Morality constrains the default representation of what is possible

Jonathan Phillips[a,1] **and Fiery Cushman**[a]

[a]Harvard University, Department of Psychology, William James Hall, 33 Kirkland Street, Cambridge, MA 02138

This manuscript was compiled on March 16, 2017

The capacity for representing and reasoning over sets of possibilities, or modal cognition, supports diverse kinds of high-level judgments: causal reasoning, moral judgment, language comprehension, and more. Prior research on modal cognition asks how humans explicitly and deliberatively reason about what is possible but has not investigated whether or how people have a default, implicit representation of which events are possible. We present three studies that characterize the role of implicit representations of possibility in cognition. Collectively, these studies differentiate explicit reasoning about possibilities from default implicit representations, demonstrate that human adults often default to treating immoral and irrational events as impossible, and provide a case study of high-level cognitive judgments relying on default implicit representations of possibility rather than explicit deliberation.

Modality | High-level Cognition | Morality | Norms | Possibility

Human thought often involves alternative possibilities: how things could have been, might be, or may turn out. In other words, it involves *modal* cognition, a capacity to construct and reason over sets of non-actual events. Past research demonstrates a role for modal cognition in assessing causal responsibility [1–3], making moral judgments [4–6], inferring what others intend to communicate [7–9], and much more [10].

To date, research on modal cognition has not distinguished between implicit and explicit representations of possibility. In fact, past methods have overwhelmingly asked how we *explicitly* and *reflectively* represent and reason about different possibilities. We now know a great deal about how and when people engage in explicit counterfactual reasoning, or reasoning about about what else would have happened if a given alternative event had occurred (see [11] for a recent review). There is also a small but growing body of research on how human adults deliberate and decide whether a given event is possible or impossible [12]. In addition, there is an emerging literature on the neural substrates recruited when participants are instructed to engage in episodic counterfactual reasoning or simulation of possible future events [13, 14].

This research has been important for building an understanding of the way that human adults explicitly represent and reason about particular possibilities. But are these same processes recruited when humans make high-level judgments that are known to involve representations of possibilities? There is reason for doubt. Causal and moral judgments, for example, are often made quickly and effortlessly, and appear early in human development [15, 16]. Thus, these judgments are unlikely to rely exclusively on explicit deliberation about, or simulation of, alternative possibilities. Moreover, deliberative judgments about possibility often provide a poor fit to the patterns of high level judgments they supposedly inform (see, e.g., [17] on the dissociation between causal judgments and explicit counterfactual reasoning).

In light of this, an intriguing possibility is that people rely on a different kind of modal representation—one that is rendered quickly, automatically, and based on a distinctive set of constraints. This could reflect a dissociable "system", or instead a graded contrast between default and deliberative judgments rendered by a common system. In either event, the aim of the present study is to test for a form of non-deliberative modal cognition, identify its signature properties, and determine its influence on higher-level cognition.

Critically, two independent research programs point toward a candidate signature property of the default representation: the relevance of prescriptive norms, such as morality and rationality. This stands in contrast to prior research on adult's deliberative reasoning about the (im)possibility of an event, which shows sensitivity to *descriptive* norms, such as the likelihood of the event occurring (see, e.g., [12]), but not *prescriptive* norms (e.g., [18, 19]). Thus, for instance, we tend to think 'It isn't possible for Lewis to open the safe' when it is highly unlikely he will guess the combination (a descriptive feature). In contrast, we tend not to think 'It isn't possible for Lewis to take the money' when it is sitting right in front of him, but belongs to somebody else (a prescriptive feature).

Remarkably, however, circumstantial evidence indicates that less explicit modal representations may be sensitive to both prescriptive and descriptive norms. That is, our default assessment of Lewis's possible actions may exclude what would be wrong in the same way that it excludes what would be unlikely.

One source of evidence comes from studies of young children. While children, much like adults, say that events that violate

Significance Statement

As humans, we think not only about what is, but also what could be. These representations of alternative possibilities support many important cognitive functions, such as predicting others' future actions, assigning responsibility for past events, and making moral judgments. We perform many of these tasks quickly and effortlessly, which suggests access to an implicit, default assumption about what is possible. What are the default features of the possibilities that we consider? Remarkably, we find a default bias towards representing immoral or irrational actions as being impossible. Although this bias is diminished upon deliberative reflection, it is the default judgments that appear to support higher-level cognition.

JP and FC conceived of and designed the studies together. JP conducted the studies and analyzed the data with input from FC. JP drafted an initial version of the manuscript, which was then revised jointly.

The authors declare that they have no conflicts of interest.

