Representations of what's possible reflect others' epistemic states

Lara Kirfel (kirfel@mpib-berlin.mpg.de)
Center for Humans and Machines, Max Planck Institute for Human Development

Matthew Mandelkern (mandelkern@nyu.edu)
Department of Philosophy, New York University

Jonathan Phillips (Jonathan.S.Phillips@dartmouth.edu)
Cognitive Science Program, Dartmouth College

Abstract

People's judgments about what an agent can do are shaped by various constraints, including probability, morality, and normality. However, little is known about how these representations of possible actions—what we call modal space representations—are influenced by an agent's knowledge of their environment. Across two studies, we investigated whether epistemic constraints systematically shift modal space representations and whether these shifts affect high-level force judgments. Study 1 replicated prior findings that the first actions that come to mind are perceived as the most probable, moral, and normal, and demonstrated that these constraints apply regardless of an agent's epistemic state. Study 2 showed that limiting an agent's knowledge changes which actions people perceive to be available for the agent, which in turn affects whether people judged an agent as being "forced" to take a particular action. These findings highlight the role of Theory of Mind in modal cognition, revealing how epistemic constraints shape perceptions of possibilities.

Keywords: modal reasoning, modal cognition, force judgment, Theory of Mind, knowledge

Introduction

Chris McCandless was a young adventurer who, inspired by a desire to live simply and escape from societal norms, ventured into the Alaskan wilderness in the early 1990s. As famously told in Jon Krakauer's book "Into the Wild", McCandless spent 113 days living in an abandoned bus in the Alaskan bush. When he decided to return to civilization, the river he had previously crossed had become impassible due to seasonal flooding. Because McCandless only had a basic, out of date map of the region, he was completely unaware that there was an abandoned, hand-operated cable car that crossed the river just $\frac{1}{2}$ mile (800 m) downstream from where he had previously crossed. Considering his options, McCandless decided to head back and re-establish his camp to wait for rescue, with tragic consequences.

Reasoning About What's Possible

Much of human judgment and decision-making – whether judging whether someone is responsible for a negative outcome, determining whether someone was forced to do a given action, or deciding on the best place to have dinner – requires reasoning about the possible actions available to an agent (Byrne, 2016; Phillips & Knobe, 2009, 2018; Phillips et al., 2015). Studies show that throughout human development (Alderete & Xu, 2023; Shtulman & Phillips, 2018), the capacity for representing and reasoning over sets of possibilities supports diverse kinds of high-level judgments. Consider the case of McCandless: Was he forced to head back to his bus? Intuitively, judgments of whether an agent had to take a particular action, or whether they could have done otherwise, reflect the alternative actions perceived as available to them. More precisely, the extent to which we judge that McCandless had no choice but to stay put is directly influenced by the alternative possible actions we represent as available (Mandelkern & Phillips, 2018; Morris et al., 2021; Phillips & Knobe, 2018). Extensive research on *modal reasoning* – the capacity to reason about possibilities – shows that representations of what is possible are highly sensitive to the perceived descriptive and prescriptive normality of actions (Bernhard et al., 2022; Phillips & Knobe, 2018; Phillips & Kratzer, 2024). That is, people tend to judge an agent as having acted freely when there were possible alternative actions available that were more morally acceptable, statistically likely, or normal (Bear & Knobe, 2017).

Modal Space

When reasoning about what's possible, what actions actually come to people's minds? In a recent paper, Hecht and Phillips (2022) develop a method to empirically measure the possible actions that people take to be relevant in a particular situation. By sequentially sampling participants' generated actions that an agent could do for a variety of scenarios, they quantitatively describe the space of possibilities – "the modal space" – that participants perceive to be available in a certain situation. They go on to show that their model of contextual possibility representations successfully predicts higher level judgments, such as judgments about whether the agent was forced to do a given action: the more there were better alternative options available in a contextual modal space, the less likely participants were to agree that the agent was forced to do the actual action they did. Interestingly, this distributional representation of what kinds of options are likely available out-performed participants' own individually generated alternatives when predicting their high-level judgments.

