If it works we didn't need it: Intuitive judgments of 'overreaction'

Jonathan F. Kominsky[1,2], Daniel Reardon[1], & Elizabeth Bonawitz[1,2]

[1]Rutgers University – Newark; [2]Harvard Graduate School of Education

Abstract

When laypeople decide if a costly intervention is an overreaction or an appropriate response, they likely base those judgments on mental simulation about what could happen, or what would have happened without an intervention. To narrow down from the infinite set of possibilities they could consider, they may engage in a process of sampling. We examine whether judgments of overreaction can be explained by a utility-weighted sampling account from the JDM literature, or a norm-weighted sampling account from the causal judgment literature, both, or neither. Three experiments test whether these judgments are overly influenced by low-risk bad outcomes (utility-weighted sampling), or by what is likely and prescriptively good (norm-weighted sampling). Overall, participants' judgments indicate that they disregard low-risk bad outcomes, and even when a high-risk outcome is successfully avoided, the intervention is an overreaction. These results favor a norm-weighted sampling account in the specific case of evaluating overreactions.

Keywords: Causal judgment; Decision-making; Counterfactual reasoning; Mental simulation; Overreaction

Introduction

"If it looks like you're overreacting, you're probably doing the right thing." – NAIAD Director Dr. Anthony S. Fauci, quoted in the Washington Post, April 9, 2020.

How are intuitive judgements of overreaction made? Prospective judgments of risk and retrospective judgments of causality have both been found to rely on some form of mental simulation (Kahneman & Tversky, 1982; Kahneman & Miller, 1986; Lewis, 1973). These judgments are made not just based on events that have actually occurred, but events that might occur in the future (hypothetical reasoning), or other events that could have occurred instead of those that did (counterfactual reasoning). However, any type of reasoning that is based on simulating alternative possibilities must address an obvious problem: there are an infinite number of possibilities one could consider, but our minds have very limited processing capacity.

One way to address the capacity issue -- proposed in both the causal reasoning and decision-making literatures -- is that the mind engages in a process of *sampling*, selecting only a few of the relevant possibilities in a probabilistic manner, weighted by certain factors. While both literatures have recognized that sampling offers a solution to an otherwise intractable problem, the samples generated seem to be weighted by different factors depending on the judgment at hand.

One proposal from the decision-making literature is that people engage in 'utility-weighted sampling' (Lieder, Griffiths, & Hsu, 2018), in which extreme outcomes, both extremely good and extremely bad, are given disproportionate weight relative to their likelihood. That is, people are very likely to consider the best-case and/or worst-case scenario, even if those outcomes are relatively unlikely.

Alternatively, a proposal from the causal judgment literature is that people engage in 'norm-weighted sampling' (Icard, Kominsky, & Knobe, 2017; Kominsky & Phillips, 2019; Phillips, Morris, & Cushman, 2019). Under this proposal, people consider possibilities that are some combination of *good* and *likely.* That is, they think about what "should" have happened.

There is evidence for each of these types of sampling within their respective domains, but what about judgments that sit at the intersection of these two areas? To address the COVID-19 pandemic, governments across the world have had to engage in extremely costly public health interventions, ranging from closing businesses to limiting travel to instituting huge infrastructures for testing and contact-tracing. The public reaction to these measures has been mixed, and particularly toward the beginning of the pandemic, there were some who thought that these costly interventions were an overreaction (e.g., Karson, 2020). This raises an obvious question: how do people make these judgments of overreaction?

There is a substantial body of work about whether stock investors or government policy-makers *objectively* overreact to the inputs they receive (see Maor, 2014 for a summary). In economics it is straightforward to look in hindsight at how investors reacted to various signals, and determine whether those responses were rational or optimal. Surprisingly, we found no literature that has investigated how *intuitive* judgments of overreaction are made.

This open question has both theoretical and practical importance. From a theory standpoint, these judgments offer an opportunity to extend two different accounts of mental simulation and better understand how the mind generates and makes use of information about events that did not actually occur. From a practical standpoint, understanding these judgments is critical to public policy messaging. If an intervention is considered an overreaction rather than an appropriate response, the public may be less willing to comply with it (though this also needs to be tested).