[1]To whom correspondence should be addressed. E-mail: phillips01@g.harvard.edu

DRAFT

physical laws cannot happen (e.g., eating lightning), young children also say that events that violate moral or social norms cannot happen (e.g., stealing candy) [20–24]. Not only do they judge these kinds of immoral events are 'impossible', they even say that such events would require 'magic'[18, 19].

Additionally, some contemporary models of high-level cognition in adults argue that the underlying representation of possibility must be sensitive to both descriptive and prescriptive norms. This feature has played a critical role, for example, in understanding human causal reasoning [25, 26], modality in natural language [27], and judgments of force and freedom [28], among others [10]. These models do not, however, provide evidence for how this modal representation is computed.

In sum, this evidence suggests that there may be a non-deliberative and early-emerging form of modal cognition that incorporates information about both descriptive and prescriptive norms. According to this hypothesis, the default representation would support not only judgments of what is 'possible', but also diverse modal judgments concerning what 'ought', 'should', 'might', 'may' or 'could' be done. Moreover, the default representation would serve as input to diverse high-level cognitive judgments. We pursue these hypotheses in three sets of studies.

Study 1: Deliberative vs. default representations of possibility

To investigate whether there are differences in more and less deliberative forms of modal cognition, we began by focusing specifically on judgments of possibility, asking participants to make judgments of whether various events were 'possible' or 'impossible', while manipulating the amount of time available for this judgment. These judgments were made in the context of some background information, such as (1):

(1) Josh is on the way to the airport to catch a flight for a hunting safari in Africa. He leaves with plenty of time to make it there, but his car breaks down on the highway. Now Josh is sitting in his car near a busy intersection, and knows he needs to get to airport soon if he is going to catch his flight.

Participants read six different background contexts, and after each one they were shown a series of candidate events one at a time. As each event was presented, participants pressed a key to indicate whether they thought that the event was 'possible' or 'impossible'. Crucially, participants were either forced to make these judgments very quickly ($\leq 1500ms$) or were asked to reflectively deliberate on the possibility of each event ($\geq 1500ms$). We used a total of 144 different events, which were designed to fall into five categories: 48 ordinary events that did not violate any norms, e.g., (2-a), 24 events that violated statistical norms, e.g., (2-b), 24 events that violated physical laws, e.g., (2-c), 24 events that violated moral rules, e.g., (2-d), and 24 events that violated norms of rationality, e.g., (2-e). (The a priori categorization of these events were confirmed in independent ratings of the morality, probability, and rationality of each of the events, see *SI: Event Ratings.*)

(2) Is it possible or impossible for Josh to...
   a. hail a taxi at the intersection.
   b. fix his car by banging on it.
   c. teleport himself to the airport.
   d. sneak onto public transportation.
   e. sell car for ride to airport.

This paradigm uses judgments of whether an event is 'possible' as an indicator of whether participants represent that event as being available in the context. By changing the amount of time participants have to deliberate, we ask whether (and how) participants' representation of the set of available events changes as a result of more or less deliberation.

Results. We first used a series of linear mixed-effects models to test whether participants' judgments of the possibility of the five different kinds of events differed across time pressure/delay conditions (see *SI: Statistical Approach*). We observed an effect of whether participants were reflectively deliberating, $\chi^2(1) = 20.653$, $p < .001$, and an effect of event-type $\chi^2(4) = 860.48$, $p < .001$. Critically, however, these main effects were qualified by a deliberation $\times$ event-type interaction, $\chi^2(4) = 64.093$, $p < .001$. We decomposed this interaction using a series of generalized linear models. These revealed that deliberation did not significantly affect participants' judgments of the possibility of *ordinary* events ($z = -0.164$, $p = 0.87$), or *improbable* events ($z = -1.056$, $p = 0.291$). While not predicted, participants more tended to judge that events that violated physics were impossible after deliberating ($M = 85.22$, $SD = 17.44$), than when forced to answer quickly ($M = 79.37$, $SD = 15.82$), ($z = -6.628$, $p < .001$). Critically, however, we found that participants' judgments of the possibility of immoral and irrational events was affected by deliberation in the opposite direction. They tended to judge that immoral events were impossible when they were not able to reflectively deliberate ($M = 37.30$, $SD = 30.21$), but more judged them to be possible after deliberating ($M = 21.41$, $SD = 31.26$), ($z = 5.423$, $p < .001$). Similarly, they tended to judge that irrational events were impossible more when they were not able to deliberate ($M = 42.10$, $SD = 25.89$), than after deliberating ($M = 37.86$, $SD = 30.89$), ($z = 1.963$, $p < .05$) (see **Fig.** 1).

