Reason-Based Constraint in Theory of Mind

Corey Cusimano (cusimano@princeton.edu)
Natalia Zorrilla (zorrilla@princeton.edu)
Department of Psychology, Princeton University, Peretsman Scully Hall
Princeton, NJ 08540 USA

David Danks (ddanks@cmu.edu)
Departments of Philosophy and Psychology, Carnegie Mellon University, Baker Hall 161
Pittsburgh, PA 15213 USA

Tania Lombrozo (lombrozo@princeton.edu)
Department of Psychology, Princeton University, Peretsman Scully Hall
Princeton, NJ 08540 USA

Abstract

In the face of strong evidence that a coin landed heads, can someone simply choose to believe it landed tails? Knowing that a large earthquake could result in personal tragedy, can someone simply choose to desire that it occur? We propose that in the face of strong reasons to adopt a given belief or desire, people are perceived to lack control: they cannot simply believe or desire otherwise. We test this "reason-based constraint" account of mental state change, and find that people reliably judge that evidence constrains belief formation, and utility constrains desire formation, in others. These results were not explained by a heuristic that simply treats irrational mental states as impossible to adopt intentionally. Rather, constraint results from the perceived influence of reasons on reasoning: people judge others as free to adopt irrational attitudes through actions that eliminate their awareness of strong reasons. These findings fill an important gap in our understanding of folk psychological reasoning, with implications for attributions of autonomy and moral responsibility.

Keywords: Theory of mind; Autonomy; Belief; Desire; Free will

Introduction

Two boat captains are dispatched to transport cargo to a distant port. Their VHF radios announce that there is a terrible storm brewing. Captain Ahab's vessel is well-maintained, and she knows the local weather projections are often exaggerated. Captain Barbosa, on the other hand, has a poorly-maintained ship, and he knows his weather station is rarely wrong. Each captain forms the belief that they will hit a storm and each forms the desire to avoid it, and so each pilots their vessel back to harbor.

Did Captain Barbosa have a choice about whether to turn back? Did Captain Ahab? Despite their identical reactions, it seems that Ahab could have done otherwise while Barbosa could not (cf. Chernyak, Kushnir, Sullivan, & Wang, 2013; Kushnir, Gopnik, Chernyak, Seiver, & Wellman, 2015; Reeder, 2009; Woolfolk et al., 2006; Young & Phillips, 2011). Put differently, Ahab had greater *control*.

Judgments about an agent's control have important consequences: They affect whether people are held responsible for their behavior, the inferences others draw about their character, and whether they are punished or helped (Malle, Guglielmo, & Monroe, 2014; Martin & Cushman, 2017; Reeder, 2009; Weiner, 1995). If Ahab had greater control over her decision, she is likely to be held responsible for the costs of turning back, while Barbosa is not. But what explains why Barbosa seems to have less control than Ahab? After all, neither of them was *physically* forced to act as they did. We suggest that they seem differently constrained because their situations differently affect what beliefs, desires, and intentions they are free to form.

Here, we take first steps towards a novel explanation of people's judgments of situational constraint. According to the *reason-based constraint* account, people attribute less control to Barbosa than to Ahab in part because Barbosa's reasons support reasoning to only one rational course of action while Ahab has multiple reasons-based actions available to her. Below we discuss how lay perceptions of others' rationality can be marshalled to explain perceptions of situational constraint. We then present an experiment to test our account.

Rationality and Constraint in Theory of Mind

People interpret and predict others' behavior in part by assuming that others are rational (Jara-Ettinger, Gweon, Schulz, & Tenenbaum, 2016). By and large, people assume that others' choices reflect their preferences (Baker, Jara-Ettinger, Saxe, & Tenenbaum, 2017), that beliefs reflect rational inferences based on their evidence (Ross & Ward, 1996), and that people tend to behave efficiently (Gergely & Csibra, 2003). Accordingly, observers only consider a limited range of rational attitudes as plausible explanations for behavior, and only a limited range of rational attitudes and behaviors as likely or predictable reactions. However, these observations about everyday *inference* and *prediction* do not entail anything about how people conceptualize *control*. After all, people could predict that others will almost always eat salad with a fork instead of a toothpick (and infer that most everyone would prefer to) but still believe that someone could easily choose to eat with a toothpick instead. This raises the question: Are beliefs, desires, and intentions – like salad consumption – subject to inviolable voluntary control, or are they in some way constrained?