Sensitivity to Theory of Mind

What shapes and constrains the modal space of actions when considering a particular situation, such as that of McCandless? While an agent's physical environment is one factor

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

that influences what options comes to mind, another aspect that has been notably absent in discussions on modal reasoning is the role of Theory of Mind (but see Kirfel & Lagnado, 2021; Kirfel & Phillips, 2021, 2023; Papafragou, 2008) in reasoning about the possible actions of third-party agents.

McCandless' case beautifully illustrates the importance of Theory of Mind for representations of modal space. To elaborate: typically the presence of a nearby hand-operated cable car would shift the shape of the distribution of possible actions people think of as available to McCandless. As a result, staying at the campsite to wait for help might rank low in the set of options available to the agent in this context. And intuitively, if McCandless had known about the cable car, it seems obvious that he wouldn't have had to return to his bus to wait. However, McCandless was, in fact, unaware of the cable car's existence. In this case, the option of using the cable car does not seem obviously relevant. Consequently, returning to his bus to wait may seem like one of the better options available to McCandless, and accordingly, we might also be more inclined to judge him to be forced to do so. Here, we hypothesize that modal reasoning intersects with Theory of Mind, and specifically, representations of what other agents *know*. While being aware that crossing the river via cable car is a generally viable option, people might still not consider it a 'possible action' for the agent due to his ignorance about the cable car. Whether our representations of the knowledge states of others affects representations of modal space remains an open and unexplored question. Will people judge an agent like McCandless less forced to perform a certain action – such as staying behind – given his partial ignorance about aspects of the environment?

Epistemic Shifts in Modal Space

In this study, we investigate whether modal space representations are sensitive to the agent's epistemic state. Specifically, by manipulating aspects of an agent's knowledge of their environment, we test whether modal space representations (of what it is possible for the agent to do) shift in a corresponding manner. In other words, we examine whether limiting an agent's knowledge about their surroundings alters the set of possible actions people consider for that agent. Additionally, we explore whether such shifts in modal space also result in the predicted effects on high-level judgments of force – specifically, whether changes in the agent's knowledge affect the extent to which people judge the agent as being forced to take a particular action.

To address these questions, we adopt the approach developed by Hecht and Phillips, 2022 to empirically measure the set of generated options deemed relevant in a given context. In Study 1, we use Hecht and Phillips (2022)'s methodology to characterize the modal space representation in different epistemic contexts – comparing cases in which an agent has full versus partial knowledge of their situation. In Study 2, we investigate whether varying the agent's epistemic context not only shifts this modal representation, but also also shifts people's judgments of whether the agent was "forced" to do a given action.

Study 1: Possibility Generation

In this study, we want to measure the set of possibilities that people take to be relevant in a given situation. We are particularly interested in characterizing the space of possible actions available to an agent when this agent is aware vs. ignorant about certain aspects of their situation.

Participants We recruited 420 participants (*age*: M = 42.80, SD = 13.43; *gender*: Female = 231, Male = 187, Other = 2; *race*: Asian = 50, Black/African American = 46, Multiracial = 41, White = 327, Other = 17), via Prolific (Palan & Schitter, 2018). Participants were paid $\approx \$12.00/hr$.

Design Our study employed a 2 *epistemic state* (knowledge vs. ignorance; *within*) $\times$ 16 (*scenario*; *within*) design.

Procedure Participants were randomly presented with 6 out of 16 open-ended decision contexts and asked to generate possible actions that could be taken by the agent in each context. The decision contexts ranged in a variety of settings as well as the severity of the agent's decision, e.g., risking missing a flight, being stranded on the open sea, raising funds for a school concert, etc. For each of the 16 scenarios, there was a "knowledge" and an "ignorance" version, only one of which participants were randomly assigned to read. In the knowledge version, the agent has full awareness of the relevant aspects and features of the situation. To illustrate, consider the knowledge version of decision context 5 ("Brad"):

> ***Knowledge:*** Brad and some friends are hiking through the mountains in the Canadian wilderness. A couple of days into their hike, Brad realizes that they are lost. He knows that a rescue crew could arrive before long, but it is extremely cold and they don't have much food or water left. He also knows that they are 5 miles from an inhabited town.