In short, we find evidence that default representations of what is possible differ from more reflective representations in that prescriptive norms selectively constrain the default representation of possibility. Moreover, two key contrasts suggest an especially strong role for moral norms: judgments of the possibility of immoral events were more affected by deliberation than either improbable events $\chi^2(1) = 28.035$, $p < .001$, or irrational events $\chi^2(1) = 24.937$, $p < .001$, despite the fact that all of these events involved clear norm violations.

Study 2: Does this default generalize across modal judgments?

Linguistically, different modal auxiliaries (e.g., 'might' vs. 'ought') are known to select different sets of possibilities [7, 8]. All of these modal auxiliaries are alike, however, in that they select some set of contextually relevant possibilities. This is also true in non-linguistic cognition: we think differently about what a person 'could' or 'might' do than we do about what they 'should' or 'ought' to do. In all of these cases though, these judgments require us to reason over the set of events that are represented as being available in the context. Accordingly, we can ask whether distinct modal judgments are supported by distinct default sets of possibilities, or whether instead they share a common set of default possibilities and only become differentiated after additional processing.

DRAFT

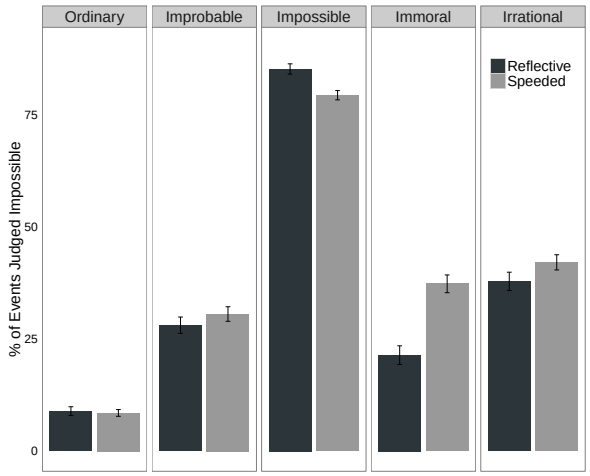

Fig. 1. Judgments of impossibility for five different kinds of events when participants made judgments after deliberating (dark bars) or without time to deliberate before responding (light bars). Error bars indicate +/- 1 SEM.

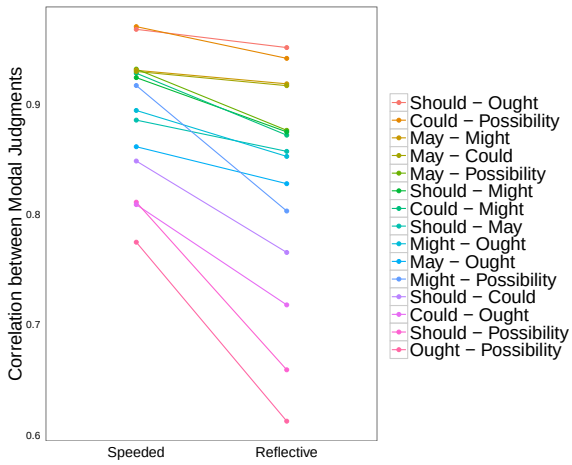

Fig. 2. Correlation coefficient of each pair of modal questions, either when participants were forced to answer quickly (left) or when they were given time to reflect before answering (right)

We propose that there is a common mechanism responsible for rapidly constructing a default representation of what is possible, which supports diverse cognitive functions. This proposal predicts that any deviation from the default representation will require additional processing, and thus the selection of *different* sets of possibilities for different modal judgments is likely the result of more deliberative cognition. This prediction is testable: we can ask whether the correlation between judgments of what a person 'could' do, 'might' do, 'ought' to do, etc. is higher when people do not have time to reflect (and thus more rely on a default representation of possibility), and lower when people do have time to reflect before making a modal judgment.

An obvious alternative prediction is that reducing the amount of time that participants have to respond should increase the noise observed in their responses. This predicts that modal judgments will instead be less correlated with one another when they have to be made extremely quickly (compared to when they are made with time to reflect). We hypothesized that this effect would be outweighed by the influence of a common default modal representation.

Motivated by the results of Experiment 1, we further hypothesized that the default representation underlying diverse modal judgments would be sensitive to prescriptive norm violations. In other words, we hypothesized that under time pressure participants' judgments of what a person could do, may do, might do, etc. would be sensitive to whether that action is immoral or irrational. This common sensitivity to prescriptive norms should thus be largely responsible for heightened similarity among the default representations elicited by diverse modals.