One account of perceived mental state control is highly permissive: it could be that people view others as free to form whatever beliefs, desires, and intentions they please. On such a view, rationality just affects what observers think is sensible or likely for others to do; it has no bearing on what observers believe those others can, in principle, choose to do. Some scholars have theorized that this is how people conceptualize others' agency in circumstances involving situational constraint (e.g., Kalish, 1998; Reeder, 2009). These views explain away intuitions about Ahab and Barbosa: the intuition that Barbosa is constrained does not reflect a judgment that he is literally unable to do otherwise, but only that he has a narrower range of clearly sensible or rational choices to pick from.

Instead, we hypothesize that in people's theory of mind, rationality is believed to literally constrain and restrict what others can believe, desire, and do. According to this view, when people conceptualize the process by which others' beliefs, desires, and intentions form during reasoning, they conceptualize this process as both (1) *rational* (in line with prior research, as noted above), and also (2) partly *involuntary* in that others cannot help but form mental states that rationally follow in light of their available reasons. On such a view, people believe that, when others are deliberating in reaction to their circumstances, those others cannot simply adopt whatever beliefs, desires, or intentions they please; they are constrained to forming attitudes that plausibly rationally reflect those circumstances. We call this the *reason-based constraint* model.

Two observations speak in favor of people conceptualizing others' mental states as partially constrained by rational reasoning. First, as a matter of psychological fact, environmental conditions do seem to spontaneously cause and constrain the attitudes that others form (Kunda, 1990; Lazarus, 1991). For example, people regularly experience cravings for foods and drugs they otherwise do not want to desire in response to situational cues (such as being exposed to the food or drug; Boswell, Sun, Suzuki, & Kober, 2017; Kober & Mell, 2015). Likewise, people frequently change their beliefs to align with new evidence even when doing so runs counter to their preferred ideology or self-concept (Epley & Gilovich, 2016; Kunda, 1990).

And second, people often have the introspective experience of constraint during reasoning when one choice appears dominant (Cusimano & Goodwin, 2020; Kouchaki, Smith, & Savani, 2018; cf. Wolf, 1980). For instance, Kouchaki et al. (2018) asked participants to make a series of choices between two options of different value. When the choice was between clearly preferable and clearly non-preferable options, people reported experiencing a lack of freedom over their choice. Cusimano and Goodwin (2020) observed a similar result in the domain of belief. When participants thought about the (strong) evidence they held for their beliefs, they reported an inability to form different beliefs. People's experience tracking others' attitudes, as well as their experience with their own belief and desire formation, suggests that, in certain situations, they will treat strong reasons as literal constraints on other's mental states.

In line with the reason-based constraint account, we predict that people expect exposure to certain kinds of information to uncontrollably cause—in line with the demands of rationality—corresponding mental states. For instance, people should expect exposure to strong evidence to uncontrollably affect others' beliefs (in line with that evidence). Similarly, they should expect that information about an outcome's utility will uncontrollably affect others' desire for that outcome (in line with the expected utility). These expectations about mental state change help explain why Ahab seems more free than Barbosa. Ahab's weak evidence for a storm and the minor negative impact a storm would have on her ship provide little constraint on the beliefs and desires she can (rationally) form. Hence, she could believe or desire otherwise, and she is therefore not compelled to return to port. Barbosa is not so lucky: his strong evidence means he cannot help but believe there is a storm coming, and the incredible damage it would do to his boat means he cannot help but want to avoid it.

The current study

To summarize, we propose that, as a component of theory of mind, people possess an intuitive theory of mental state change such that reasons (provided by someone's situation or environment) constrain mental state change through a process of reasoning that is largely rational but also partially outside of voluntary control. Two predictions follow from this proposal:

1. *Reason correspondence*. Participants should judge evidence as constraining belief formation, and utility as constraining desire formation.
2. *Reasoning dependence*. Participants should judge that reason-based constraints impact mental state change only when they cannot help but be aware of those reasons. Thus, participants should judge others as free to form irrational mental states through means that remove awareness of strong reasons.

