's focus first on the effect for *forced*. There, our example was *ship*, which we claimed had the modal proxy *ship*^{M}.

ship	The captain was forced to throw the (a) cargo/(b) his wife overboard.
ship^{M}	Given the circumstances, the captain had to throw (a) the cargo/(b) his wife overboard.

On the account of modality discussed in the previous section, the proxy can be understood as quantifying over a certain set of possibilities. (We will use the superscript 'Q' to indicate the quantificational paraphrases of the modal proxies of our target sentences).

ship^{Q}	Among the relevant possibilities compatible with the circumstances, in all the highest ranked ones the captain throws (a) the cargo/(b) his wife overboard.

To capture the intuitive judgments, we need to guarantee that the inner domain (at the world of evaluation) does not include possibilities in which the captain refrains from throwing the cargo overboard but does include possibilities in which the captain refrains from throwing his wife overboard. The former is easy. It seems bizarre even to consider the possibility that the captain might choose not to throw the cargo overboard — given the circumstances, this seems like a rather far-fetched possibility. What we need is a principle that prevents us from handling the wife and the cargo analogously.

Note that the issue here is not whether the captain would ignore a possibility but rather whether we who are evaluating the sentence would do that. Even if the captain is a hardened wife-hating psychopath, the possibility of that he might not throw his wife overboard will still be deemed pertinent by most people confronted with the scenario. What matters is not the probabilistic claim that he was in any way likely to do otherwise but the deontic claim that his actual behavior violated a moral rule.

Let's now say that a possibility is *hopeful* in a context just in case it is a possibility where none of the events at issue are salient norm-violations in the context. We propose the following principle:

HOPE The inner domain (at the world of evaluation) contains a hopeful possibility.[25]

Hope accounts for the contrast in *ship*^{Q}. Murder is a salient norm-violation even in the context of discussing what to do in a life-threatening storm, while destruction of property is not.[26] *Hope* ensures that in the context of the vignettes some possibilities where the captain refrains from throwing his wife overboard are in the inner domain and so *ship/wife*^{Q} comes out as false. By contrast, *Hope* makes no predictions about *ship/cargo*^{Q}. It is judged true presumably because possibilities where the captain refuses to toss the cargo overboard are deemed sufficiently far-fetched in the circumstances. They are excluded from the outer domain, and consequently are not in the inner domain either.

You can think of *Hope* as a reflection of an *ought implies can* principle.[27] Salient norm-violations ought not to occur and — if the inner domain is bound to contain a possibility where they don't — salient norm-violations can fail to occur. Accordingly, the principle runs into difficulties when it comes to genuine dilemmas. Suppose the captain has a choice: he can save the boat by throwing any one of the passengers overboard. Is he forced to do so? As long as *Hope* is in effect, we predict that people will say no. Perhaps they imagine that there is a way to avoid the killing and still save the ship even if this is ruled out explicitly. Within the semantic framework we are working with

25 *Hope* establishes a connection among different features determined by the same context — the events at issue that are salient norm-violations, the outer domain, and the ranking. In this regard, it is similar to the principle that says that the speaker of the context is located at the place and time of the context. This principle is responsible for *I am here now* coming out as true in any context. Similarly, *Hope* ensures that *This did not have to happen* comes out as true in any context where *this* refers to a contextually salient norm-violation. Just as there are cases when *I am not here right now* comes out intuitively true (think of answering machines) there are also cases when it appears we can truthfully say *This crime had to happen* (think of defense attorneys). Such cases may be handled by allowing special contextual features to override the relevant principles at the level of what is said, or perhaps at the level of what is communicated.

26 Note that the claim is not that the captain does not violate a salient norm when he throws the cargo overboard. He surely does, which is why it is proper for him to deliberate before doing so. But given the circumstances, what he chooses to do is not a salient norm-violation.