Following the basic procedure introduced Experiment 1, we collected deliberative vs. non-deliberative judgments of what agents 'may' do (Study 2a), 'might' do (Study 2b),'could' do (Study 2c), 'ought' to do (Study 2d), or 'should' do (Study 2e). We also included judgments of what it is 'possible' for them to do (from Study 1). These different modal terms are known to express different kinds of modality (metaphysical, circumstantial, deontic), and the overall patterns of responses demonstrate that participants tracked these differences (see *SI: Differences across modal auxiliaries*)

Results. Our analyses treat the 144 different events as the primary unit of analysis. For each of the six modal judgments, we calculated the proportion of the time that the modal judgment was accepted vs. rejected for each event, doing so independently for trials when participants had time to reflect and when they did not. These scores reflect the likelihood that an event will be included in the set of possibilities relevant to each type of modal concept (e.g., what should be done, what could be done, and so on). We then calculated the correlation between the events' representations for each possible pair of different modal judgments, both when participants were given time to reflect before answering and when they were not.

As hypothesized, modal judgments were significantly more correlated when participants did not have time to reflect ($M_r = 0.893$, $SD_r = 0.060$) than when they did ($M_r = 0.830$, $SD_r = 0.101$), $t(14) = 5.131$, $p < .001$, $d = 0.754$ (see Fig. 2). This finding supports the view that participants relied on a common default representation of the set of available events when answering quickly. By contrast, when participants reflected on each judgment, their responses were more strongly dictated by unique features of each modal concept, and thus became less correlated with one another.

We next asked whether a principal dimension on which default judgments became more similar was sensitivity to prescriptive norms. To do this, we divided the events that participants judged into two categories: those for which descriptive norms were primarily relevant (the improbable and physically impossible events), and those for which prescriptive norms were primarily relevant (the irrational and immoral events). We then again calculated the correlation between each pair of modal judgments, first focusing on events where descriptive norms were relevant, and then second focusing on events where prescriptive norms were relevant. This analysis allows us to ask whether it was the presence of one of these kinds of norms that was responsible for the difference in correlations when participants reflected before answering.

Analyzing these correlations with linear mixed effects mod-

DRAFT

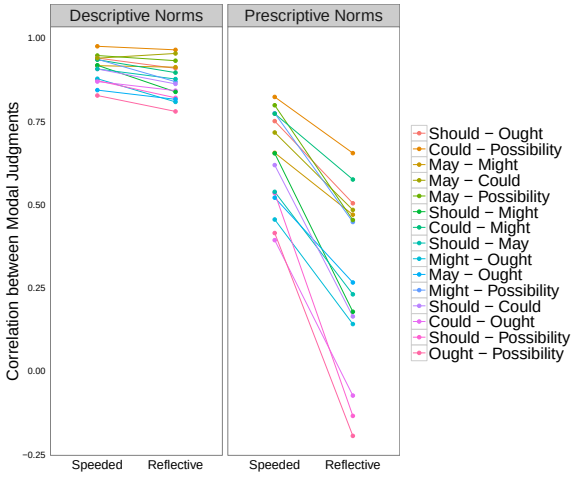

Fig. 3. Correlation coefficient for each pair of modal judgments either for items where descriptive norms were primarily relevant (left panel) or for items where prescriptive norms were primarily relevant (right panel). Correlations between modal judgments made without time to reflect are plotted on the left-hand side of the panel; correlations between modal judgments made after reflecting are plotted on the right-hand side of the panel.

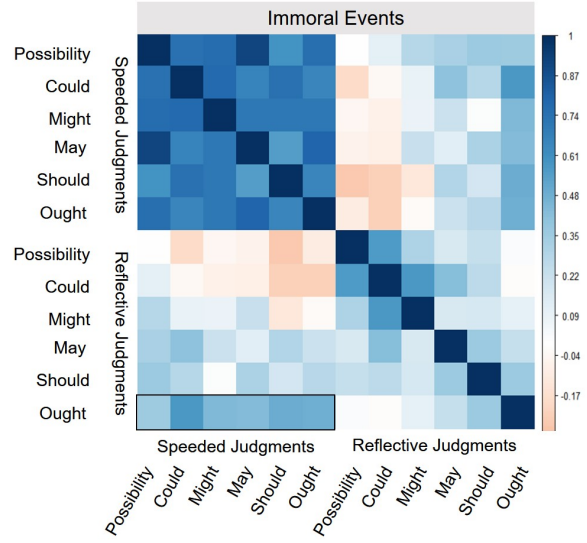

Fig. 4. Graphical depiction of the correlation matrix for speeded and reflective modal judgments of events that involved immoral actions. Squares within the black box depict correlations of speeded modal judgments with reflective judgments of 'ought'.