In the absence of existing work on judgments of overreaction, these simulation-based processes from the decision-making and causal judgment literatures seem like good candidates for how such judgments are made. When participants are presented with a costly intervention and try to determine if it is reasonable, they might evaluate what will happen as a result of the intervention and what would happen without it. For example, if someone thinks that shutting indoor dining will mean that their town will have very few cases of COVID-19, while allowing indoor dining will mean that their town has many cases of COVID-19, then it seems

like a reasonable intervention. If instead they think that there will be few cases of COVID-19 even if indoor dining continues, then it seems like an overreaction. The two different sampling proposals we describe make different predictions about the judgments people will make.

Under the utility-weighted sampling account (Lieder et al., 2018), we would expect that people should often view costly interventions as appropriate, even when the disaster they are trying to prevent is relatively unlikely, because this account gives disproportionate weight to extreme outcomes, particularly costly outcomes. For example, if a dam had a 5% chance of failing catastrophically, this theory predicts that people would be very likely to consider the possibility that it will fail catastrophically even if they only consider a few possibilities, justifying a costly intervention.

In contrast, according to the norm-weighted sampling account (Icard et al., 2017), people should frequently judge that interventions are overreactions, particularly when the risk of a bad outcome is low. For example, in the case of the failing dam, this theory predicts that people should almost never consider the possibility that the dam will fail, as it is both unlikely and strongly negative. In fact, they might be more likely to view a costly intervention as an overreaction all the time, because they are less likely to consider bad outcomes even when they are relatively likely (Sytsma, Livengood, & Roese, 2012).

Of course, there could be many other factors that also influence these judgments, such as whether they are prospective or retrospective, whether the intervention has a clear causal mechanistic link to the outcome, whether the goal is prevention of a bad outcome or mitigation of the consequences, and more. Examining these judgments will not only help us understand the judgments themselves, it may provide novel insights into these sampling processes.

The current experiments

We therefore designed three experiments manipulating different features of two different scenarios in order to gain an initial understanding of these judgments of overreaction. In all three experiments we asked for both prospective judgments (judging the intervention before the outcome is known) and retrospective judgments (judging the intervention after learning what happened).

Shared Methods

Participant recruitment

All participants were recruited from Prolific Academic, restricted to users from the USA who had not participated in any prior version of the experiment. In each experiment we pre-registered a sample of 40 participants in each between-subjects condition. Participants were compensated $1.12 for a ~7-minute task.

Materials

We created two scenarios. One, the "Dam" scenario, involved a town with a dam that could potentially fail and flood the town, with the intervention of a costly construction project that required displacing half of the town. The other, the "Fire" scenario, involved a power company using rolling blackouts during the hottest weeks of the year to avoid destructive wildfires. Each experiment varied different parameters of these scenarios in order to test different hypotheses about what factors influence judgments of overreaction.

In all experiments, we asked participants to rate the interventions (the construction project and the blackouts) on a scale that went from 0 to 100, with 0 labeled "didn't do enough", 50 labeled "appropriate response" and 100 labeled "complete overreaction". Participants made two ratings, a prospective rating before knowing the outcome, and a retrospective rating after knowing the outcome. The slider always started at 0, and participants were never given information about their prospective rating when making their retrospective rating.

Following each retrospective rating, participants completed check questions that served as both exclusion criteria and manipulation validation. If participants answered any question that had an objective answer (e.g., a question about whether an event happened, but not a question about the 'severity' of an event), they were excluded from analyses and replaced. Which questions were used for exclusions was preregistered separately for each experiment.

All stimuli were presented and responses recorded using Qualtrics (2005). We used comprehension checks as preregistered exclusion criteria, customized for each experiment. The materials, data, and analyses for all three experiments, as well as preregistrations, can be found at https://osf.io/k4cbq. (Note that Experiment 2 in the repository is Experiment 3 in this manuscript, and vice versa.)

Experiment 1

In Experiment 1, we manipulated three factors: The risk of the bad outcome (high or low), whether or not the bad outcome occurred, and whether there was an explicit mechanistic causal link between the intervention and the outcome. Under a simulation account, risk should affect prospective judgments of overreaction by changing how likely people are to consider possibilities that the outcome occurs. Similarly, whether the outcome occurs should affect retrospective judgments because if it does not, people may not consider counterfactual possibilities in which a bad event happened. However, as we are asking for judgments about *interventions*, another key factor in retrospective judgments might be whether the intervention is directly connected to the outcome: If the bad outcome occurs for unrelated reasons, does the intervention look less justified, or is it unaffected?