The current study directly tests these predictions by presenting participants with vignettes that vary the presence of strong or weak evidence, as well as strong or weak changes in utility, for some individual. Participants then reported whether the individual could adopt beliefs or desires that violated the reasons they had access to in that situation. We varied two kinds of control: attempting to change one's mind through reasoning, and attempting to change one's mind by removing awareness of the evidence or utility (by taking a pill that would lead one to forget).

Experiment

Methods

Participants. We recruited 939 adults from Prolific (47% reported male, 51% reporting female, 2% reporting other or non-reporting; mean age = 38 years). An additional 61 participants were recruited but failed at least one (of two) attention checks and so were excluded from the analyses reported below.

Design and vignette construction. Participants read a vignette about a target character who is anticipating some outcome that will either pose a minor or a major threat to their well-being. Additionally, the character in the vignette either has weak or strong evidence that the outcome will be realized. We constructed four vignettes that replicated this design yielding a 2 (Evidence: weak evidence vs strong evidence) x 2 (Utility: bad outcome vs very bad outcome) x 4 (Vignette) between-participants design.

For instance, in the "Storm" vignette, participants read about a captain, Jeremiah, who is sailing his boat and hears on the radio that there is a storm brewing. In the "bad outcome" condition, Jeremiah's boat was recently repaired and outfitted for hard weather; a storm would be unpleasant, but not catastrophic. By contrast, in the "very bad outcome" condition, Jeremiah's boat has not been repaired in a while, and so a storm would be very dangerous. The vignette also varied Jeremiah's evidence about whether there was going to be a storm. In the "weak evidence" condition, Jeremiah is aware that the weather station is unreliable at predicting storms; by contrast, in the "strong evidence" condition, the weather station has a near-perfect prediction record. Within each of the four cells of this design, participants reported how much control Jeremiah had to adopt beliefs and desires that went against the evidence and utility present in the situation.

The remaining vignettes described characters in different situations. In "Scholarship", a college student wants to change whether she believes or desires that she will get a scholarship; in "Advertising", an advertising executive wants to change whether she believes or desires that her recent ad campaign succeeded; and in "Heirloom", a man wants to change whether he believes or desires that a recently recovered heirloom is worth a lot of money.

Procedure. Figure 1 provides a graphical overview of the experimental procedure. At the start of the study, participants were randomly assigned to one of the four Evidence × Utility conditions described above, and to one of four vignettes. Participants read the vignette in full prior to responding to any dependent measures. Embedded in the vignette were two comprehension questions that tested whether participants read and understood important details of the character's situation. For instance, participants who read the Storm vignette were asked whether Jeremiah's boat was recently outfitted for bad weather, as well as whether the weather station was reliable or unreliable. Participants who answered at least one of these two questions incorrectly were excluded.

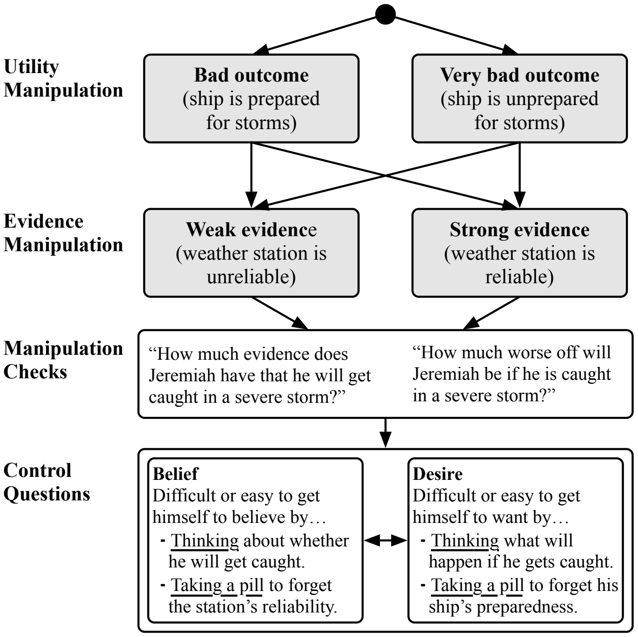
Utility Manipulation
Bad outcome (ship is prepared for storms)
Very bad outcome (ship is unprepared for storms)
Evidence Manipulation
Weak evidence (weather station is unreliable)
Strong evidence (weather station is reliable)
Manipulation Checks
"How much evidence does Jeremiah have that he will get caught in a severe storm?"
"How much worse off will Jeremiah be if he is caught in a severe storm?"
Control Questions
Belief
Difficult or easy to get himself to believe by…
- Thinking about whether he will get caught.
- Taking a pill to forget the station's reliability.
Desire
Difficult or easy to get himself to want by…
- Thinking what will happen if he gets caught.
- Taking a pill to forget his ship's preparedness.

