ARTICLE IN PRESS

Journal of Applied Research in Memory and Cognition xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Applied Research in Memory and Cognition

journal homepage: www.elsevier.com/locate/jarmac

COVID-19 Express article

Intuitive Judgments of "Overreaction" and Their Relationship to Compliance with Public Health Measures

Jonathan F. Kominsky *
Harvard University Graduate School of Education, USA
Rutgers University – Newark, USA

Daniel Reardon
Rutgers University – Newark, USA

Elizabeth Bonawitz
Harvard University Graduate School of Education, USA
Rutgers University – Newark, USA

How does the general public determine if a policy intervention is appropriate or an overreaction, and how do such judgments influence compliance? In four studies, we found that prospective judgments of overreaction are influenced by how likely a bad event is to occur, and retrospective judgments are influenced by whether the intervention is successful. In Studies 1–3, we investigated the mechanics of these judgments and found that if the bad event is low-risk, or the intervention is successful in preventing it, people judge the intervention to be an overreaction. In Study 4, a survey of 450 US participants showed that opinions of the risks and outcomes of the COVID-19 pandemic correlated with overreaction judgments, and critically, those judgments of overreaction predicted non-compliance with public health measures.

Keywords: Judgment, Overreaction, COVID-19, Policy Communication

General Audience Summary

We investigated how people determine whether a costly public policy is an "overreaction" or an appropriate response. Using fictional examples where we could manipulate the facts, we found evidence that when people make these judgments before the outcome is known, they are based on the risk of something bad happening, but even when that risk is high, people think costly interventions are overreactions. Furthermore, when making these judgments after the outcome is known, if the intervention is successful, that is, if the bad outcome does not occur, then these interventions are seen as overreactions. We then looked at the judgments of COVID-19 public health policies and found that the same factors influence people's judgments of these real-world policies, and most importantly, that judgments of overreaction predict whether people will comply with these policies. Understanding how people make judgments of overreaction can help create more effective public policy messaging for future crises.

* Correspondence concerning this article should be addressed to Jonathan F. Kominsky, Harvard Graduate School of Education, Cambridge, United States. Contact: jonathan.f.kominsky@gmail.com (J.F.K.).

Please cite this article as: Kominsky, J.F., Reardon, D., & Bonawitz, E. Intuitive Judgments of "Overreaction" and Their Relationship to Compliance with Public Health Measures, *Journal of Applied Research in Memory and Cognition* (2021), https://doi.org/10.1016/j.jarmac.2021.11.001

In order to face public health crises like the SARS-CoV-2 pandemic, governments are often forced to impose strict and costly policy measures. To be effective, these measures must often be imposed well ahead of the negative outcomes they seek to prevent. This creates a challenge for communicating with the general public because the public is being asked to take substantial costs to avoid something that has not happened yet. On March 15, 2020, the director of NAIAD, Dr. Anthony Fauci, summed up the problem in an interview: "If it looks like you're overreacting, you're probably doing the right thing." (CBS News, 2020).

Past work in behavioral economics and policy has investigated cases of *objective* overreactions, that is, the magnitude of financial decisions made in relation to new input (Maor, 2014; Peters, Jordan, & Tosun, 2017), but for policy adherence by the general public, what may matter more is *subjective* judgments of whether something is an overreaction. In polls from both April and December of 2020, a consistent 15–20% of respondents said they felt that the policies used to combat the pandemic were overreactions (Karson, 2020; B Research, 2020), but it is unclear whether judging that they were "overreactions" changed how likely people were to comply with them. There is no prior research that addresses this question. In fact, there is no prior research that investigates how people determine whether something is an overreaction in the first place.

This presents two critical questions for policy and science communicators. First, what affects intuitive judgments of overreaction? Second, do judgments of overreaction predict compliance? By answering both of these questions together, we can determine the most effective communications strategies for future crises (Vermeulen, 2014).