Experiment-specific methods

Participants We manipulated three factors between-subjects: Risk (high risk of bad outcome vs. low risk of bad outcome), Outcome (good vs. bad), and Causality (explicit mechanistic link w/intervention vs. unrelated mechanism), yielding a 2x2x2 design. We recruited 320 participants, 40

per cell. In addition, another 154 participants failed the preregistered exclusion criteria (32% attrition).

Stimuli For each scenario, participants first read background information about the scenario and the Risk manipulation, followed by a description of the intervention. They then made their prospective judgment of the intervention. These prospective judgments were therefore only subject to the Risk manipulation. On a separate page, participants then read the outcome of the event, which included the Outcome and Causality manipulations. They then made their retrospective judgment of the intervention.

Results and discussion

Results can be found in Fig. 1, collapsed across scenario. We preregistered an analysis plan in which we first conducted a mixed-model ANOVA with scenario as a within-subjects factor, to determine if there were any interactions between scenario and our factors of interest. For this experiment alone there were, so each scenario was analyzed separately.

Prospective ratings The only factor manipulated prior to the prospective ratings was Risk, and the initial mixed-model analysis found a significant interaction between Risk and Scenario (p<.001). We therefore conducted independent-samples t-tests separately for the Dam and Fire scenarios.

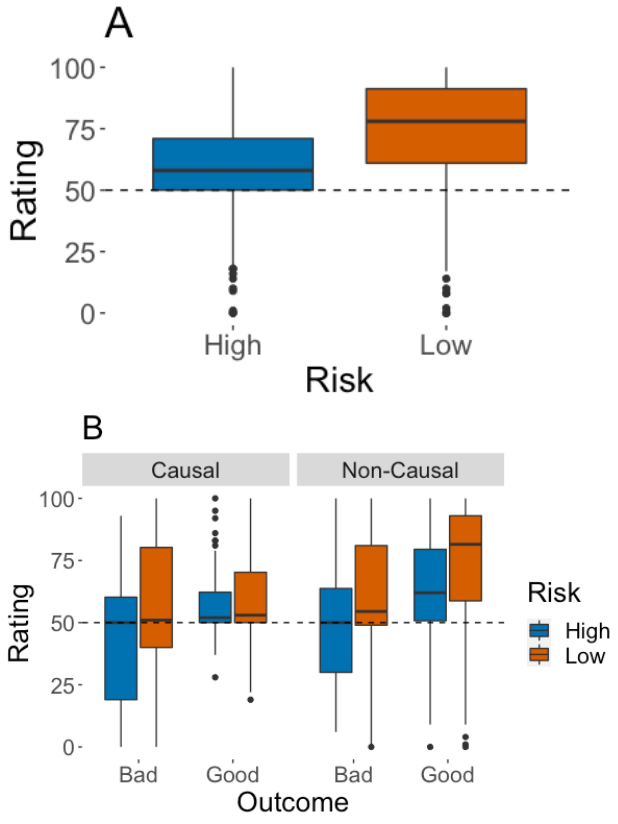

Figure 1. Results of Experiment 1, collapsed across the two scenarios. A) Prospective ratings. B) Retrospective ratings. The dashed line at 50 marks the point on the scale labeled "appropriate response".

There was a significant effect of Risk in the Dam scenario such that ratings in the low-risk condition (M=76.2, SD=20.6) were significantly higher (assessed as a larger overreaction) than those in the high-risk condition (M=55.5, SD=11.7), t(318)=11.03, p<.001, d=1.23. There was a similar, but smaller, effect in the Fire scenario (low-risk: M=71.4, SD=22.5; high-risk: M=62.4, SD=22.7), t(318)= 3.56, p<.001, d=.40. Notably, single-sample t-tests showed that all ratings were significantly above the midpoint of the scale (all ps<.001), indicating that participants tended to think these interventions were overreactions.

Retrospective ratings The initial mixed-model analysis with Scenario found a significant four-way interaction (p=.023), and so each scenario was analyzed separately. For brevity only significant statistical tests are reported here; the full analysis plan is available at the repository linked above.