Critically, in this version, Brad knows the fact that there is a town nearby. Now consider, by contrast, the ignorance version of this scenario:

> ***Ignorance:*** Brad and some friends are hiking through the mountains in the Canadian wilderness. A couple of days into their hike, Brad realizes that they are lost. He knows that a rescue crew could arrive before long, but it is extremely cold and they don't have much food or water left. He also incorrectly believes that they are nowhere near civilization; unbeknownst to Brad, they are 5 miles from an inhabited town.

Here, Brad is unaware of the fact that there is an inhabited town nearby – information that might relevant for the range of action options he might consider. Across all scenarios, we varied the agent's knowledge about a specific aspect about their situation or their environment. This information was never directly relevant as a solution to the agent's problem

(e.g. location of rescue team); rather, it would affect what action options the agent could consider (e.g. seeking help from a nearby town). Consequently, this information could also affect the relevance of certain actions, such as staying put.

Participants were asked to generate five actions that could be pursued by the agent in the context:

> "In this situation, what are some things you believe [Agent] could do? Please list 5."

After the participant wrote down five actions in five separate open text fields, we then sequentially re-presented each option to the participant and asked them to evaluate it along the dimensions of probability, morality and normality on a 0 ("least") to 100 ("most") point scale, following Hecht and Phillips (2022).

> *Probability Question*: "How probable is it that [Agent] will do that thing?"

> *Morality Question*: "How morally acceptable would it be for [Agent] to do that thing?"

> *Normality Question*: "How normal would it be if [Agent] did that thing?"

Results

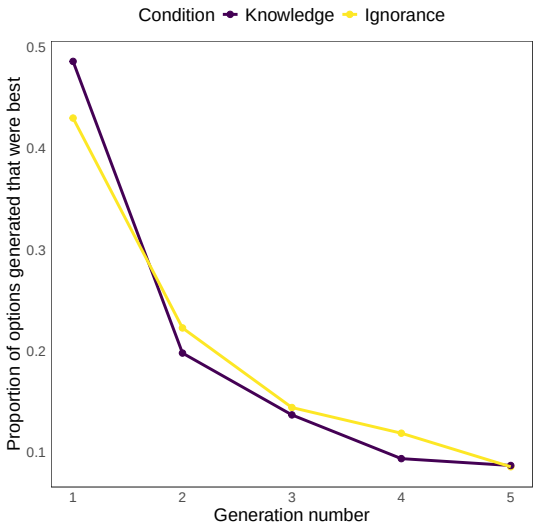

Figure 1: **Proportion of generated actions that were evaluated as best** for each sequential action generation (1-5) split by Knowledge and Ignorance conditions.

Participants' first responses for each scenario tended to be the one they rated highest, or tied for highest, based on a composite score calculated by averaging their *probability*, *morality*, and *normality* ratings. In the knowledge condition, the first action option generated was retrospectively evaluated as the best or tied for the best 49% of the time, compared to 43% in the ignorance condition (Figure 1).[1]

[1]All data, code, and materials are available at https://github.com/LaraKirfel/possibilities_ignorance.

More generally, the earlier an action option tended to be generated by a participant, the more highly it was evaluated. This was the case when participants generated actions in knowledge contexts, $F(1,652.1) = 1207.2$, $p < 0.001$, as well in the contexts in which the agent was ignorant about certain features of the scenario $F(1,7683) = 553$, with $p < 0.001$ (Figure 2). All three rating dimensions, the probability, morality and normality of an action, were independently predictive of how likely this action option was generated early in the generation process. In the knowledge condition, the more probable a generated option was rated, the more likely it was to be generated earlier, $\chi^2(1) = 91.48$, $p < 0.001$ independent of normality and morality. Likewise, the more normal an answer was perceived, the more likely it was generated earlier, independent of the other ratings, $\chi^2(1) = 6.92$, $p < 0.01$. Morality was not found to have an independent predictive effect, $\chi^2(1) = 2.96$, $p = 0.08$. In the ignorance condition, probability ($\chi^2(1) = 30.82$, $p < 0.001$), normality ($\chi^2(1) = 22.36$, $p < 0.001$) and morality ($\chi^2(1) = 5.58$, $p = 0.02$) were all independently predictive of how early an action option was generated in the process.