els, we found an interaction effect between the type of norms that were relevant and whether or not participants had time to reflect, $\chi^2(1) = 24.323$, $p < .001$. Decomposing this interaction, we first focused on modal judgments made in scenarios for which descriptive norms were primarily relevant (e.g., winning the lottery). Here, we observed a modest decrease in the correlation between the modal judgments when participants reflected before answering ($M_r = 0.873$, $SD_r = 0.055$) from when they answered before reflecting ($M_r = 0.908$, $SD_r = 0.042$), $t(14) = 5.336$, $p < .001$, $d = 0.724$. Next, we focused on judgments made in scenarios for which prescriptive norms were relevant (e.g., theft) and found a much larger decrease in the correlations between the modal judgments made when participants reflected before answering ($M_r = 0.278$, $SD_r = 0.265$), as compared to when they answered before reflecting ($M_r = 0.628$, $SD_r = 0.145$), $t(14) = 8.841$, $p < .001$, $d = 1.642$ (see Fig. 3). A corollary finding is that deontic modals show less of a difference between deliberative and reflective responses than metaphysical or circumstantial modals (see *SI: Within-modal correlations*).

Together, these analyses indicate that prescriptive norms constitute a principal dimension differentiating specific modal concepts from their common default. In contrast, the use of descriptive norms appears to be relatively more consistent across distinct modal concepts, even upon reflection. This pattern is most evident for moral norms. Focusing specifically on the immoral events, which clearly distinguish different kinds of modal reasoning, we created a correlation matrix of all of the different judgments. This approach revealed that all of the different modal judgments were highly correlated when participants were not able to reflect ($0.558 \leq r \leq 0.919$), but were not when participants took time to reflect ($-0.016 \leq r \leq 0.572$) (see Fig. 4). A corollary finding is that participants' non-reflective modal judgments are all highly correlated with *reflective* judgments of what agents 'ought' to do ($r_{Mean} = 0.465$), but not other reflective judgments ($-0.091 < r_{Mean} > 0.264$). Put simply, when answering quickly, modal judgments become more ought-like (see box in Fig. 4).

In summary, we have demonstrated thus far that default modal judgments show a common sensitivity to both descriptive and prescriptive norms. In contrast, reflective modal judgments can deviate away from this default, allowing for some kinds of modal cognition (e.g., thoughts of what 'could' happen) to focus primarily on what is physically possible, without regard to immorality or irrationality.

Study 3: Default modal representations in high-level cognition

Existing evidence suggests that the modal representations used in other domains of high-level cognition—for instance causal attribution, language comprehension and moral judgment—are similar to the default representations identified in Experiments 1 and 2. For instance, causal judgments of an event are sensitive to the moral status of alternative events (see, e.g., [25, 29]), which is consistent with the demonstrated role of prescriptive norms in default modal representations. This current evidence is indirect, however, and the role of alternative events in causal cognition remains controversial (see, eg., [30]). Accordingly, Study 3 provides a direct test of the hypothesis that the default representations of the set of events available in a given context are exported to other cognitive domains.

To do this, we used judgments of 'force' as a case study. This is an ideal test because empirical [28, 31] and theoretical [32, 33] studies concur that judgments of 'force' rely on representations of alternative possibilities. Put simply, if an agent was 'forced' to act in some way, this implies that there was no relevant alternative; and if she was not forced, this implies that some alternative existed. We therefore asked: when people decide whether an agent was 'forced' to do something, do they rely on a set of possibilities more like a deliberative set, or instead the implicit set that we identified in Study 1?

Participants in this study once again read the six different background contexts used in the previous studies (see (1)). However, in this case, the agent asked another person for

DRAFT

advice about what to do. This person then directly suggested that the agent pursue one of the ordinary, immoral, irrational or improbable events used previously (2-a)-(2-e), as in (3).

(3) Josh calls his father who lives a few states away and tells him about about his problem. Not really knowing how to help him, Josh's father makes a suggestion. His father says Josh could *sneak onto public transportation.*

Participants were then told that, notwithstanding this advice, the agent decided to pursue some specified alternative course of action, as in (4).

(4) Josh ignores his father's suggestion and decides to book the next available flight, even though it is quite expensive.

Within each context, the agent was always described as pursuing the same course of action, regardless of the advice given.

After reading all of this information, participants were asked whether they agreed or disagreed that the agent was forced to do the action he or she actually pursued, as in (5).

(5) Josh was forced to book the next available flight.