27 We thank Tad Brennan for this observation.

this would amount to adjusting the outer domain: focusing on an otherwise irrelevant possibility and thereby including it in the outer domain (relative to the world of evaluation); since the possibility is high-ranked, it will be part of the inner domain (relative to the world of evaluation) as well. Alternatively, people might deem possibilities where the captain kills no one and saves no one better than possibilities where he kills one and saves everyone else. Within the semantic framework we are working with this may amount to adjusting the ranking. We suspect that a hearer who conforms to *Hope* has a choice between these two options, although further features of the context may rule out one or the other. But there are limits to this — if there is no way to make the hopeful possibility highest ranked among the relevant options, many will abandon *Hope* and interpret the sentence in violation of the principle.

Thinking of *Hope* in this manner means rejecting the classical idea that interpretation of context-sensitive sentences invariably yields a determinate proposition relative to any context of utterance. Never mind — the classical idea was never anything more than a useful idealization anyway. When someone utters *It is cold here* there is nothing in the surrounding situation or the mental life of the speaker and hearer that would single out a fully precise region such that the sentence expresses the proposition that it is cold *exactly* at that region. When someone utters *All the beer is gone* there is no fact that would determine for each quantity of beer in the universe what would have to be the case regarding it so that it would count as being among the beer quantified over. There is therefore nothing out of the ordinary in the indeterminacy of interpretation we envision due to the fact that *Hope* can be adhered to by adjusting either the outer domain or the ranking, and due to the fact that whichever of these two options we choose, there remains a further choice as to how exactly the adjustment is to be performed. How people *actually* handle *Hope* is an important question for psychology, how they *should* handle it is an important question for ethics. We take no stance on either. What we claim is that people do not normally abandon *Hope* and this fact can explain the results in this experiment.

The status of *Hope* is similar to the rules Lewis (1996) introduces in his work on knowledge ascriptions (*Rule of Actuality*, *Rule of Attention*, *Rule of Reliability*, etc.). Like those, *Hope* is a principle that governs the domain of possibilities we quantify over when we make overt or covert modal claims. Lewis's rules turned out to be making some pretty bad predictions, so they tend to be rejected today even by those who are largely sympathetic towards

his general outlook.[28] This may well happen in time to *Hope* as well, which would be unfortunate but not terribly so. What matters for us is that there be *some* principle that privileges possibilities where certain norm-violations do not occur. This is what breaks the apparent symmetry in the *ship* example.

4.2 Causation

Let's turn now to the effect for *cause*. There, our example was *pen*, which we claim has the modal proxy pen^{M}, which in turn has the quantificational paraphrase pen^{Q}:

pen	a) The professor/(b) The assistant caused the problem.
pen^{M}	Given the action of (a) the professor/(b) the assistant, the problem had to arise.
pen^{Q}	Among the relevant possibilities compatible with (a) the professor's/(b) the assistant's action, in all the highest ranked ones the problem arises.

The experimental data show that people agree with the claim that the professor caused the problem but disagree with the claim that the assistant caused the problem. The task now is to explain this asymmetry in terms of the inner domain generated by the outer domain and the ranking. More specifically, we need to show that relative to the world of evaluation this domain contains no possibility where the professor takes a pen and the problem fails to arise, but contains a possibility where the assistant takes a pen and the problem fails to arise.

The latter seems easy: surely it could have happened that the professor refrains from taking a pen, in which case whether or not the assistant takes one there would have been no problem at the desk. But the former is a puzzle: even if the professor does take a pen, the assistant might still refrain from taking one, in which case, again, the problem would not arise. *Hope* alone is clearly no help — it guarantees that relative to the world of evaluation the inner domain includes certain worlds but what we need here is a guarantee that it *fails* to contain certain worlds. To provide an explanation we need further resources.

28 Section 6 of Stanley 2005 discusses some particularly acute problems with the *Rule of Belief* and the *Rule of Actuality*.

Our strategy will be to make use of an approach that has proven helpful in numerous other areas: an appeal to economy. The economy principle is motivated by the fact that in assessing modal claims the domain has to be surveyed, which takes genuine cognitive effort:[29]

ECONOMY The outer domain (at the world of evaluation) is the smallest one that satisfies the largest number of principles.