Factors That Could Influence Judgments of Overreaction

In the absence of existing literature on intuitive judgments of overreaction, we formulated hypotheses about what could influence such judgments based on literatures about other judgments we felt were likely to play a role in judgments of overreaction: Prospective judgments of risk and retrospective judgments of causality. In both of these literatures, many have argued that 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) (Kahneman & Miller, 1986; Kahneman & Tversky, 1982; Lewis, 1973). We believe the same is likely true of judgments of overreaction. When judging whether an intervention is an appropriate response or an overreaction prospectively, the outcome is not yet known. When making a retrospective judgment, except in certain cases where the intervention and the outcomes are all easily and precisely quantifiable or the causal structure is very clear, one must consider what would have happened without the intervention, or with a less costly intervention, to determine if the actual intervention was an appropriate response.

If intuitive judgments of overreaction are made on the basis of the possibilities that people consider, then these judgments should be influenced by things that make people more or less likely to consider certain possibilities. Recent work in the causal judgment literature has suggested that people generally consider outcomes that are a combination of likely and good (Icard, Kominsky, & Knobe, 2017; Kominsky & Phillips, 2019; Phillips, Morris, & Cushman, 2019). Particularly when reasoning about what could have happened, people are reluctant to undo events that turned out well (e.g., De Brigard & Giovanello, 2012). On the other hand, recent work in the judgment and decision-making literature has suggested that people might give disproportionate attention to unlikely but extremely bad outcomes (Lieder, Griffiths, & Hsu, 2018). In either case, this presents two clear candidates for factors that could influence judgments of overreaction: the prospective likelihood of a bad outcome and whether the bad outcome actually occurs.

The Current Studies

We conducted four studies to understand judgments of overreaction and their impact. We operationalize "overreaction" in this project on a 100-point scale where 0 is labeled "didn't do enough," 50 is labeled "appropriate response," and 100 is labeled "complete overreaction" (for a discussion of the use of graded scales in studies of causal judgment, see O'Neill et al., 2021).

Studies 1 and 2 were designed to identify what factors control judgments of overreaction *in general*, by systematically manipulating fictional scenarios of (realistic) crises that were answered with costly public policies and asking participants to make both prospective judgments (before the outcome is known) and retrospective judgments.[1] In particular, these studies both manipulated the risk of a bad outcome and whether it actually occurred. In addition, each study tested one further factor to explore different possible influences on judgments of overreaction. Study 1 manipulated whether there was an explicit causal mechanism (Ahl, Amir, & Keil, 2020; Ahn & Kalish, 2000; Buchanan & Sobel, 2011; Lombrozo, 2010) between the intervention and the outcome, while Study 2 manipulated whether the intent of the intervention was to *prevent* a bad outcome or *mitigate* its consequences.

Study 3 validates that the risk manipulation affected the possibilities that participants considered in the intended way and asked whether they could easily generate less costly alternative interventions to better understand what possibilities they were considering.

Study 4, conducted January 14-16, 2021, shortly before the approval of the first SARS-CoV-2 vaccine in the US, examined judgments about real-world public health policies to see if they correlated with the factors found in Studies 1 and 2 and also tested whether there was a link between judgments of overreaction and *compliance* with these policies. Other work during the COVID-19 pandemic has identified a link between perceived risk and compliance with these regulations (Sinclair,

[1] We elected to use these fictional scenarios in order to avoid presenting misinformation about an ongoing public health crisis while manipulating our factors of interest.

Please cite this article as: Kominsky, J.F., Reardon, D., & Bonawitz, E. Intuitive Judgments of "Overreaction" and Their Relationship to Compliance with Public Health Measures, *Journal of Applied Research in Memory and Cognition* (2021), https://doi.org/10.1016/j.jarmac.2021.11.001

Hakimi, Stanley, Adcock, & Samanez-Larkin, 2021) or the role of trust in political leaders on judgments of perceived efficacy of these kinds of regulations (Mækelæ et al., 2020). This study seeks to determine whether judgments that these regulations are or are not overreactions influence compliance with them, above and beyond the factors that have been studied in these recent projects.