Figure 1. Schematic representation of experimental procedure. Gray boxes represent between-participants experimental manipulations. See text for exact wording.

After reading the vignette, participants responded to four manipulation check questions presented in a random order. Two questions measured perceived utility (e.g., "How much worse off will Jeremiah be if he is caught in a severe storm?"; "How much do you agree or disagree with the following statement? If Jeremiah is caught in a severe storm, this will have a large negative impact on him."). The other two questions measured how much evidence the target agent had ("How much evidence does Jeremiah have that he will get caught in a severe storm?"; "How much do you agree or disagree with the following statement? Based on the weather report, Jeremiah will probably be caught in a severe storm."). Participants responded using 7-point rating scales.

Participants then reported how much control the character had over their own mental states[1]. We probed participants' judgments regarding two kinds of control, which differed based on what kind of process the character might try to use in order to change their mental state. One control process was for the character to reason about what to believe or desire ("reasoning" condition). The other process was for the character to eliminate their access to reasons by taking a pill that would make them forget the relevant evidence or utility information ("pill" condition).

To illustrate these two different kinds of control processes, consider desire control in the Storm vignette.

[1] We also measured participants' judgments of perceived control using items that did not specify the process ("reasoning" vs. "pill") by which control would be attempted. We omit these items and their analyses in the interest of space.

Participants reported Jeremiah's reasoning-based control over his desire by responding to the question:

> "How difficult or easy would it be for Jeremiah to form the desire to get caught in a severe storm by… *thinking about what will happen* if he gets caught in a severe storm?"

By contrast, when participants evaluated Jeremiah's pill-based control over his desire, they responded to the question:

> "How difficult or easy would it be for Jeremiah to form the desire to get caught in a severe storm by… *taking a pill that makes him forget* that ______?"

In the desire condition, the missing text in the pill question was filled in with information relevant to the utility of the outcome and was based on the participant's utility condition. For instance, in the bad outcome condition, the underline was filled in with "his ship has been outfitted for hard weather." However, in the very bad outcome condition, the underline was filled in with "his ship is old and in need of repair." Thus, taking the pill ensures that the target character forgets that the outcome is associated with either low or very low utility, respectively.

The belief control questions shared the same format, but focused on belief formation and evidence. For instance, when reporting whether the character could easily exercise reasoning-based control over their belief, participants read:

> "How difficult or easy would it be for Jeremiah to form the belief that he will not get caught in a severe storm by…
> *thinking about whether* he will get caught in the storm?"

And when reporting whether Jeremiah could exercise pill-based control over their belief, participants read:

> "How difficult or easy would it be for Jeremiah to form the belief that he will not get caught in the storm by… *taking a pill that makes him forget* that _____?"

In this case, the missing text described the evidence that Jeremiah had for the belief he was trying to change. So in the low evidence condition, participants read "the local weather station is extremely unreliable," while the high evidence participants read, "the local weather station is extremely reliable."

The belief and desire control questions were shown on separate pages in random order for each participant. Within a block of belief or desire questions, the "reasoning" and "pill" variants were presented in random order. Participants responded to all questions on the same 7-point rating scale, anchored at 1 (extremely difficult) and 7 (extremely easy).