Discussion

In line with previous research (Hecht & Phillips, 2022), the action option that came to participants' minds were constrained by probability, morality, and normality. We replicated previous studies showing that the first options that came to mind were also considered the most moral, probable, and normal to do (Phillips & Knobe, 2018). Valuable action options came to mind early, with decreasing value the later they arose in the generation process. Controlling for the two other ratings, the later an option was generated, the less likely, moral or normal it was perceived. This was the case for contexts in which the agent knew about all relevant features of their situation or was ignorant about certain aspects. Having shown that the modal space representation follows similar patterns across different epistemic conditions, we now want to test whether the change in the agent's epistemic state results in a change in the set of possible actions participants represented as available to the agent in the context.

The core of our approach is to make use of the fact that judgments of whether an agent was *forced* to do a given action depend on the alternative actions represented as available to the agent. Thus, if the agent's epistemic state affects the possible actions represented as available to the agent, then, even holding fixed the actual action done, this change should result in differences in judgments of whether the agent was forced to do a given action.

Study 2: Epistemic Shifts in Modal Space and Force Judgments

In our second study, we were interested in whether a change in the representation of epistemically-relevant possibilities drives a change in *force* judgments.

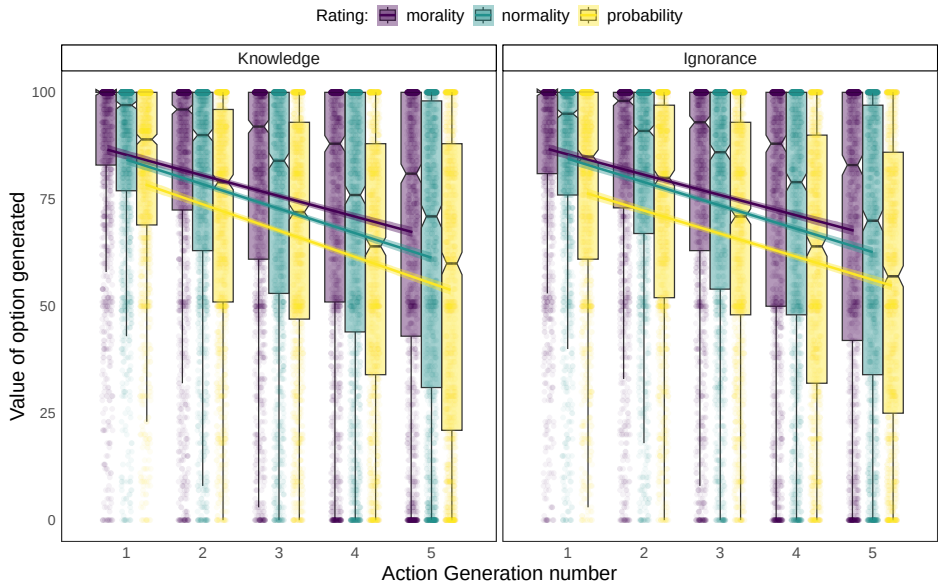

Figure 2: **Action Ratings by Epistemic Condition**: Boxplots of participants' subjective ratings of the morality (violet), normality (green), and probability (yellow) of each generated option they generated as a function of generation number (1-5).

Participants We recruited 401 participants (*age*: M = 43.58, SD = 13.45; *gender*: Female = 168, Male = 184, Non-binary = 8, Undisclosed = 1, *race*: Asian = 27, Alaskan/American Indian = 3, Black/African American = 41, Hawaiian/Pacific Islander = 1, Multiracial = 10, Latinx = 13, White = 263, Other = 10; Undisclosed = 3, demographics of 40 participants are missing due to a programming error), via Prolific. Participants were paid $\approx \$12.00/hr$.