For the Dam scenario, a 2 (Risk) x 2 (Outcome) x 2 (Causality) fully between-subjects ANOVA found all three main effects were significant (ps<.003), but also significant interactions between Outcome and Causality, F(1, 312)=7.65, p=.006, η_p^2=.024, and a significant three-way interaction, F(1, 312)=6.11, p=.014, η_p^2=.019. To summarize our follow-up analyses, in the high-risk conditions there were two main effects, indicating participants gave higher (overreaction) ratings to good outcomes and higher ratings when there was no explicit causal link, ps≤.038. In the low-risk conditions, there was a significant interaction between Outcome and Causality: there was a strong effect of Causality for good outcomes wherein an explicit causal link led to lower ratings (p<.001), but no effect of Causality for bad outcomes.

Notably, ratings were significantly *below* 50 in the high-risk bad-outcome case (uncorrected ps<.02), not significantly different from 50 in the low-risk bad-outcome no-causal-connection condition (p=.13), and significantly above 50 in every other condition (uncorrected ps<.03). This suggests that when the bad outcome actually occurs, the intervention is seen as more appropriate, and if the bad outcome was likely the intervention is seen as inadequate, at least in this case.

For the Fire scenario, the 2 (Risk) x 2 (Outcome) x 2 (Causality) ANOVA revealed only main effects of Outcome, F(1, 312)=16.39, p<.001, η_p^2=.050, and Causality F(1, 312)=9.89, p=.002, η_p^2=.031. As in the Dam scenario, overreaction ratings were overall higher when the outcome was good, and when there was no explicit causal link between the intervention and the outcome. Ratings were significantly higher than 50 in all high-risk good-outcome conditions and all low-risk conditions except when there was a bad outcome and a direct causal link with the intervention (i.e., the bad outcome directly overwhelmed the intervention; ps<.004). All other ratings were not significantly different from 50 (ps>.4). This pattern is similar to the dam scenario but with higher ratings overall.

These results are largely in line with the norm-weighted sampling account, in which people tend to consider possibilities that are a combination of good and likely. If people tend to ignore extreme bad outcomes, they should

regard costly interventions as overreactions except when the bad outcome has undeniably occurred. Overall, participants were likely to judge an intervention to be an overreaction, except when the bad outcome actually occurred. This was mitigated somewhat when there was a very clear causal mechanism, but from a public policy perspective it is an unsettling finding nonetheless.

Experiment 2

Experiment 2 is similar to Experiment 1, but manipulates inevitability instead of causal mechanism. If the goal of an intervention is *mitigation* rather than *prevention*, how does it affect people's judgments of the intervention? One can think of it as turning the bad outcome into something like an immutable background condition (McGill & Tenbrunsel, 2000). If the bad outcome is guaranteed, then the intervention may always seem more justified. Therefore, this experiment manipulated the goal of the intervention (prevention or mitigation), the risk of the bad outcome, and whether the bad outcome occurred.

Experiment-specific methods

Participants We manipulated three factors between-subjects: Risk (high risk of bad outcome vs. low risk of bad outcome), Intent (prevention vs. mitigation), and Outcome (good vs. bad). yielding a 2x2x2 design. We aimed to recruit 320 participants, but due to imperfect randomization caused by the online platform ended up with 321, with slightly uneven distributions across cells. The low-risk mitigation bad-outcome cell had 39 participants, while both prevention good-outcome conditions had 41. Another 311 participants were excluded based on preregistered exclusion criteria (49.2% attrition).

Stimuli The stimuli were very similar to Experiment 1 with the following modifications: Participants read both the Risk and Intent manipulations prior to giving their prospective ratings. The 'prevent' conditions were identical to their corresponding Risk condition from Experiment 1, but the 'mitigate' conditions described the goal of the intervention being to redirect the flood to minimize damage (dam scenario) or limit the spread of wildfires rather than prevent them altogether (fire scenario). After prospective ratings, participants read that the worst outcome occurred or that the prevention or mitigation was successful, and gave their retrospective ratings.

Results and discussion

We preregistered an analysis plan in which we first conducted a mixed-model ANOVA with scenario as a within-subjects factor, to determine if there were any interactions between scenario and our factors of interest. There was only one significant interaction, between Scenario and Outcome for retrospective ratings alone. For consistency across analyses, we elected to collapse across Scenario by averaging the ratings for the two scenarios together. The results of a follow-up analysis by scenario are not meaningfully different from those reported here. All results are shown in Fig. 2.