At the end of the study, participants reported their age and sex and were debriefed.

Results

Target sample size, exclusion criteria, and data analyses for this study were preregistered on AsPredicted (*link*). We created composite measures of perceived evidence (r = .65) and utility (r = .73) from our two-item manipulation checks. We subjected these measures to 2 (Evidence) x 2 (Utility) x 4 (Vignette) between-participant ANOVAs. As expected, participants rated the outcome as worse for the character in the very bad outcome condition (M = 6.12, SD = 1.03) compared to the bad outcome condition (M = 3.59, SD = 1.52), F(1, 923) = 1081.63, p < .001; however, perceived utility was unaffected by the evidence manipulation, F(1, 923) = 0.21, p = .649. Likewise, participants judged that the character's evidence of the bad outcome was stronger in the strong evidence condition (M = 5.91, SD = 1.19) compared to the weak evidence condition (M = 4.12, SD = 1.31), F(1, 923) = 537.87, p < .001. And as expected, perceived evidence was unaffected by the utility manipulation, F(1, 923) = 0.68, p = .409.

We next analyzed participants' judgments that the character could change their current beliefs and desires to adopt opposing ones. To this end, we aggregated participants' "reasoning" and "pill" control judgments into new dependent variables, "belief control" and "desire control," each predicted by the independent variable "control process." We then submitted participants' belief and desire control judgments to two 2 (Evidence) x 2 (Utility) x 4 (Vignette) x 2 (Control process) mixed ANOVAs. Figure 2 shows the mean judgments across the four vignettes.

Control over desires. We observed no differences in average desire control judgments across utility conditions, F(1, 923) = 0.14, p =.705, nor evidence conditions, F(1, 923) = 1.62, p = .204. However, we did observe a main effect of process, F(1, 923) = 76.08, p < .001: participants reported that it would easier for the character to change their desire through reasoning (M = 4.05, SD = 1.83) than by taking the pill (M = 3.32, SD = 1.93). As expected, we observed no interaction between evidence and control process, F(1, 923) = 0, p =.984. However, we did observe the expected utility × control process interaction, F(1, 923) = 21.89, p < .001 (Figure 2, left panel). This interaction revealed that, when evaluating the character's control through reasoning, participants thought it would be easier for the character to change their desire in the bad outcome condition (M = 4.24, SD = 1.76) compared to the very bad outcome condition (M = 3.86, SD = 1.88), F(1, 923) = 9.24, p =.002. By contrast, when thinking about the character taking a pill to forget the stakes of the outcome, participants judged it to be more difficult for the character to change their desire in the bad outcome condition (M = 3.11, SD = 1.91) compared to the very bad outcome condition (M = 3.52, SD = 1.94), F(1, 923) = 11.62, p = .001.

Control over beliefs. On average, participants thought that it would be easier for the character to adopt the irrational