Design

Study 2 employed a 2 *epistemic state* (knowledge vs. ignorance; *within*) $\times$ 16 (*scenario; within*) $\times$ 6 (*action*; *between*) design.

Procedure For Study 2, we created six new actions for each of the 16 contexts. These actions were selected to vary in the extent to which they would be seen as probable, moral, and normal, and thus to be distributed widely across the range of options that naturally come to mind in each context. In order to have actions that spanned beyond the normal modal space for a context, we added actions that were designed to violate both descriptive and prescriptive norms. In our running example of Brad who gets lost in the Canadian wilderness, the six actions were: (1) boost morale with a three-legged race, (2) try to find enough dry wood to start a fire, (3) huddle to conserve energy, (4) attempt to search for a water source, (5) steal all of the food and water, (6) throw away all of the food.

Participants were presented with 8 out of the 16 scenarios used in Study 1. After reading each scenario, participants were randomly presented with 1 out of the 6 actions. They were told that this action was what the agent decided to do (e.g. "Brad decided to try to find enough dry wood to start a fire."). While each scenario was again presented either in the knowledge or the ignorance version, the action remained constant. Participants were then asked to make a force judgment, i.e. to rate the extent to which they thought the agent was forced to perform this action on a 100 point scale, from 0 ("completely disagree") to 100 ("completely agree"):

> *Force Judgment*: How much do you agree with the following statement: "[Agent] had to do [action]."

After having made a force judgment, participants were again presented with the action the agent decided to do and asked to rate it on the same three 100-point scales used in Study 1, measuring the probability, morality and normality of the agent doing that action.

Method To what extent should people think an agent was forced to do a given action? According to Hecht and Phillips (2022), this will depend on their representation of the probability of there being better actions available to the agent. For example, if Brad decided to try to find dry firewood, participants should judge that he did not have to do that action to the extent that they represented there being better alternatives in that situation. To test this hypothesis, we next need to situate the 6 actual decisions for each scenario within the set of contextually-relevant possibilities that participants generated for those scenarios in Study 1.

In order to do this, we first computed the average rating for each of the six actual action in Study 2 across these three dimensions of *probability*, *morality* and *normality*. Then, we wanted to locate that action relative to the space of possible actions that participants generated for that specific context when the agent was in that specific epistemic condition. Think of this as the context-specific epistemic modal space PS^c_e. To do this, we compare the rating of the actual action to the distribution of ratings for other possible actions that people thought of for the corresponding decision context and epistemic condition in the open-ended possibility generation

task in Study 1. We then calculate what Hecht and Phillips (2022) call the "modal distance": for each actual action, we again average the values on the same three dimensions and then estimate the probability that a randomly sampled action from the context-specific epistemic modal space ($a_s \in PS_e^c$) has a higher rating than the actual action, a_a. In other words, $ModalDistance_c(a_a) = P(\text{Value}(a_s \in PS_e^c) > \text{Value}(a_a))$. To compute this, we first sampled with replacement 10,000 possible actions from those participants generated in the knowledge or ignorance version of a decision context from Study 1. Then, for each actual action in that epistemic condition and context in Study 2, we calculated the probability of sampling an action that had a higher value than the actual action. This generated a unique value for each actual action that reflected the probability that a better option would come to mind in that same context for an agent with the same epistemic state.

To illustrate the central idea, consider the example of Brad instead simply deciding to hold a three-legged race. The modal distance for this action will capture the extent to which the distribution of actions in the context-specific epistemic modal space have a higher value than this. Obviously, the same action can have a different modal distance value depending on the context in which the action occurs. For instance, holding a three-legged race might have a high modal distance in the context of a wilderness misadventure (where many better options are possible for that context), but a low distance in the context of a kids' birthday party (where there may not be many actions that are actually better options). Similarly, we propose that an action's modal distance can also vary depending on the agent's knowledge: while holding a three-legged race is unlikely to ever be a high value action in an outdoor emergency, it may not be wildly worse than the other options when the agent doesn't know about the nearby town. However, if the agent is aware that an inhabited town is nearby, it seems like a particularly bad option to have chosen. We use the modal distance measure to model the predicted impact of alternative possibilities on whether an agent had to do a action.