All studies were preregistered, and all registrations and materials can be found at the project repository (https://osf.io/k4cbq).

Study 1

In Study 1, we manipulated three features of the scenario: The a priori risk of the bad outcome (i.e., the dam failing or a destructive wildfire) with two levels (high vs. low), whether the bad outcome actually occurred (bad vs. good outcome), and whether there was an explicit causal mechanism between the intervention and the outcome (explicit link vs. ambiguous). Presenting participants with a direct causal link between the intervention and the outcome could make it less likely that they would consider alternatives in which the intervention does not occur, but the outcome is good nonetheless. Therefore, we predicted that providing an explicit causal mechanism might reduce judgments of overreaction.

Methods

Participants

All participants were recruited from Prolific Academic, restricted to users from the USA who had not participated in any prior version of the study. In studies 1 and 2, we preregistered a sample of 40 participants in each between-subjects condition. Participants were compensated \$1.12 for a ~7-min task. We recruited 320 participants. In addition, another 154 participants failed the preregistered exclusion criteria (32% attrition, see below).

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 study varied different parameters of these scenarios in order to test different hypotheses about what factors influence judgments of overreaction.

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 not given information about their prospective rating when making their retrospective rating. Participants completed the prospective and retrospective ratings for one scenario before reading the other, and the order of scenarios was counterbalanced.

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 2 × 2 × 2 design. In this study, the risk manipulation occurred prior to the prospective rating. Participants in the high-risk condition read a version of each scenario that specified that there was a high risk of a serious negative event (e.g., "Last year, the engineers found that the dam had developed tiny cracks and was at high risk of catastrophic failure in the near future.") or low risk of that same event (e.g., "Last year, the engineers found that the dam was in good shape and unlikely to fail anytime soon."). After making the prospective rating, participants then saw the second half of the vignette on a separate page, with the outcome and causality manipulations. Four different versions were made, one for each combination of outcome and causality. For example, the good outcome/mechanistic link for the dam scenario said, "Near the end of the renovation there was a historic rainstorm, which filled the reservoir to the absolute brim, and more than the dam was originally designed to handle. But, because of the renovations, the dam was fine," while the good outcome/no mechanistic link instead described a drought, leaving the reservoir low and therefore not threatening the dam. See Table 1 for the full set of variants for the "dam" scenario and the repository for the full materials. Participants made their retrospective ratings immediately after reading the rest of the vignette.

Following each retrospective rating, participants completed check questions that served as both exclusion criteria and manipulation validation. However, not every question was used in the exclusion criterion; if the question had any potentially subjective or unclear components, it was not used as an exclusion criterion. For example, in this study, there was one question about the risk, one about the outcome, and one about the mechanistic link. However, in the no-explicit-mechanistic-link conditions, it is left deliberately unclear whether the intervention had a mechanistic link rather than explicitly denying the presence of one. Therefore, only the check questions for the risk and outcome were used as exclusion criteria. Participants who answered either of those questions incorrectly were excluded from analyses and replaced.

Results

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 study (but not any other study), there was a significant interaction between the prospective "risk" manipulation and scenario, $F(1, 318) = 18.99$, $p < .001$, so each scenario was analyzed separately.

Prospective Ratings. The only factor manipulated prior to the prospective ratings was risk. We therefore conducted independent-samples t-tests separately for the Dam and Fire scenarios. The mean ratings can be found in Figure 1a.