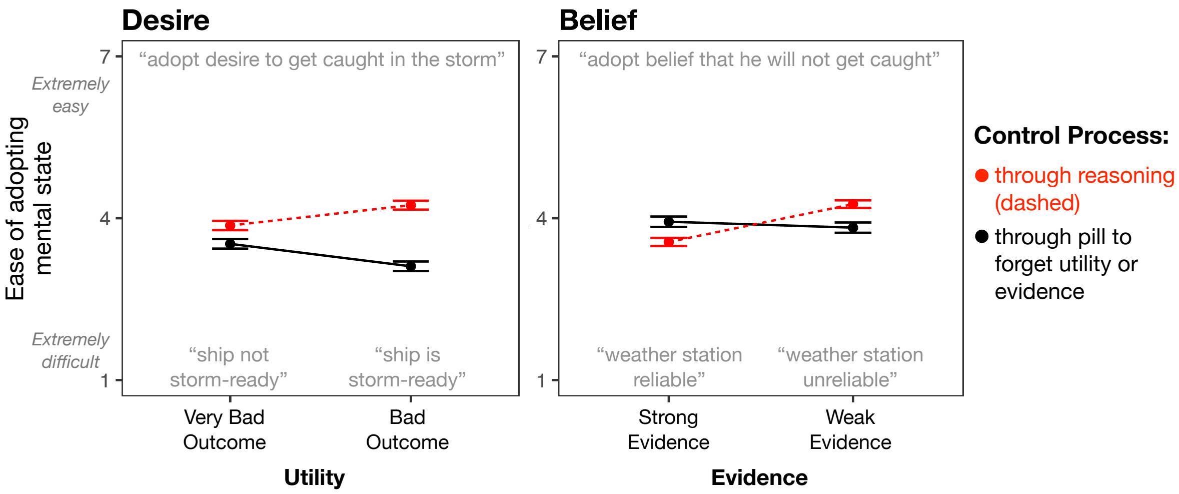

Figure 2. Mean (and standard error of mean) control ratings for desires and belief based on control process, the utility manipulation, and evidence manipulation.

belief in the weak evidence condition (*M* = 4.04, *SD* = 1.85) compared to the strong evidence condition (*M* = 3.75, *SD* = 1.87), *F*(1, 923) = 23.71, *p* < .001. Additionally, on average, there was no difference based on whether the characters tried to execute this control through reasoning (*M* = 3.91, *SD* = 1.63) or through the pill (*M* = 3.88, *SD* = 2.07), *F*(1, 923) = 0.17, *p* = .683. However, as predicted, we observed a significant evidence × control process interaction, *F*(1, 923) = 22.04, *p* < .001 (Figure 2, right panel). This interaction revealed that, when considering the pill, participants reported no difference in ability to change belief across the weak (*M* = 3.83, *SD* = 2.08) and strong (*M* = 3.94, *SD* = 2.07) evidence conditions, *F*(1, 923) = 0.32, *p* = .57. However, when considering whether the character could do so via reasoning, participants judged it harder to change beliefs in the strong evidence condition (*M* = 3.56, *SD* = 1.63) compared to the weak evidence condition (*M* = 4.26, *SD* = 1.55), *F*(1, 923) = 52.78, *p* < .001.

Unexpectedly, we also observed an overall main effect of utility, such that participants assigned to the very bad outcome condition rated it overall more difficult for the character to change his belief (*M* = 3.71, *SD* = 1.86) compared to participants assigned to the bad outcome condition (*M* = 4.10, *SD* = 1.85), *F*(1, 923) = 23.71, *p* < .001. However, as expected, utility did not interact with process control or any of our other predictors (*p*s > .19).

Discussion

We hypothesized that people intuitively treat others' mental states (and mental state changes) as constrained by reasons and rationality, such that it would be more difficult for people to adopt beliefs or desires that violate rationality as opposed to conform to it. As predicted, participants reported that the presence of stronger evidence against a belief made it more difficult for someone to adopt that belief. In parallel, participants reported that someone's knowledge that an outcome would dramatically lower their well-being made it more difficult for that person to desire for that outcome to occur. These results suggest that people do not only assume that others are rational and guided by reasons, but also that, to some degree, those others are constrained by rationality and reasons.

We also observed that participants' judgments that the characters were constrained depended on whether the characters were trying to change their minds through reasoning, or whether they could do something to remove awareness of the reasons from their minds. Participants reported that strong reasons constrained mental state change only when the character was aware of those reasons during reasoning. By contrast, when participants considered the possibility of intentionally forming an irrational belief or desire by eliminating knowledge of the evidence or utility from their mind (the "pill" condition), the difference in control either attenuated or reversed. These results support the idea that people conceptualize reason-based constraint in a way that reflects their lay theory of reasoning-based mental state formation.

These findings extend and help explain prior work documenting that adults and children believe that situational and environmental forces can limit others' choices and autonomy (e.g., Chernyak et al, 2013; Kushnir et al, 2015; Woolfolk et al., 2006). For instance, people regard extremely high monetary awards for participating in medical research as constraining others' autonomy (Baron, 1998): when the incentives are too high, people think others cannot say no. However, besides noting that situational pressures seem to constrain behavior, prior work has not fully explained why people intuitively treat situations as constraining. The work presented above suggests one contributing explanation: Situational pressures operate by uncontrollably causing others to hold reasonable attitudes. This explanation neatly explains why people see punishment, social norms, and morality (the most commonly used stimuli in prior studies) as constraining. Punishment, social ostracism, and the risk of harm provide extremely strong reasons to desire certain outcomes – on our view, so strong that a rational agent could not form alternative, overriding desires through their normal process of reasoning and mental state formation.