There are presumably a variety of principles governing relevance. Some of these say that certain possibilities are irrelevant — e.g. other things being equal far-fetched options are excluded from the outer domain. Others do the opposite, ensuring certain possibilities are not ignored — e.g. other things being equal, possibilities that are explicitly mentioned are included in the outer domain. These are first-order principles because they tell us whether possibilities of a certain type are relevant or not. *Hope* is one of the first-order domain principles. But first-order principles by themselves cannot fix the outer domain. They can and often do come into conflict with one another, and when they don't, they typically severely underdetermine what relevant possibilities there are. So, we need some meta-principles in addition. *Economy* is one of these; it says that we should select the smallest outer domain that satisfies the largest number of principles.

It should now be possible to see, at least in broad outline, how one might explain the majority judgments in the pen vignette. The action of the professor is a salient norm-violation, so by *Hope*, the domain must include at least one possibility in which it does not occur. For this reason, we include in the inner domain a possibility in which the professor does not take a pen. Given all we are told about the situation in the vignette, we know this must be a world where the receptionist does not run out of pens, so *pen/assistant*^{Q} comes out false. However, there appears to be no principle mandating that we include in the domain a possibility in which the administrative assistant does not take a pen. Thus, by *Economy*, some of these possibilities are eliminated from the domain, which leads to the prediction that *pen/professor*^{Q} is true.

Let's see how this works in detail. We have three binary choices — whether the professor takes a pen, whether the assistant takes a pen, and whether

29 Note that the principle essentially relies on distinguishing between the inner and outer domains and could not even be formulated in a theory that treats domain restriction on modals as dependent on a single contextual parameter. Thanks to Kai von Fintel pressing us on the need to abandon the overly simple semantic framework we employed in an earlier version of the paper.

the problem at the desk arises. These generate eight types of possibilities:

	w1	*w2*	*w3*	*w4*	*w5*	*w6*	*w7*	*w8*
Professor takes pen	+	+	+	+	–	–	–	–
Assistant takes pen	+	+	–	–	+	+	–	–
Problem at the desk	+	–	+	–	+	–	+	–

While the vignette does not *say* this, it *strongly suggests* that four of these eight possibilities (*w2*, *w3*, *w5*, and *w7*) are irrelevant, i.e. not included in the outer domain. This is so because if any of these worlds is actual the vignette is intuitively incomplete. Take *w2*. It is indeed possible for both the professor and the assistant to take a pen and for the receptionist still having one left to take a note. Perhaps she has a secret stash of pens in her drawer which she regularly relies on in cases of emergency. Alas, on the Monday morning described she inexplicably found her drawer empty, which is why she was unable to take the message. Of course, if all this is true it is decidedly odd that the vignette is silent about this crucial detail. Assuming that the vignette is not misleading, there cannot be a secret stash and we can rule out *w2* as a relevant option. In *w3*, the secretary runs out of pens even though the assistant declines to take one. What happens to the pen the assistant took in the world the vignette describes? Perhaps it ran out of ink and the receptionist discarded it before she received the phone call. But then we should wonder why the vignette fails to mention this. So again, assuming the vignette is not unduly reticent, *w3* must be an irrelevant possibility. Similar broadly Gricean considerations rule out *w5* and *w7* as well.

Of the remaining possibilities *w1* must be surely be included in the inner domain on the grounds that it is the one that according to the vignette occurs. *w6* and *w8* are the only remaining ones that are hopeful, i.e. where the professor does not take a pen. According to *Hope*, at least one of them must be included in the inner domain and since there appears to be no principled basis to select one over the other, presumably both are in. Finally, *w4* is not hopeful and — assuming there is no further principle that requires that we take it into account — by *Economy* it is irrelevant; i.e. it is not in the outer domain. Thus, the domain consists of *w1*, *w6* and *w8*, which means that *pen/professor*$^{\mathrm{Q}}$ comes out true and *pen/assistant*$^{\mathrm{Q}}$ false.[30]

30 It is worth noting that the explanation we provided is completely neutral with respect to the question as to whether the outcome itself is a salient norm-violation. In this particular case, the outcome is something bad (the receptionist having a problem), but the explanation would