Table 1
"Dam" Scenario From Study 1

1. Background: There was a town below a large dam. The dam held back a big reservoir, and if it ever failed, it would flood the whole town. Engineers inspected the dam regularly.	
2a. Low risk: Last year, the engineers found that the dam was in good shape and unlikely to fail anytime soon.	**2b. High risk**: Last year, the engineers found that the dam had developed tiny cracks and was at high risk for catastrophic failure in the near future.
3. Intervention: The state decided to do a major renovation to reinforce the dam. The renovation would be so large that they would have to displace half of the town for a year to make room for all of the equipment and workers it would take and eat a lot of the state budget. The townspeople were very unhappy about this but ultimately were forced to move away for two years so they could renovate the dam.	
4. Prospective judgment: What do you think of the following action in this story? The state's decision to displace half the town and renovate the dam [0–100 scale, 0 marked "did not do enough," 50 marked "appropriate response," 100 marked "overreaction"]	
5a. Good outcome/Explicit Mechanism: Near the end of the renovation, there was a historic rainstorm, which filled the reservoir to its absolute brim, and more than the dam was originally designed to handle. But, because of the renovations, the dam was fine.	**5b. Bad outcome/Explicit Mechanism**: While the dam was being renovated, the foundation rapidly eroded past the point of no return, and the dam failed despite the engineers' best efforts. The flood killed dozens and wiped out most of the town.
5c. Good outcome/Ambiguous: Near the end of the renovation there was a major draught, and the reservoir level dropped by several feet. With so little water to hold back, the dam was fine.	**5d. Bad outcome/Ambiguous**: While the dam was being renovated, there was a sudden earthquake that cracked the dam and caused it to fail. The flood killed dozens and wiped out most of the town.
6. Retrospective judgment: What do you think of the following action in this story? The state's decision to displace half the town and renovate the dam [0–100 scale, 0 marked "did not do enough," 50 marked "appropriate response," 100 marked "overreaction"]	

Note. The manipulations of Risk (2a/b), Outcome (5a/c-b/d), and Causality (5a/b-c/d) were all between-subjects. Bolded text was not presented to participants.

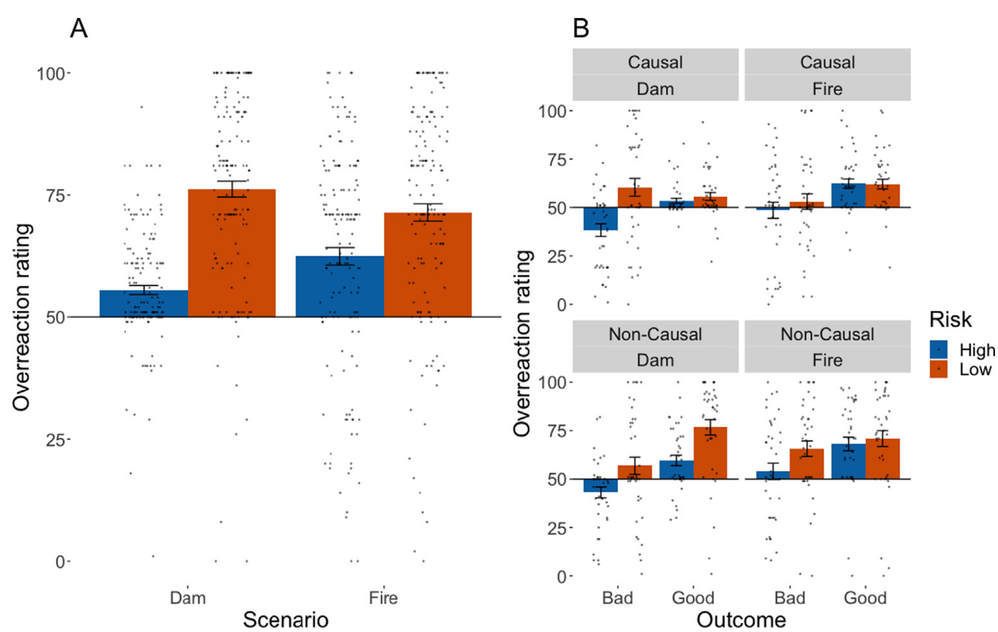

Figure 1. Overreaction ratings in Study 1. Bars show mean ratings divided by condition and scenario for prospective (A) and retrospective (B) ratings. Error bars represent ±1 SEM, dots are individual ratings. 50 on this scale was marked "appropriate response," indicated by the location of the x-axis.