Alternative explanations. The observation that attributions of low control were specific to reasoning provides evidence against recent accounts of people's judgments of constraint. For instance, some recent work has suggested that people implicitly conflate rationality and possibility (Phillips & Cushman, 2018; Phillips & Knobe, 2009). Accordingly, people intuitively treat rational decisions as easy for others to execute and irrational decisions as difficult or impossible. A related alternative suggests that people conceptualize situations as "soft constraints" that do not literally limit others' freedom (the way that so-called "hard" constraints like physical limitations do) but merely appeal to different motives (e.g., Kalish, 1998; Reeder, 2009). Accordingly, this work predicts that people report what others "can" and "cannot" do merely as a way of indicating what they think would be sensible or insensible to do, respectively. Both of these alternative accounts suggest that judgments of rationality and irrationality act as a kind of heuristic when judging control.

However, if participants used rationality as a heuristic for control, then they should have attributed similar amounts of control in both the "reasoning" and "pill" conditions. The choice to adopt an irrational mental state is equally irrational regardless of the specific method used to do so. The characters in the vignettes know that the resulting belief or desire would be equally incorrect and counterproductive regardless of whether they consciously ignored the strong reasons or took a pill that made them forget those strong reasons. And, moreover, taking a pill to forget strong reasons is arguably more irrational because it is irreversible. However, our results do not reflect this line of reasoning: participants thought the strong reasons were not constraining when the characters could (still irrationally) take a pill to forget them. Put another way: participants judged others as free to think irrationally when they could manipulate the reasons available as inputs to reasoning. These results speak in favor of the reason-based constraint model that we propose, and against these heuristic explanations.

Everyday reasoning about mental state constraint. In our study, the character's evidence and expected utility were salient to participants. The vignettes were relatively simple and participants made explicit judgments about the strength of evidence and utility prior to making judgments about control. Absent these cues, it is possible that participants would not readily appreciate the strength of reasons, and accordingly, not judge the characters in these situations to be constrained. Indeed, prior work shows that people tend not to think about the constraining evidence that others have for their beliefs (Cusimano & Goodwin, 2020). This may explain, in part, why in other studies people tend to attribute high mental state control to others (Cusimano & Goodwin, 2019). These observations help resolve the apparent tension between our results and prior findings. While people think others are constrained by reasons, they by default do not probe deeply enough into others' reasoning to override their initial judgments that those others think and act freely.

Future work on attributions of autonomy and rational constraint. The work presented here focuses on autonomy over beliefs and desires, but it provides a clear direction for future work investigating perceived constraint over observable, physical behaviors. Constraints on behavior can also be conceptualized as constraints on mental state formation, and specifically as constraints on the formation of intentions to perform those behaviors. To wit: someone cannot intentionally speak if they cannot intend to speak. Our proposal generates a straightforward and testable prediction about situational constraint over observable behaviors: namely, that people judge others' actions as situationally constrained when they judge intentions to perform those actions as restricted by what it is rational to intend. Of course, this proposal is speculative and remains a question for future research, but it would draw an even tighter connection between reasons, autonomy, and action.

And finally, this work also reveals a tension between attributions of rationality and autonomy. On the one hand, people tend to believe that others act most autonomously only when those others are capable of rational thought and have full knowledge of their situations. On the other hand, we have shown that rationality, as well as full knowledge of one's situation, can also limit one's perceived autonomy. For instance, prior work has shown that others are seen as constrained when they lack relevant knowledge of their situation (e.g., Chernyak et al., 2013). By contrast, in our study, participants judged others as less constrained when they thought that the agents were ignorant of relevant features of their situation. We may expect that if participants viewed the characters as lacking sanity, they would judge them as more capable of adopting irrational mental states simply because they want to, while perhaps judging them as lacking autonomy in other ways or in other situations. In sum, the relationship between rationality and autonomy is more complex than prior empirical work has suggested, and an ongoing goal of our research is to more fully understand this relationship.