One may well be suspicious of *Economy* on the grounds that it too easily falsifies possibility claims. Take, for example, the sentence *Given the action of the professor, it could still be that the assistant doesn't take a pen*. Intuitively, this is true in the context of our vignette. But if *Economy* allows us to exclude *w4* from the outer domain, we should predict that it is false.[31] Above we said that *w1* must be included in the outer domain because according to the vignette, it is the actual possibility. Arguably, this was an appeal to a more general first-order principle, according to which possibilities that are mentioned or otherwise raised to salience cannot be ignored. *Economy* does not allow us to ignore possibilities willy-nilly: we can shrink the outer domain only as far as satisfying the maximum number of first-order domain principles permits. *Given the action of the professor, the problem had to arise* does not talk about what the assistant does, but *Given the action of the professor, it could still be that the assistant doesn't take a pen* definitely does. This is why *w4* can be ignored in interpreting the former but not in interpreting the latter.

But doesn't this then show that applying *Economy* is, after all, a mistake? Here is a way to bring out the worry. Consider someone who has the majority intuition in the *pen* case. Now imagine that after this person insists that given the action of the professor the problem at the desk had to arise, we *remind* him that the assistant might have refrained from taking a pen. It is fairly obvious that this person will acknowledge that this is indeed a real possibility, and that if the assistant does not take a pen then the problem won't arise, even if the professor does. Won't he then have to *retract* his earlier claim that given the action of the professor the problem at the desk had to arise? And if he does retract it, doesn't that just show that he was indeed in error?

---

go through in exactly the same way even if the outcome had not been bad at all. We can still exclude *w2*, *w3*, *w5*, and *w7* from the outer domain on Gricean grounds; we must still include *w1* in the inner domain because it is actual; we get still have only *w6* and *w8* as the only hopeful possibilities; and we can still eliminate *w4* an economy grounds. This neutrality is an important virtue of the explanation. Suppose we modified our case in such a way that the outcome ended up being something good (e.g., it turned out to be extremely helpful that there were no pens on the desk). The theory now generates the seemingly paradoxical prediction that the person who acted wrongly will specifically be singled out as the cause of the good outcome. In fact, that is precisely the result obtained in experimental studies using cases of this form (Hitchcock & Knobe 2009).

31 Thanks to Josh Dever for the observation. Brian Weatherson and Stewart Shapiro have also raised a similar point in conversation.

Here, as elsewhere, we are skeptical of the force of such arguments. Following Lewis (1996), we believe that certain possibilities can be properly ignored even if they cannot be properly ignored while attending to those very possibilities. This assumption introduces instability in our modal judgments, but instability needn't be understood as a manifestation of cognitive carelessness. The situation is not that different from how ordinary quantificational domains are handled. When I say that all the bottles are empty and you remind me of a bottle you left in the car but could get in no time, I may well retract my earlier claim. And I may retract it even if the bottle was not among the ones I originally quantified over. It is one of the relevant bottles *now* and that is what counts.

As our example illustrates, *Hope* and *Economy* are not the only principles governing domain selection. We relied on the substantive principle that a possibility raised to salience must be relevant and on the meta-principle which forbids making arbitrary distinctions among possibilities. Presumably, a variety of other principles are also at work in domain selection.

4.3 Intentionality

The effect for attributions of intentionality was illustrated with *rifle*, which has the negative modal proxy $rifle^{M}$ paraphraseable as $rifle^{Q}$:

rifle	Jake hit (a) the bull's-eye/(b) his aunt intentionally.
$rifle^{M}$	It is not the case that given the fluke, Jake had to hit (a) the bull's-eye/(b) his aunt.
$rifle^{Q}$	It is not the case that among the relevant possibilities compatible with the fluke, Jake hit (a) the bull's-eye/(b) his aunt in all the highest ranked ones.

What needs explaining is why the inner domain (relative to the world of evaluation) contains no possibility where the fluke occurs and Jake fails to hit the bull's-eye (making *rifle/bull's-eye* false) but does contain a possibility where the fluke occurs but Jake fails to hit his aunt (making *rifle/aunt* true).