Following previous work on judgments of force [28, 32, 33], participants should only judge the agent to have been forced if they believe that there were not other options available to the agent—in other words, if the set of relevant alternative possibilities was empty. Thus, in the context of the specific advice given to the agent, participants should only agree that the agent was forced if they did not represent the proposed option as an available possibility.

Crucially, in all cases, the judgments of force were made with unlimited time available. Thus, our question is whether *deliberative* judgments of force for each scenario are better predicted by deliberative or speeded judgments of possibility for the same scenario in Study 1. We hypothesized that force judgments would be best predicted by speeded possibility judgments, reflecting a dominant influence of the default representation of the set of available events.

Results. We computed the average agreement that the agent was forced to do a given action when the alternative proposed was one of the events used in the previous studies. We then first asked whether participants' non-reflective judgments were predictive of judgments of force in a simple linear model, and found that they were, $F(1, 118) = 46.12$, $p < .001$, $\eta^2 = 0.281$ (see **Fig**. 5). Next, we asked whether reflective or non-reflective judgments of possibility were a better predictor of judgments of force by comparing a series of linear mixed-effects models. We found that including reflective judgments of possibility did not significantly improve a model that already included non-reflective judgments of possibility, $\chi^2(1) = 0.030$, $p = .863$. In contrast, including non-reflective judgments of possibility did significantly improve a model that already included reflective judgments of possibility $\chi^2(1) = 27.640$, $p < .001$. Thus, above and beyond reflective judgments of what is possible, the default representation of possibility is predictive of deliberative judgments of force.

We next asked whether morality had an effect on judgments of force (as found in previous work [28, 31]) and whether this effect was mediated by non-reflective judgments of possibility. We found that this was the case: participants judged agents to be more forced when the proposed alternative was immoral vs. ordinary, $t(70) = 7.309$, $p < .001$, $d = 1.827$. Moreover, this effect was mediated by non-reflective judgments of possibility, 95%CI [0.018,0.542], $p = 0.04$. A similar pattern of results was also observed for events that involved violations of rational norms ($t(70) = 5.462$, $p < .001$, $d = 1.366$; 95%CI [0.045,0.864], $p = 0.04$). These results fit well with ongoing work that offers a semantics for 'force' according to which prescriptive norms impact the set of actions available to an agent [34]. (See *SI: Study 3b* for a test of an alternative semantics for 'force' that proposes to account for the impact of norms through inferences about the agents' desires.)

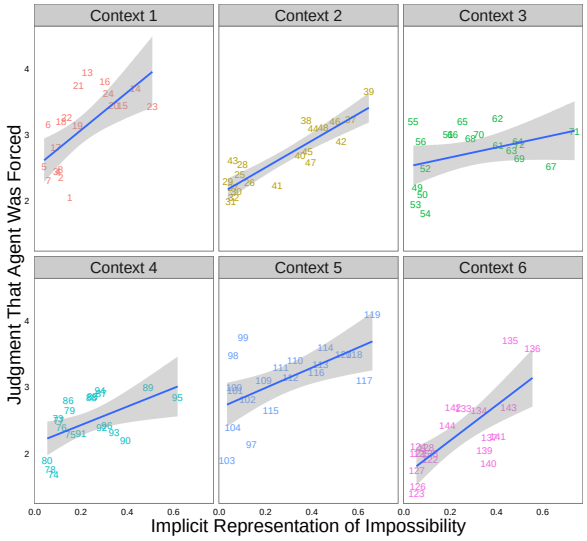

Fig. 5. Relationship between the implicit representation of possibility and high-level judgments of whether or not an agent was forced to do a given action in six different contexts. Shaded areas represent 95% confidence intervals.

Discussion

We find evidence for a default representation of what is "possible"–i.e., a set of events, specific to a context, that constitutes a starting point for understanding non-actual alternatives. A hallmark of this default representation is that it tends to exclude immoral actions. First, we found that time pressure makes people less likely to judge it possible to act immorally or, to a lesser extent, irrationally. Second, we found that this effect is reflected across diverse modal concepts: what a person ought to do, may do, could do, should do, or might do. These judgments appear to share a common default basis, and become more differentiated only after deliberative processing. Finally, as a case study of the role of implicit modal cognition in reflective high-level judgments, we found that default modal judgments are better predictors than reflective modal judgments of people's decisions about whether a person was 'forced' to act in a particular way. A mediation analysis indicates that, to the extent that we are "forced" to act morally, it is because our default representation of what is possible tends to exclude immoral actions.