Results

The correlation between modal distance values and force judgments from contexts in which the agent was knowledgeable about their situation was $r = -0.89, p < 0.001$. Likewise, the correlation in the ignorance scenarios was $r = -0.87, p < 0.001$. That is, for both epistemic condition, the more distant the action was from the center, the less people thought the agent was forced to do that action.

More critically, we want to investigate whether force judgments reflect differences in modal space based on the agent's epistemic state. To illustrate the kind of patterns we predict, start by focusing on a couple of simple comparisons. Figure 3 (left panel) shows the distribution of modal distance in scenario 5 ("Brad") for the actions "steal all of the food and water" and "attempt to find a water source", depending on the agent's epistemic state. It also shows people's judgment about whether the agent was forced to do that action (right panel). The action "stealing all of the food and water" is perceived as an overall rather abnormal activity in this scenario. In the knowledge condition (Brad knows about a nearby town), this action has an overall higher modal distance, that is, people perceive more better options than in the ignorance condition. Consequently, people see Brad in the knowledge condition as slightly less forced to steal all food and water, compared to when Brad is ignorant about the town. In contrast, searching for a water source is perceived as an overall normal action, and as a result people judge the agent as much more forced to do so in both conditions.

Generalizing, the key analysis concerns whether we can show that, across all 6 actions for all 16 contexts, judgments of force are uniquely well-predicted by the modal distances that correspond to the agent's actual epistemic state. To do this we used model comparison of linear mixed-effects models. First, we predicted force judgments in the knowledge contexts using either the knowledge modal distance value or ignorance modal distance value for each action. There was no significant interaction between ignorance and knowledge modal distance $\chi^2(1) = 0.68,\ p = 0.41$. As expected, the model performed worse without the modal distance values from the knowledge contexts $\chi^2(1) = 46.69,\ p < 0.001$. However, the model also performed somewhat worse without the modal distance values from ignorance contexts $\chi^2(1) = 6.20,\ p = 0.01$, but the decrease in model fit was notably smaller in the latter case. Second, we also predicted force judgments from the ignorance condition using modal distance values from both epistemic conditions. Once again, no significant interaction between both modal distance terms was found $\chi^2(1) = 1.20,\ p = 0.27$. The model performed significantly worse when knowledge modal distance values were removed $\chi^2(1) = 40.13,\ p < 0.001$. It was also somewhat affected by the removal of the modal distance values from the knowledge contexts, although to a lesser degree, $\chi^2(1) = 4.34,\ p = 0.04$. Figure 4 shows Chi-square values reflect the contribution of modal distance predictors explaining the variance in people's *had to* judgments. We also found that differences in modal distance values for actions between epistemic conditions were significantly correlated with differences in force judgments between epistemic conditions $r = -0.50, p < 0.001$.

Discussion

In Study 2, we examined whether people's modal space representations are sensitive to an agent's knowledge about the world. Specifically, we investigated whether changes in the agent's knowledge about their environment led to shifts in contextual modal space, which, in turn, should result in shifts in the modal distances for the same actual actions. These shifts in modal distance should correspondingly affect judgments of whether or not the agent *had to* do the chosen action. Model comparisons revealed that the modal distance values from the same epistemic context predicted force judgments better than modal distance values from the altered