In the bull's-eye case, intuitions are easy to explain. We normally don't consider possibilities which diverge from what happens at a time earlier than the occurrence of events we explicitly hold fixed. Thus, relevant worlds where the fluke occurs are worlds where events unfold as they do according

to the vignette, which means that in all such worlds Jake aims at the bull's-eye. Then, if the fluke occurs only a miracle could prevent Jake hitting the bull's-eye. Thus *rifle/bull's-eye*^{Q} is true.

The aunt case is different. Since it was morally wrong of Jake to aim at his aunt, the inner domain must contain a hopeful possibility, where he fails to do so. Thus, *Hope* forces consideration of possibilities that diverge from the scenario described before the fluke happens. Once these possibilities are at play there will definitely be possibilities where Jake fails to hit his target despite the occurrence of the fluke.

Let us now spell this argument out in more detail for the case of the aunt. Think of the shooting scenario as involving three events: Jake aiming at his aunt, the fluke, and Jake's aunt being hit. The first of these is a salient norm-violation, the second is not; their joint occurrence is sufficient for the third. Here is the possibility chart:

	w1	*w2*	*w3*	*w4*	*w5*	*w6*	*w7*	*w8*
Jake aims at his aunt	+	+	+	+	–	–	–	–
Jake's hand slips, gun stays on target	+	+	–	–	+	+	–	–
Jake's aunt is hit	+	–	+	–	+	–	+	–

w1 is in the inner domain because it has been raised to salience in the vignette. *w2* is a world where Jake's aunt is miraculously escapes being hit despite being shot at from a gun that is on target; *w5* and *w7* are worlds where she is hit even though Jake doesn't even aim at her. These are certainly far-fetched possibilities, so they are outside of the outer domain. *w3* is a world where Jake aims at his aunt, the fluke does not occur, and he hits her. *w4* is similar, except that Jake misses his aunt. Unless there are independent reasons to consider them, *w3* and *w4* are excluded from the outer domain on grounds of economy. Of course, in this case there may well be independent reasons to consider *w3* and *w4*: the fluke is an extremely unlikely event, so perhaps possibilities where it does not occur cannot be ignored willy-nilly.

Be that as it may, the key point is that *w6* and *w8* are in the inner domain. Jake's actual behavior is a salient norm-violation, and *Hope* says that we need to include in the inner domain at least one possibility in which this norm is not violated. Both *w6* and *w8* are hopeful, and choosing between them would be arbitrary. We thereby arrive at an inner domain that includes *w6* — a possibility in which the fluke does occur but Jake does not hit his aunt. For this reason, we predict that *rifle/aunt*^{Q} is true.

4.4 Remarks on explanatory strategy

Perhaps it will be helpful here to pause for a moment and reflect on the general explanatory strategy we have been pursuing. In characterizing the modals under discussion here, one might have expected to find a simple and exceptionless rule that would easily handle all cases. We believe such a rule does not exist. Accordingly, we have adopted a somewhat different explanatory strategy. We have posited a heterogeneous set of principles that, together, purport to account for people's intuitions in these cases. This explanatory strategy is then, by its very nature, an open-ended one. Although we have described certain principles here it should be clear that there are numerous others still to be described.

At this point, one might well complain that our explanatory strategy gives us too much wiggle room. That is, one might say: 'These principles you introduce never make any definite, falsifiable predictions. Whenever a potential counterexample comes up, you can always wiggle out of it by positing a new principle or by manipulating the set of initial possibilities.'

This complaint is in one way accurate and in another completely misguided. It is true that our explanatory strategy allows us to escape a certain kind of burden. Since we claim that the determination of the inner domain of certain modals cannot be captured by a simple rule, we do not take on the burden of giving a single rule that will capture the data in all cases. But at the same time, we take on another burden that earlier accounts have shirked. When we are trying to explain the data about, say, *cause*, we cannot introduce ad hoc principles that apply just to this one expression. Instead, we are forced to explain all of the data in terms of general principles that will have testable implications for theories about the inner domain of modals.