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 = 0.40. Notably, single-sample t-tests showed that ratings were significantly greater than 50 for all four conditions (Dam High: $t(159) = 6.01$, $p < .001$; Dam Low: $t(159) = 16.06$, p < .001; Fire High: $t(159)$ = 6.92, p < .001; Fire Low: $t(159)$ = 12.05, p < .001), indicating that participants tended to think these interventions were overreactions *even when the risk was high*.

Retrospective Ratings. The initial mixed-model analysis with scenario found a significant four-way interaction ($p = .023$), and so each scenario was analyzed separately. The mean ratings can be found in Figure 1b.

For the Dam scenario, a 2 (risk) x 2 (outcome) x 2 (causality) fully between-subjects ANOVA found significant main effects of risk, $F(1, 312) = 34.54$, $p < .001$, $\eta_p^2 = .100$, outcome, $F(1, 312)$ = 24.92, p < .001, η_p^2 = .074, and causality, $F(1, 312)$ = 9.22, p = .003, η_p^2 = .029 but also significant interactions between outcome and causality, $F(1, 312)$ = 7.65,

Please cite this article as: Kominsky, J.F., Reardon, D., & Bonawitz, E. Intuitive Judgments of "Overreaction" and Their Relationship to Compliance with Public Health Measures, *Journal of Applied Research in Memory and Cognition* (2021), https://doi.org/10.1016/j.jarmac.2021.11.001

$p = .006$, $\eta_p^2 = .024$, and a significant three-way interaction, $F(1, 312) = 6.11$, $p = .014$, $\eta_p^2 = .019$. Neither interaction between risk and outcome, $F(1, 312) = 3.05$, $p = .08$, nor the interaction between risk and causality $F(1, 312) = 0.49$, $p = .49$, reached significance. To understand this three-way interaction, we first split the data by risk condition and conducted separate Outcome x Causality ANOVAs. In the high-risk condition, we found main effects of outcome, $F(1, 156) = 36.53$, $p < .001$, $\eta_p^2 = .19$, and causality, $F(1, 156) = 4.40$, $p = .038$, $\eta_p^2 = .027$, but no interaction, $F(1, 156) = 0.69$, $p = .79$. In short, participants gave higher (overreaction) ratings to good outcomes and higher ratings when there was no explicit causal link. In the low-risk conditions, we found no significant main effect of outcome, $F(1, 156) = 3.82$, $p = .052$, a significant main effect of causality, $F(1, 156) = 5.06$, $p = .026$, $\eta_p^2 = .031$, and critically, a significant interaction $F(1, 156) = 9.95$, $p = .002$, $\eta_p^2 = .060$. Post-hoc *t*-tests revealed that there was a strong effect of causality for good outcomes wherein an explicit causal mechanism led to *lower* ratings (Explicit mechanism: $M = 55.7$, $SD = 12.9$; Ambiguous: $M = 76.75$, $SD = 25.1$), $t(78) = 4.72$, $p < .001$, $d = 1.05$, but no effect of causality for bad outcomes, $t(78) = 0.55$, $p = .58$.

Notably, as reported in Table 2, preregistered single-sample *t*-tests revealed ratings were significantly below 50 in the high-risk bad-outcome conditions regardless of causality (uncorrected *p*s < .02), not significantly different from 50 in the low-risk bad-outcome no-causal-connection condition ($p = .13$), and significantly above 50 in every good outcome condition (uncorrected *p*s < .03). In short, in the cases where the intervention worked, with or without an explicit causal mechanism, participants' judgments were on the overreaction side of the scale.