Conclusion

Assumptions that others are rational play a foundational role in folk theory of mind. Our results extend this role to the impact of reason-based constraints and rationality on mental state change. People seem to believe that, at least when reasoning normally, others cannot help but be rational. This may explain why people view others as lacking autonomy in certain situations: those situations strongly constrain what is rational for someone to think and want.

Acknowledgments

This work was supported by a grant from the Templeton World Charity Foundation. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Foundation.

References

Baker, C. L., Jara-Ettinger, J., Saxe, R., & Tenenbaum, J. B. (2017). Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nature Human Behaviour, 1*(4), 1-10.

Boswell, R. G., & Kober, H. (2016). Food cue reactivity and craving predict eating and weight gain: a meta-analytic review. *Obesity Reviews, 17*(2), 159-177.

Chernyak, N., Kushnir, T., Sullivan, K. M., & Wang, Q. (2013). A comparison of American and Nepalese children's concepts of freedom of choice and social constraint. *Cognitive Science, 37,* 1343–1355.

Cusimano, C., & Goodwin, G.P. (2019). Lay beliefs about the controllability of everyday mental states. *Journal of Experimental Psychology: General*, *148*(10), 1701-1732.

Cusimano, C., & Goodwin, G.P. (2020). People judge others to have more voluntary control over beliefs than they themselves do. *Journal of Personality and Social Psychology, 119*, 999-1029

Epley, N., & Gilovich, T. (2016). The mechanics of motivated reasoning. *Journal of Economic Perspectives*, *30*(3), 133-40.

Gill, M. J., & Cerce, S. C. (2017). He never willed to have the will he has: Historicist narratives, civilized blame, and the need to distinguish two notions of free will. *Journal of Personality and Social Psychology, 112*, 361-382.

Jara-Ettinger, J., Gweon, H., Schulz, L. E., & Tenenbaum, J. B. (2016). The naïve utility calculus: Computational principles underlying commonsense psychology. *Trends in cognitive sciences, 20*, 589-604.

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

Kouchaki, M., Smith, I. H., & Savani, K. (2018). Does deciding among morally relevant options feel like making a choice? How morality constrains people's sense of choice. *J. of Personality and Social Psych, 115*, 788–804.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480-498.

Kushnir, T., Gopnik, A., Chernyak, N., Seiver, E., & Wellman, H. M. (2015). Developing intuitions about free will between ages four and six. *Cognition, 138*, 79–101.

Lazarus, R. S. (1991). *Emotion and adaptation*. Oxford University Press

Martin, J. W., & Cushman, F. (2016). Why we forgive what can't be controlled. *Cognition*, *147*, 133-143.

Monroe, A. E. & Malle, B. F. (2010). From uncaused will to conscious choice: The need to study, not speculate about people's folk concept of free will. *Review of Philosophy and Psychology, 9,* 211-224.

Murray, D., & Lombrozo, T. (2017). Effects of manipulation on attributions of causation, free will, and moral responsibility. *Cognitive science*, *41*(2), 447-481.

Phillips, J., & Cushman, F. A. (2017). Morality constrains the default representation of what is possible. *Proceedings of the National Academy of Sciences*, *114*, 4649–4654.

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

Reeder, G. (2009). Mindreading: Judgments about intentionality and motives in dispositional inference. *Psychological inquiry*, *20*(1), 1-18.

Weiner, B. (1995). *Judgments of responsibility: A foundation for a theory of social conduct*. Guilford Press.

Wolf, S. (1980). Asymmetrical freedom. *The Journal of Philosophy*, *77*(3), 151-166.

Wood, T., & Porter, E. (2019). The elusive backfire effect: Mass attitudes' steadfast factual adherence. *Political Behavior, 41*(1), 135-163.

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

Young, L., & Phillips, J. (2011). The paradox of moral focus. *Cognition, 119*(2), 166-178.