5 Conclusion

The aim of this paper was to present an explanation for the impact of normative considerations on people's assessment of certain seemingly purely descriptive matters. A number of experiments in the last few years have shown that people's judgments about whether an action was free or forced, whether it caused a certain outcome, and whether it was performed intentionally often depend on whether the action violates a norm. The explanation we provided is unified and charitable: we argued that there is a common core of the phenomenon and that these judgments are not in error.

The explanation is based on two main claims. First, a large category of expressions of prime philosophical concern are contextually equivalent to modal proxies. Second, natural language modals can have impure flavors: the inner domain over which they quantify is shaped by heterogeneous considerations, including normative ones.

The present study suggests something surprising about the relationship between people's judgments about how things are and their judgments about how things ought to be. Hume famously claimed that it is "altogether inconceivable" that a proposition where the subject is connected to the predicate by an *ought* or an *ought not* could be derived from propositions where the connections are made by an *is* or an *is not* (*Treatise* 3.1.1.). While many philosophers would dispute that the chasm is that deep, it is received view that normative and descriptive considerations usually are, and always should be sharply distinguished from one another. If morality impacts our sound judgment about matters of freedom, causation, and intentionality, the received view is called into question. The challenge is not whether we can coherently draw the line between the normative and the descriptive. It is, rather, whether the distinction we know and cherish is as deeply rooted in ordinary thinking as it is often assumed. If our explanation of the phenomena is on the right track the answer to this question appears to be negative.

References

Adams, Fred & Annie Steadman. 2004. Intentional action in ordinary language: Core concept or pragmatic understanding? *Analysis* 64(2). 173–181. http://dx.doi.org/10.1093/analys/64.2.173.

Alicke, Mark. 2008. Blaming badly. *Journal of Cognition and Culture* 8(1–2). 179–186. http://dx.doi.org/10.1163/156770908X289279.

Bach, Kent. 1994. Conversational impliciture. *Mind and Language* 9(2). 124–162. http://dx.doi.org/10.1111/j.1468-0017.1994.tb00220.x.

Cappelen, Herman & Ernest Lepore (eds.). 2005. *Insensitive semantics*. Oxford: Blackwell Publishing Ltd. http://dx.doi.org/10.1002/9780470755792.fmatter.

Cariani, Fabrizio. 2011. *Ought* and resolution semantics. *Noûs*. http://dx.doi.org/10.1111/j.1468-0068.2011.00839.x.

Charlow, Nate. 2011. What we know and what to do. *Synthese*. http://dx.doi.org/10.1007/s11229-011-9974-9.

Cushman, Fiery, Joshua Knobe & Walter Sinnott-Armstrong. 2008. Moral appraisals affect doing/allowing judgments. *Cognition* 108(1). 281–289. http://dx.doi.org/10.1016/j.cognition.2008.02.005.

Driver, Julia. 2008. Attributions of causation and moral responsibility. In Walter Sinnott-Armstrong (ed.), *Moral psychology volume 2: The cognitive science of morality: Intuition and diversity*. Cambridge, MA: MIT Press.

Egan, Andy, John Hawthorne & Brian Weatherson. 2005. Epistemic modals in context. In Gerhard Preyer & Georg Peter (eds.), *Contextualism in philosophy: Knowledge, meaning, and truth*, 131–168. Oxford: Oxford University Press.

Falkenstien, Kate. forthcoming. Explaining the effect of morality on intentionality of lucky actions: The role of underlying questions. *Review of Philosophy and Psychology*.

Gillies, Anthony. 2010. Iffiness. *Semantics and Pragmatics* 3(4). 1–42. http://dx.doi.org/10.3765/sp.3.4.

Guglielmo, Steve & Bertram F. Malle. 2010. Enough skill to kill: Intentionality judgments and the moral valence of action. *Cognition* 117(2). 139–150. http://dx.doi.org/10.1016/j.cognition.2010.08.002.

Guglielmo, Steve & Bertram F. Malle. 2011. The timing of blame and intentionality: Testing the moral bias hypothesis. unpublished ms.