For the Fire scenario, the 2 (risk) x 2 (outcome) x 2 (causality) ANOVA revealed only main effects of outcome, $F(1, 312) = 16.39$, $p < .001$, $\eta_p^2 = .050$, and causality $F(1, 312) = 9.89$, $p = .002$, $\eta_p^2 = .031$, but no effect of risk, $F(1, 312) = 3.26$, $p = .07$, and no interactions, *p*s > .19. 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; *p*s < .004). All other ratings were not significantly different from 50 (*p*s > .4). This pattern is similar to the dam scenario but with higher ratings overall.

Study 2

Study 1 identified key factors that affect judgments of overreaction: how likely people think the bad outcome is to occur and whether it actually does. Interventions that were effective (i.e., successfully prevented the bad outcome) were judged to be overreactions, though to a lesser degree when there was an explicit causal mechanism. Study 2 was designed to replicate the effects of risk and outcome from Study 1 and test a different factor that might change the possibilities people consider. If the goal of an intervention is to mitigate the consequences of a bad event rather than prevent it altogether, people may be less inclined to consider possibilities in which the bad outcome does not occur, regardless of its severity. That is, it may imply that the bad outcome is a foregone conclusion. In this case, the intervention may seem more appropriate, especially in prospective judgments, though Study 1 found that rat-

Table 2
Overreaction Ratings and t-tests Against 50 in Study 1

			Dam, prospective		Dam, retrospective		Fire, prospective		Fire, retrospective	
Risk	Outcome	Causal Mech.	Mean (SD)	*t*-test vs. 50	Mean (SD)	*t*-test vs. 50	Mean (SD)	*t*-test vs. 50	Mean (SD)	*t*-test vs. 50
High	Bad	Expl.	55.63 (8.72)	4.08***	38.33 (20.70)	3.57***	62.83 (21.90)	3.70***	48.55 (26.24)	0.35
		Ambig.	53.20 (14.04)	1.44	43.13 (17.83)	2.44*	60.33 (23.87)	2.74**	54.00 (26.62)	0.95
	Good	Expl.	56.93 (9.43)	4.64***	53.45 (8.08)	2.70*	64.43 (22.49)	4.06***	62.30 (15.31)	5.08***
		Ambig.	56.40 (13.56)	2.98**	59.63 (16.84)	3.62***	62.08 (23.09)	3.31**	68.10 (22.30)	5.13***
Low	Bad	Expl.	77.88 (18.25)	9.66***	60.38 (29.06)	2.26*	70.85 (20.06)	6.57***	53.03 (25.23)	0.76
		Ambig.	78.20 (18.24)	9.78***	56.85 (28.03)	1.55	73.93 (19.97)	7.58***	65.70 (25.33)	3.92***
	Good	Expl.	72.58 (22.20)	6.43***	55.70 (12.85)	2.81**	69.58 (22.73)	5.45***	62.10 (15.88)	4.82***
		Ambig.	76.18 (23.61)	7.01***	76.75 (25.14)	6.73***	71.28 (27.01)	4.98***	70.90 (25.91)	5.10***

Note. The *t* values presented in italics indicate that the mean is significantly less than 50. Expl. = Explicit causal mechanism condition, Ambig. = Ambiguous causal mechanism condition.
**p* < .05. ***p* < .01. ****p* < .001, all p-values are uncorrected.

Please cite this article as: Kominsky, J.F., Reardon, D., & Bonawitz, E. Intuitive Judgments of "Overreaction" and Their Relationship to Compliance with Public Health Measures, *Journal of Applied Research in Memory and Cognition* (2021), https://doi.org/10.1016/j.jarmac.2021.11.001

ings leaned toward overreaction even when the risk of a bad outcome was high.

Methods

Participants

We aimed to recruit a new set of 320 participants from Prolific Academic, but due to imperfect randomization ended up with 321, with slightly uneven distributions across cells. Another 311 participants were excluded based on preregistered exclusion criteria (49.2% attrition). Participants were once again paid \$1.12 for a ~7-min task.