Prospective ratings A 2 (Risk: high vs. low) x 2 (Intent: prevent vs. mitigate) ANOVA found only a main effect of Risk such that ratings were higher in the low-risk conditions (M=71.1, SD=17.4) than high-risk conditions (M=58.4, SD=17.7), $F(1, 317)=42.02$, $p<.001$, $\eta_p^2=.117$. These ratings are in line with what we observed in Experiment 1, and follow the same pattern. The average ratings in both risk conditions were significantly above 50, ps<.001, indicating that participants generally regarded all interventions as overreactions even in the high-risk conditions.

Retrospective ratings We conducted a 2 (Risk) x 2 (Intent) x 2 (Outcome: good vs. bad) ANOVA. There was a significant main effect of Risk such that, as in the prospective ratings, ratings were higher in the low-risk conditions (M=53.0, SD=19.6) than the high-risk conditions (M=45.9, SD=20.6), $F(1, 313)=11.40$, $p<.001$, $\eta_p^2=.035$. There was also a significant main effect of Outcome such that ratings were higher for good outcomes (M=56.9, SD=13.0) than bad outcomes (M=42.0, SD=23.5), $F(1, 313)=50.83$, $p<.001$, $\eta_p^2=.140$. There were no other significant main effects or interactions, ps>.25.

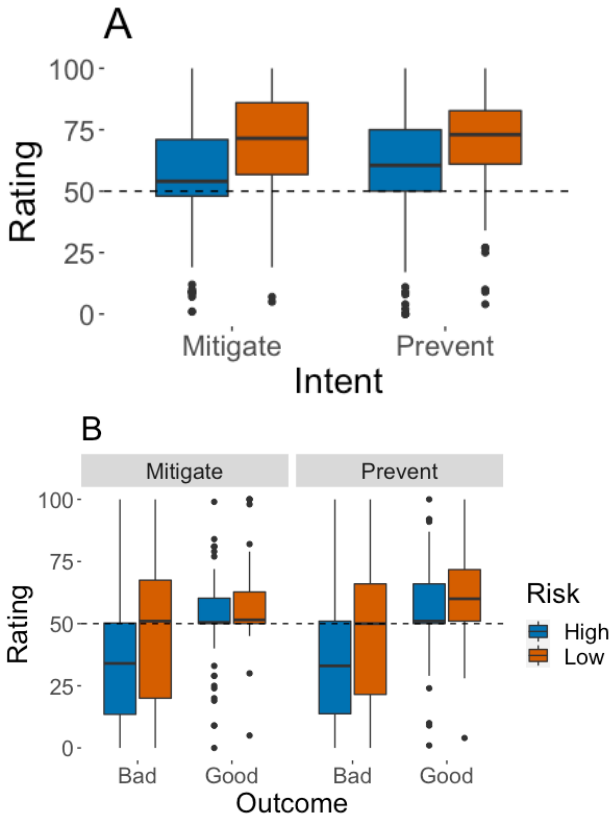

Figure 2. Results of Experiment 2. A) Prospective ratings. B) Retrospective ratings. The dashed line at 50 marks the point on the scale labeled "appropriate response".

We examined whether ratings in each of the Risk x Outcome cells differed from 50 with four single-sample *t*-tests. This analysis found that the mean rating in both good-outcome conditions were significantly higher than 50, *p*s<.007, the mean rating in the high-risk bad-outcome condition was significantly below 50, *p*<.001, and the mean rating in the low-risk bad-outcome condition was not significantly different from 50, *p*=.15. When the outcome was good, participants judged the intervention to be an overreaction, but when the outcome was bad, they judged the intervention to be either appropriate or insufficient.

In short, whether an intervention was intended to mitigate or prevent a bad outcome had no detectable impact on judgments of whether it was an overreaction. However, we replicated two key findings from Experiment 1: interventions against low-risk bad outcomes and interventions that succeed are judged to be overreactions.