Although all forms of modal cognition involve constructing and reasoning over sets of possibilities, it is remarkable that they apparently begin with a common default template. By analogy, all forms of social cognition—whether about in-groups

DRAFT

or out-groups, cooperators or competitors—involve reasoning about sets of other people, but it would be remarkable to discover that under time pressure, all forms of social cognition begin with a representation of the very same individuals. In modal cognition, much like in social cognition, one might have initially thought that even rapid categorizations of what it is 'possible' for a person to do and what a person 'ought' to do would depend on separate processes that pick out separate events. After all, most people would say that it is *possible* to run a red light at an empty intersection, but not that you *ought* to. In contrast to this intuition, our findings suggest that it takes time to realize that it is possible to run a red light. We begin with a default representation of possibility that tends to excludes this action, along with other immoral or irrational acts.

The purpose of this default mechanism is a key area for further study. One appealing possibility is that the default representation serves the function of proposing actual candidate actions during decision-making—in other words, it helps us to construct a choice set [35]. It is natural to suppose that screening out immoral or irrational actions from one's own decision-making would tend to be adaptive.

Importantly, the default representation we have uncovered bears a striking resemblance to the pattern of possibility judgments observed early in human development [18, 19, 21, 24]. In both cases, judgments of possibility are sensitive not only to descriptive norms but also to prescriptive ones. The similarity of these patterns presents the intriguing possibility that the modal representation observed early in development is retained in adult cognition alongside a separate later-developing capacity for deliberatively reasoning about possibilities. Future work should continue to explore this connection directly, and ask which other factors serve to constrain both of these representations of possibilities.

Finally, it is worth noting a potential connection between our research and the widespread, surprising presence of moral norms in disparate corners of the human mind: e.g, in causal reasoning and mental state attribution [10, 25, 26, 29, 36]. These processes also involve reasoning over a set of possibilities. Given the present results, an exciting possibility is that morality shapes how we think about many things because it constrain the very possibilities that come to mind.

Materials and Methods

Participants. All participants were recruited from Amazon Mechanical Turk (www.mturk.com) through TurkPrime (www.turkprime.com), which was used to prevent repeat participation across the studies. Sample sizes and demographic information is as follows. Study 1a (judgments of possibility), 498 participants ($M_{age} = 33.85$, $SD_{age} = 10.82$, 233 females). Study 1b (event ratings), 61 participants ($M_{age} = 31.53$, $SD_{age} = 7.87$, 25 females). Study 2a (ought), 301 participants ($M_{age} = 33.40$, $SD_{age} = 9.87$, 147 females). Study 2b (might), 201 participants ($M_{age} = 33.79$, $SD_{age} = 10.20$, 92 females). Study 2c (could), 304 participants ($M_{age} = 34.87$, $SD_{age} = 11.32$, 144 females). Study 2d (may), 200 participants ($M_{age} = 32.80$, $SD_{age} = 9.21$, 86 females). Study 2e (should), 297 participants ($M_{age} = 34.05$, $SD_{age} = 9.86$, 153 females). Study 3 (force judgments), 400 participants ($M_{age} = 38.05$, $SD_{age} = 12.73$, 218 females). Study 3b (desire judgments), 400 participants ($M_{age} = 36.61$, $SD_{age} = 11.80$, 200 females). These studies were approved by the Harvard University Institutional Review Board, IRB14-2016, and informed consent was acquired from all participants.

Access to data, analysis code, and materials. All materials used to conduct these studies and analyze the results are available at: github.com/phillipsjs/implicitModality. Studies 1a-b and 2a-e were conducted through Testable (www.testable.org) and Study 3a-b were conducted in Qualtrics. Stable links to experiments are as follows. Study 1a-b: possibility: testable.org/t/3111b018, event ratings: testable.org/t/31041bbc; Study 2a-e: ought: testable.org/t/31092d3e, might: testable.org/t/3120b96c; could: testable.org/t/31bbed30, may: testable.org/t/31f78b37, should: testable.org/t/317a2d7c; Study 3a-b: force: goo.gl/6POVuY, desire: https://goo.gl/30OjG1.

ACKNOWLEDGMENTS. This research was supported by Grant N00014-14-1-0800 from the Office of Naval Research.