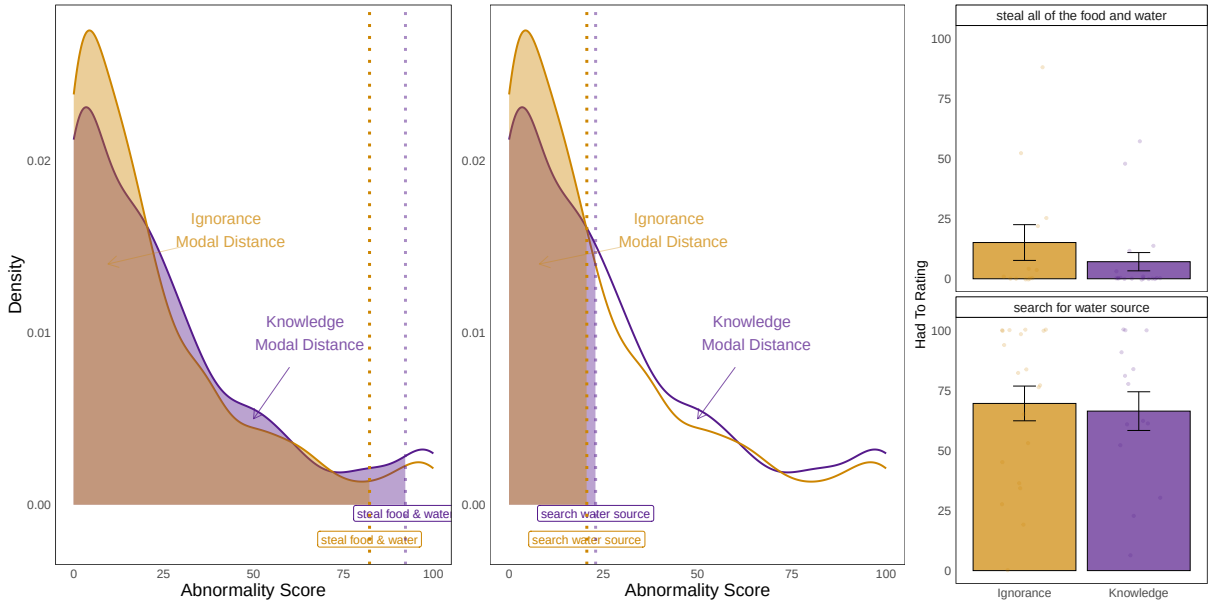

Figure 3: **Modal Distances and Force Judgments for two actual actions, split by Epistemic Condition in Scenario 5**. ***Left Panel:*** Normality Density curve in each epistemic condition (knowledge vs. ignorance) of Scenario 5 ("Brad"). The Abnormality score on the X Axis is the reverse coded value rating (composite score of normality, probability and morality rating). Dotted lines mark the average rating of an action that has been rated as rather abnormal (i.e. high on the X Axis) – "steal food and water" – or as rather normal action "search for a water source" –. The density curve (Y-Axis) to the left of the actual action displays the amount of actions that would have been a "better", i.e. more normal option. ***Right Panel:*** Force judgment for each action, split by epistemic condition.

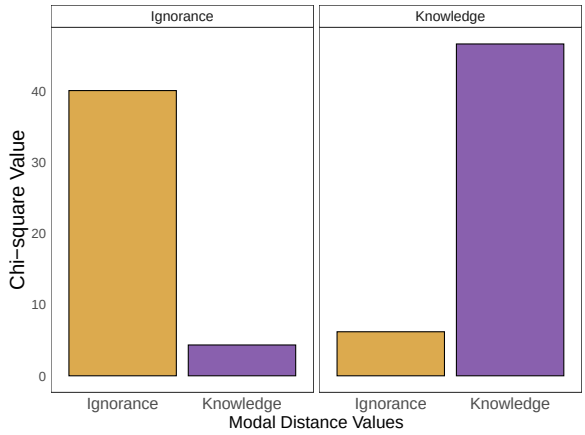

Figure 4: **Chi-square values for model comparisons across epistemic conditions**: Chi-square values from likelihood ratio tests assessing the effect of modal distance predictors on *had to* judgments in different epistemic conditions. The x-axis represents the predictor included in the model.

epistemic context. Differences in force judgments between knowledge and ignorance contexts were predicted by differences in modal distance between the two epistemic contexts. In sum, changes in contextual modal space account for differences in judgments of the same actions across different epistemic conditions of the agent.