Experiment 3

Experiment 3's primary goal was to provide independent evidence for the role of possibilities in these judgments, by explicitly providing participants with information about *will happen* as a result of the intervention and information about what *would have happened* without it. This is analogous to work in the causal judgment literature that has asked participants to explicitly consider certain counterfactual possibilities in order to validate the role of counterfactual reasoning in those judgments. (Kominsky & Phillips, 2019; Phillips, Luguri, & Knobe, 2015). However, unlike previous work, we varied these possibilities on dimensions of both valence (good or bad) and *realism*, examining whether such judgments are influenced by unrealistic possibilities.

Experiment-specific methods

Pilot stimulus-generation experiment The goal of this experiment was to present participants with statements about what *will* happen as a result of the intervention, or what *would have* happened without it, and vary those outcomes on two dimensions: valence (good or bad) and realism. For realism, we wanted outcomes that were slightly good or bad, the 'realistic' best and worst case outcomes, and unrealistic best and worst outcomes that were completely implausible, in order to test whether even impossible events influenced these judgments. To that end, we asked 18 participants from Prolific to generate six possible outcomes for each scenario, and drew from these responses when constructing our stimuli. The stimuli and full set of responses can be found at the repository linked above.

Participants We manipulated three factors between-subjects, all involving the outcome: Intervention (what will happen vs. what would have happened), Valence (good vs. bad), and Realism (slightly vs. realistic vs. unrealistic). yielding a 2x2x3 design. We aimed to recruit 480 participants, but ended up recruiting 481 due to imperfect randomization, with one extra participant in the "what would have happened"/bad/realistic condition. An additional 40 participants were excluded due to failing our exclusion criteria (7.7% attrition).

Stimuli The first part of each vignette was identical to the high-risk conditions from Experiment 1, and were identical for all participants. That is, unlike Experiments 1 and 2, the prospective ratings occurred before *any* of our manipulations, so they served as a baseline rather than a meaningful measure of interest. After the prospective rating, participants read one of the 12 outcomes and gave the retrospective rating. To give a sense of what we mean for each level of realism, here are the responses we chose for the dam scenario, for the "what would have happened"/good outcome conditions:

Slightly: "The cracks aren't an immediate threat and can be dealt with at a later date."
Realistic: "No heavy rain or storms hit the town so the water level does not rise very much."
Unrealistic: "Poseidon comes out of the water and stops the dam from breaking himself. At the same time, he gives free horses to everyone."

Results

Initial analyses by Scenario revealed no effects or interactions, so we collapsed across Scenario, as in Experiment 2. We examined all results with 2 (Intervention) x 2 (Valence) x 3 (Realism) ANOVAs. Unlike Experiments 1 and 2, the prospective ratings in this experiment occurred before *any* of the manipulations. All participants gave their prospective ratings after seeing the exact same information, so any variability could only be due to random variation.

Because the prospective ratings were essentially a baseline rather than a meaningful measure unto themselves, we had the opportunity to examine the effects of our manipulations while controlling for individual variation by analyzing *difference scores*, i.e., retrospective rating – prospective rating, for each participant. These difference scores can be found in Fig. 3. Positive indicates a move towards overreaction, while negative indicates a move towards "not enough" or "appropriate".

A 2 x 2 x 3 ANOVA of difference scores revealed main effects of Valence, $F(1, 469)=4.82$, $p=.029$, $\eta_p^2=.010$, and Intervention, $F(1, 469)=7.89$, $p=.005$, $\eta_p^2=.017$, as well as an interaction between the two, $F(1, 469)=7.56$, $p=.006$, $\eta_p^2=.016$. There were no significant effects or interactions with Realism.

We conducted separate analyses of the effect of Valence for each Intervention condition. In the "what will happen" condition (i.e., when participants were told the outcome of the intervention), there was no difference between good outcomes (M=-4.2 , SD=15.5) and bad outcomes (M=- 3.4, SD=18.4), $t(238)=.41$, $p>.5$, and both were significantly different from 0, ps < .05. In the "what would have happened" condition, there was a significant difference, such that good outcomes were more positive (M=+4.8, SD=17.5) than bad outcomes (M=-3.3, SD=19.5), i.e., retrospective ratings moved more towards overreaction for good outcomes, $t(239)=3.37$, $p<.001$, $d=.434$. This difference in the good-