1. Lewis D (1973) *Counterfactuals*. (John Wiley & Sons).
2. Pearl J (2009) *Causality*. (Cambridge university press).
3. Gerstenberg T, Tenenbaum JB (in press) Intuitive theories in *Oxford Handbook of Causal Reasoning*, ed. Waldmannn M. (Oxford University Press).
4. Kane R, ed. (2011) *The Oxford handbook of free will*. (Oxford University Press).
5. Hart HLA, Honoré T (1985) *Causation in the Law*. (Oxford University Press).
6. Lagnado DA, Gerstenberg T, Zultan R (2013) Causal responsibility and counterfactuals. *Cognitive Science* 47:1036–1073.
7. Portner P (2009) *Modality*. (OUP Oxford).
8. Kratzer A (2012) *Modals and conditionals: New and revised perspectives*. (Oxford University Press) Vol. 36.
9. Stalnaker R (2002) Common ground. *Linguistics and philosophy* 25(5):701–721.
10. Phillips J, Luguri J, Knobe J (2015) Unifying morality's influence on non-moral judgments: The relevance of alternative possibilities. *Cognition* 145:30–42.
11. Byrne RM (2016) Counterfactual thought: From conditional reasoning to moral judgment. *Annual review of psychology* 67(1).
12. Shtulman A, Tong L (2013) Cognitive parallels between moral judgment and modal judgment. *Psychonomic bulletin & review* 20(6):1327–1335.
13. Schacter DL, Benoit RG, De Brigard F, Szpunar KK (2015) Episodic future thinking and episodic counterfactual thinking: Intersections between memory and decisions. *Neurobiology of learning and memory* 117:14–21.
14. De Brigard F, Spreng RN, Mitchell JP, Schacter DL (2015) Neural activity associated with self, other, and object-based counterfactual thinking. *NeuroImage* 109:12–26.
15. Hamlin JK, Wynn K, Bloom P (2007) Social evaluation by preverbal infants. *Nature* 450(7169):557–559.
16. Buon M, Jacob P, Loissel E, Dupoux E (2013) A non-mentalistic cause-based heuristic in human social evaluations. *Cognition* 126(2):149–155.
17. Mandel DR, Lehman DR (1996) Counterfactual thinking and ascriptions of cause and preventability. *Journal of personality and social psychology* 71(3):450.
18. Phillips J, Bloom P (2016) Do children believe immoral events are impossible? under revision.
19. Shtulman A, Phillips J (2016) Differentiating *Could* from *Should*: Developmental changes in modal cognition. under revision.
20. Browne CA, Woolley JD (2004) Preschoolers' magical explanations for violations of physical, social, and mental laws. *Journal of Cognition and Development* 5(2):239–260.
21. Kalish C (1998) Reasons and causes: Children's understanding of conformity to social rules and physical laws. *Child development* 69(3):706–720.
22. Kushnir T, Gopnik A, Chernyak N, Seiver E, Wellman HM (2015) Developing intuitions about free will between ages four and six. *Cognition* 138:79–101.
23. Shtulman A (2009) The development of possibility judgment within and across domains. *Cognitive Development* 24(3):293–309.
24. Shtulman A, Carey S (2007) Improbable or impossible? how children reason about the possibility of extraordinary events. *Child Development* 78(3):1015–1032.
25. Halpern JY, Hitchcock C (2015) Graded causation and defaults. *The British Journal for the Philosophy of Science* 66(2):413–457.
26. Kominsky JF, Phillips J, Gerstenberg T, Lagnado D, Knobe J (2015) Causal superseding. *Cognition* 137:196–209.
27. Knobe J, Szabó ZG (2013) Modals with a taste of the deontic. *Semantics and Pragmatics* 6(1):1–42.
28. Young L, Phillips J (2011) The paradox of moral focus. *Cognition* 119(2):166–178.
29. Hitchcock C, Knobe J (2009) Cause and norm. *Journal of Philosophy* 106(11):587–612.
30. Samland J, Waldmann MR (2016) How prescriptive norms influence causal inferences. *Cognition* 156:164–176.
31. Phillips J, Knobe J (2009) Moral judgments and intuitions about freedom. *Psychological Inquiry* 20(1):30–36.
32. Berlin I (1970) *Four essays on liberty*. (Oxford University Press New York) Vol. 969.
33. Aristotle (2002) *Nicomachean ethics*. (Oxford University Press).
34. Mandlekern M, Phillips J (2017) *Force*: Topicalization, context-sensitivity, and morality. submitted.
35. Ben-Akiva M, Boccara B (1995) Discrete choice models with latent choice sets. *International Journal of Research in Marketing* 12(1):9–24.
36. Knobe J (2003) Intentional action in folk psychology: An experimental investigation. *Philosophical Psychology* 16(2):309–325.
37. Jaeger TF (2008) Categorical data analysis: Away from anovas (transformation or not) and towards logit mixed models. *Journal of memory and language* 59(4):434–446.
38. Barr DJ, Levy R, Scheepers C, Tily HJ (2013) Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language* 68(3):255–278.
39. Bates D, Maechler M, Bolker B, Walker S, et al. (2014) lme4: Linear mixed-effects models using eigen and s4. *R package version* 1(7).

DRAFT