Halpern, Joseph & Judea Pearl. 2005. Causes and explanations: A structural-model approach – part i: causes. *The British Journal for the Philosophy of Science* 56(4). 843–887. http://dx.doi.org/10.1093/bjps/axi147.

Hitchcock, Christopher. 2007. Prevention, preemption, and the principle of sufficient reason. *Philosophical Review* 116(4). 495–532. http://dx.doi.org/10.1215/00318108-2007-012.

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

Hume, David. 2000. *Treatise of human nature*. Oxford: Oxford University Press.

Knobe, Joshua. 2003. Intentional action in folk psychology: An experimental investigation. *Philosophical Psychology* 16(2). 309–324. http://dx.doi.org/10.1080/09515080307771.

Knobe, Joshua. 2010. Person as scientist, person as moralist. *Behavioral and Brain Sciences* 33(04). 315–329. http://dx.doi.org/10.1017/S0140525X10000907.

Knobe, Joshua & Ben Fraser. 2008. Causal judgment and moral judgment: Two experiments. In Walter Sinnott-Armstrong (ed.), *Moral psychology vol.*

2: The cognitive science of morality: Intuition and diversity. Cambridge, MA: MIT Press.

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

Kratzer, Angelika. 1977. What 'must' and 'can' must and can mean. *Linguistics and Philosophy* 1(3). 337–355. http://dx.doi.org/10.1007/BF00353453.

Kratzer, Angelika. 1981. The notional category of modality. In Hans-Jurgen Eikmeyer & Hannes Rieser (eds.), *Words, worlds, and contexts*, 38–74. Berlin: de Gruyter.

Kratzer, Angelika. 1991. Modality. In Arnim von Stechow & Dieter Wunderlich (eds.), *Semantics: An international handbook of contemporary research*, 639–650. Berlin: de Gruyter.

Lewis, David. 1973. Causation. *Journal of Philosophy* 70(17). 556–567. http://dx.doi.org/10.2307/2025310.

Lewis, David. 1981. Ordering semantics and premise semantics for counterfactuals. *Journal of Philosophical Logic* 10(2). 217–234. http://dx.doi.org/10.1007/BF00248850.

Lewis, David. 1996. Elusive knowledge. *Australasian Journal of Philosophy* 74(4). 549–567. http://dx.doi.org/10.1080/00048409612347521.

Lewis, David. 2000. Causation as influence. *Journal of Philosophy* 97(4). 182–197. http://dx.doi.org/10.2307/2678389.

Mele, Alfred & Paul Moser. 1994. Intentional action. *Noûs* 28. 39–68. http://dx.doi.org/10.2307/2215919.

Nadelhoffer, Thomas. 2004. The Butler problem revisited. *Analysis* 64(3). 277–284. http://dx.doi.org/10.1093/analys/64.3.277.

Nadelhoffer, Thomas. 2005. Skill, luck, control, and intentional action. *Philosophical Psychology* 18(3). 341–352. http://dx.doi.org/10.1080/09515080500177309.

Nadelhoffer, Thomas. 2006. Bad acts, blameworthy agents, and intentional actions: Some problems for juror impartiality. *Philosophical Explorations* 9(2). 203–219. http://dx.doi.org/10.1080/13869790600641905.

Nanay, Bence. 2010. Morality or modality? What does the attribution of intentionality depend on? *Canadian Journal of Philosophy* 40(1). 25–39. http://dx.doi.org/10.1353/cjp.0.0087.

Nichols, Shaun & Joseph Ulatowski. 2007. Intuitions and individual differences: The Knobe effect revisited. *Mind and Language* 22(4). 346–365. http://dx.doi.org/10.1111/j.1468-0017.2007.00312.x.

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

Pollock, John L. 1976. The possible worlds analysis of counterfactuals. *Philosophical Studies* 29. 469–476. http://dx.doi.org/10.1007/BF00646329.

Portner, Paul. 2009. *Modality*. Oxford: Oxford University Press.