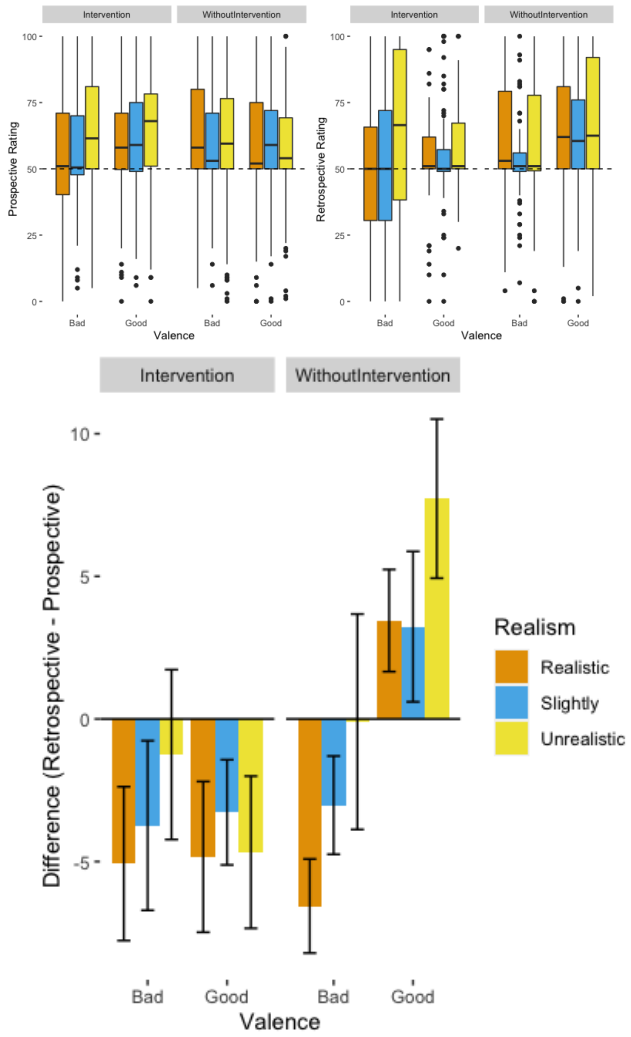

Figure 3. Results of Experiment 3. Top panels show raw prospective (left) and retrospective (right) ratings. Bottom panel shows difference scores, and error bars in this panel represent ± 1 SEM.

outcome condition was significantly different from 0, p=.003, but not in the bad-outcome condition, p=.07.

In short, the degree of outcome realism presented had little to no impact on participants' judgments, but providing a possibility in which the bad outcome would not have occurred even without the intervention led to higher ratings (i.e., more toward overreaction), while considering any other possibility led to lower ratings or no change.

General Discussion

In three experiments we provided an initial investigation of how people judge whether a costly intervention is an appropriate response, an overreaction, or not enough. Experiment 1 found that judgments tend more toward overreaction overall, but especially when the bad outcome is unlikely, and when the bad outcome does not occur. Notably, this means that a successful intervention tends to be regarded as an overreaction. However, successful interventions are judged to be more appropriate when there is a direct causal mechanistic link between the intervention and the outcome.

Experiment 2 provided a replication of the effects of risk and outcome from Experiment 1, but found no effect of whether the goal of the intervention was to *prevent* or *mitigate* the bad outcome, suggesting that it does not matter whether some negative outcome is seen as inevitable or not.

Finally, Experiment 3 showed that, when participants were directly given specific possibilities to consider, their judgments reflected those possibilities. In particular, when participants were told that the bad outcome would not have occurred even without the intervention, their judgments of overreaction increased. Notably, however, the realism of the possibilities had no detectable effect in our scenarios.

Sampling possibilities

Experiment 3 showed very clearly that participants give higher ratings when provided with a possibility that indicates the intervention was not *necessary*, i.e., the bad outcome would not have occurred even without the intervention. In Experiments 1 and 2, in every case in which the bad outcome did not actually occur, the intervention was judged to be an overreaction. This was slightly mitigated when there was an explicit mechanistic link between the intervention and the prevention of the bad outcome in Experiment 1, but even then ratings were significantly above the midpoint of the scale.