General Discussion

The present study provides evidence that modal space representations are sensitive to agents' epistemic states. Across two studies, we demonstrate that manipulating an agent's knowledge about certain aspects of their environment systematically shifts the space of actions people consider to be available for that agent. We showed that restricting the agent's knowledge about their surroundings alters people's modal space representation. Correspondingly, this shift also influences the extent to which people perceive an agent as forced to undertake a particular action. These findings demonstrate that Theory of Mind is a crucial factor in shaping modal space representations. People's representation of an agent's knowledge of the world influences the set of possible actions they consider for that agent, which, in turn, shifts the perceived relevance of a particular action. This has downstream effects on higher level force judgments.

While this work has demonstrated that an agent's epistemic state can affect representations of modal space, an important question for future work is why such an effect takes place. One possibility is that agents' epistemic states can directly shape the set of possible actions represented as available to them in a given context. An alternative possibility is that the agent's epistemic state only has a more indirect effect by changing the probability, normality, or morality of the actions, which then exert their typical influence on representations of modal space. Yet another option is that epistemic constraints might influence high-level judgments via entirely different psychological processes. For instance, epistemic constraints could affect moral judgments about the agent, which themselves influence force judgments (Clark et al., 2014, 2021). Or they might influence what people think is rational from that agent's perspective, which could in turn shape judgments of force (Cusimano et al., 2024). We hope to answer these questions in future work.

References

Alderete, S., & Xu, F. (2023). Three-year-old children's reasoning about possibilities. *Cognition*, *237*, 105472.

Bear, A., & Knobe, J. (2017). Normality: Part descriptive, part prescriptive. *Cognition*, *167*, 25–37.

Bernhard, R. M., LeBaron, H., & Phillips, J. (2022). It's not what you did, it's what you could have done. *Cognition*, *228*, 105222.

Byrne, R. M. (2016). Counterfactual thought. *Annual review of psychology*, *67*(1), 135–157.

Clark, C. J., Luguri, J. B., Ditto, P. H., Knobe, J., Shariff, A. F., & Baumeister, R. F. (2014). Free to punish: A motivated account of free will belief. *Journal of personality and social psychology*, *106*(4), 501.

Clark, C. J., Winegard, B. M., & Shariff, A. F. (2021). Motivated free will belief: The theory, new (preregistered) studies, and three meta-analyses. *Journal of Experimental Psychology: General*, *150*(7), e22.

Cusimano, C., Zorrilla, N., Danks, D., & Lombrozo, T. (2024). Psychological freedom, rationality, and the naive theory of reasoning. *Journal of Experimental Psychology: General*, *153*(3), 837.

Hecht, E., & Phillips, J. (2022). Domain-general modal spaces play a foundational role throughout high-level cognition. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *44*(44).

Kirfel, L., & Lagnado, D. (2021). Changing minds—epistemic interventions in causal reasoning.

Kirfel, L., & Phillips, J. (2021). The impact of ignorance beyond causation: An experimental meta-analysis. *Proceedings of the annual meeting of the cognitive science society*, *43*(43), 1595–1601.

Kirfel, L., & Phillips, J. (2023). The pervasive impact of ignorance. *Cognition*, *231*, 105316.

Mandelkern, M., & Phillips, J. (2018). Sticky situations:'force'and quantifier domains. *Semantics and Linguistic Theory*, *28*, 474–492.

Morris, A., Phillips, J., Huang, K., & Cushman, F. (2021). Generating options and choosing between them depend on distinct forms of value representation. *Psychological Science*, 1731–1746.

Palan, S., & Schitter, C. (2018). Prolific. ac—a subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, *17*, 22–27.

Papafragou, A. (2008). Modality and theory of mind: Perspectives from language development and autism. In *Modality and its interaction with the verbal system* (pp. 185–204). John Benjamins Publishing Company.

Phillips, J., & Knobe, J. (2009). Moral judgments and intuitions about freedom. *Psychological Inquiry*, *20*(1), 30–36.

Phillips, J., & Knobe, J. (2018). The psychological representation of modality. *Mind & Language*, 65–94.

Phillips, J., & Kratzer, A. (2024). Decomposing modal thought. *Psychological Review*, *131*(4), 966.

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

Shtulman, A., & Phillips, J. (2018). Differentiating "could" from "should": Developmental changes in modal cognition. *Journal of Experimental Child Psychology*, *165*, 161–182.