Roxborough, Craig & Jill Cumby. 2009. Folk psychological concepts: Causation. *Philosophical Psychology* 22(2). 205–213. http://dx.doi.org/10.1080/09515080902802769.

Schaffer, Jonathan & Zoltán Gendler Szabó. forthcoming. Epistemic comparativism. *Philosophical Studies*. http://www.jonathanschaffer.org/comparativism.pdf.

Soames, Scott. 1986. Incomplete definite descriptions. *Notre Dame Journal of Formal Logic* 27(3). 349–375. http://dx.doi.org/10.1305/ndjfl/1093636680.

Sousa, Paulo & Colin Holbrook. 2010. Folk concepts of intentional action in the contexts of amoral and immoral luck. *Review of Philosophy and Psychology* 1. 351–370. http://dx.doi.org/10.1007/s13164-010-0028-x.

Sripada, Chandra Sekhar & Sara Konrath. 2011. Telling more than we can know about intentional action. *Mind and Language* 26(3). 353–380. http://dx.doi.org/10.1111/j.1468-0017.2011.01421.x.

Stalnaker, Robert. 1980. A defense of conditional excluded middle. In William L. Harper, Robert Stalnaker & Glenn A. Pearce (eds.), *Ifs: Conditionals, belief, decision, chance, and time*, 87–104. Dordrecht, NL: Reidel.

Stanley, Jason. 2005. *Knowledge and practical interests*. Oxford: Oxford University Press. http://dx.doi.org/10.1093/0199288038.001.0001.

Stanley, Jason & Zoltán Gendler Szabó. 2000. On quantifier domain restriction. *Mind and Language* 15(2-3). 219–261. http://dx.doi.org/10.1111/1468-0017.00130.

Stephenson, Tamina. 2007. Judge dependence, epistemic modals, and predicates of personal taste. *Linguistics and Philosophy* 30. 487–525. http://dx.doi.org/10.1007/s10988-008-9023-4.

Swanson, Eric. 2011. On the treatment of incomparability in ordering semantics and premise semantics. *Journal of Philosophical Logic* 40. 693–713. http://dx.doi.org/10.1007/s10992-010-9157-z.

Sytsma, Justin, Jonathan Livengood & David Rose. 2011. Two types of typicality: Rethinking the role of statistical typicality in ordinary causal attributions. *Studies in History and Philosophy of Science*.

Szabó, Zoltán Gendler. 2006. Sensitivity training. *Mind and Language* 21(1). 31–38. http://dx.doi.org/10.1111/j.1468-0017.2006.00304.x.

Woodward, James. 2003. *Making things happen: A theory of causal explanation*. Oxford: Oxford University Press.

Yalcin, Seth. 2007. Epistemic modals. *Mind* 116(464). 983–1026. http://dx.doi.org/10.1093/mind/fzm983.

Yalcin, Seth. 2011. Nonfactualism about epistemic modality. In Andy Egan & Brian Weatherson (eds.), *Epistemic modality*. Oxford: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780199591596.003.0011.

Young, Liane, Fiery Cushman, Ralph Adolphs, Daniel Tranel & Marc Hauser. 2006. Does emotion mediate the relationship between an action's moral status and its intentional status? Neuropsychological evidence. *Journal of Cognition and Culture* 6(1-2). 291–304. http://dx.doi.org/10.1163/156853706776931312.

Young, Liane & Jonathan Phillips. 2011. The paradox of moral focus. *Cognition* 119(2). 166–178. http://dx.doi.org/10.1016/j.cognition.2011.01.004.

Zalla, Tiziana & Marion Leboyer. 2011. Judgment of intentionality and moral evaluation in individuals with high functioning autism. *Review of Philosophy and Psychology* 2. 681–698. http://dx.doi.org/10.1007/s13164-011-0048-1.

Joshua Knobe
Program in Cognitive Science and
Department of Philosophy
Yale University
P.O. Box 208306
New Haven, CT 06520-8306
joshua.knobe@yale.edu

Zoltán Gendler Szabó
Department of Philosophy
Yale University
P.O. Box 208306
New Haven, CT 06520-8306
zoltan.szabo@yale.edu