Together, these results indicate that judgments of overreaction do rely on the same kind of simulation process previously found in the causal judgment and decision-making literatures. We can posit that, when the bad outcome does not actually occur, participants did not consider possibilities in which the bad outcome happens, even if there is no intervention. Under that interpretation, these results are compatible with the norm-weighted sampling accounts (Icard et al., 2017; Phillips et al., 2019).

However, these experiments offer only an initial investigation into these issues, and there are many limitations that will need to be addressed in future work. We do not know whether our bad outcomes were extreme enough for the purposes of utility-weighted sampling. We also did not examine the costliness of the intervention as a factor. Furthermore, while we have focused on whether the ratings were above or below the midpoint of our scale, the scale itself was novel, and we don't know how participants interpreted or interpolated between the labeled points of "didn't do enough", "appropriate response", and "complete overreaction". One could also describe this work as investigating judgments of appropriateness or underreaction, but we treat these as a spectrum of which overreaction is part. We also did not ask for judgments of cases that would be regarded as *objective* overreactions according to rational choice theory (Maor, 2014). For that matter, we have no sense of how well-calibrated participants' judgments were, because the cases we presented did not have any objective criteria by which to determine whether they were overreactions or not. There is much more to be learned from and about these judgments.

Practical ramifications

One lesson from these studies is that Dr. Fauci was largely correct: If you intervene successfully to prevent a bad outcome, it will probably look like an overreaction. This seems, at first glance, like bad news for any major policy intervention, from public health to climate change. However, there are several important unknowns, and some good news. First, these studies did not establish a link between judgments of overreaction and behavior. Second, while we tried to use scenarios based on real events, we did not use cases that directly affected our participants. Finally, we found at least one clear way to mitigate judgments of overreaction for a successful intervention: present a clear causal mechanism.

Conclusions

Judgments of 'overreaction' provide a fertile ground for future investigations of causal judgment and decision-making processes, as well as immediately applicable lessons for policymakers and science communicators (Vermeulen, 2014). This initial investigation provides an initial foothold, but there is an urgent need for future work to rigorously test both the theoretical ramifications and practical applications of these judgments, both in relation to immediate crises like COVID-19 and ongoing crises like climate change.

Acknowledgements

This work was funded by a Social Science Research Council COVID-19 Rapid Response Grant to JFK and EB. JFK was supported by a Templeton Foundation Developing Belief Network postdoctoral fellowship.

References

Icard, T. F., Kominsky, J. F., & Knobe, J. (2017). Normality and actual causal strength. *Cognition*, *161*, 80-93.

Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological review*, *93*(2), 136-153.

Karson, K. (2020c). Large majorities of Americans back coronavirus restrictions, slower return to normal: POLL. *ABC News*. Retrieved from https://abcnews.go.com/Politics/large-majorities-americans-back-coronavirus-restrictions-slower-return/story?id=70291873

Kominsky, J. F., & Phillips, J. (2019). Immoral Professors and Malfunctioning Tools: Counterfactual Relevance Accounts Explain the Effect of Norm Violations on Causal Selection. *Cognitive Science*, *43*, e12792.

Lewis, D. (1973). Causation. *The Journal of Philosophy*, *70*(17), 556-567.

Lieder, F., Griffiths, T. L., & Hsu, M. (2018). Overrepresentation of extreme events in decision making reflects rational use of cognitive resources. *Psychological Review*, *125*(1), 1-32.

Maor, M. (2014). Policy Bubbles: Policy Overreaction and Positive Feedback. *Governance*, *27*(3), 469-487.

McGill, A. L., & Tenbrunsel, A. E. (2000). Mutability and propensity in causal selection. *Journal of Personality and Social Psychology*, *79*(5), 677-689.

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.

Phillips, J., Morris, A., & Cushman, F. (2019). How we know what not to think. *Trends in cognitive sciences, 23*(12), 1026-1040.

Qualtrics. (2005). Qualtrics online survey software. Provo, UT.

Sytsma, J., Livengood, J., & Rose, D. (2012). Two types of typicality: rethinking the role of statistical typicality in ordinary causal attributions. *Studies in History and Philosophy of Biology and Biomedical Sciences*, *43*(4), 814-820.

Vermeulen, K. (2014). Understanding Your Audience: How Psychologists Can Help Emergency Managers Improve Disaster Warning Compliance. *Journal of Homeland Security and Emergency Management*, *11*(3), 309-315.