Materials

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 $2 \times 2 \times 2$ design. However, there were slight differences from Study 1. First, the risk and intent manipulations *both* occurred before the prospective ratings, and only the outcome manipulation occurred afterward. The language of the risk manipulation was identical to Study 1. The intent manipulation either described the intervention as aiming to prevent the bad outcome from occurring at all (e.g., "…in order to stop the dam from failing altogether") or to mitigate the damage that would result from the bad outcome (e.g., "…in order to allow the flooding to be controlled and minimize damage when the dam failed."). After reading the first half of the scenario, participants made their prospective ratings.

After making their prospective ratings, participants read the second half of the vignette on a separate page. The Outcome manipulation was sensitive to the intent condition. So, for example, the good outcome/prevent condition was identical to the good outcome/mechanistic link condition of Study 1, but the good outcome/mitigate condition said, "Not long after the earthworks project finished, there was a historic rainstorm, which filled the reservoir to its absolute brim. The dam failed, but the water was redirected away from the town, and while there was a little damage to the town, nobody died." The full set of conditions from the Dam scenario can be found in Table 3.

For this study, the check questions for all three manipulations were preregistered as exclusion criteria, which likely contributed to the high attrition rate. For the intent check question, participants were asked if the goal was to mitigate or prevent the bad outcome. For the outcome question, participants were given three options, good, mitigated, and bad (e.g., for the dam, "the dam did not fail," "the dam failed but there was limited damage," "the dam failed and there was major damage").

Results

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 Figure 2.

Prospective Ratings

A 2 (risk: high vs. low) x 2 (intent: prevent vs. mitigate) ANOVA found 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), $F(1, 317) = 42.02$, $p < .001$, $\eta^2_p = .117$. There was no main effect of intent,

Table 3
"*Dam*" *Scenario From Study 2*

1. Background: There was a town below a large dam. The dam held back a big reservoir, and if it ever failed, it would flood the whole town. Engineers inspected the dam regularly.	
2a. Low risk: Last year, the engineers found that the dam was in good shape and unlikely to fail anytime soon.	**2b. High risk**: Last year, the engineers found that the dam had developed tiny cracks and was at high risk for catastrophic failure in the near future.
3a. Prevent: The state decided to do a major renovation to reinforce the dam. The renovation would be so large that they would have to displace half of the town for a year to make room for all of the equipment and workers it would take and eat a lot of the state budget. The townspeople were very unhappy about this but ultimately had to move away for a year so they could renovate the dam.	**3b. Mitigate**: The state decided to do a major earthworks project in order to allow the flooding to be controlled and minimize damage when the dam failed. The project would be so large that they would have to displace half of the town for a year to make room for all of the equipment and workers it would take and eat a lot of the state budget. The townspeople were very unhappy about this but ultimately were forced to move away for a year so they could redirect the water.
4. Prospective judgment: What do you think of the following action in this story? The state's decision to displace half the town and renovate the dam [0–100 scale, 0 marked "did not do enough," 50 marked "appropriate response," 100 marked "overreaction"]	
5a. Good outcome/Prevent: Not long after the renovation finished, there was a historic rainstorm, which filled the reservoir to its absolute brim. But, because of the renovations, the dam was fine.	**5b. Bad outcome/Prevent**: Not long after the renovation project finished, there was a historic rainstorm, which filled the reservoir to its absolute brim. The dam failed and the flood killed dozens and wiped out most of the town.
5c. Good outcome/Mitigate: Not long after the earthworks project finished, there was a historic rainstorm, which filled the reservoir to its absolute brim. The dam failed, but the water was redirected away from the town and while there was a little damage to the town, nobody died.	**5d. Bad outcome/Mitigate**: Not long after the earthworks project finished, there was a historic rainstorm, which filled the reservoir to its absolute brim. The dam failed, and the earthworks failed to redirect most of the water, so the flood killed dozens and wiped out most of the town.
6. Retrospective judgment: What do you think of the following action in this story? The state's decision to displace half the town and renovate the dam [0–100 scale, 0 marked "did not do enough," 50 marked "appropriate response," 100 marked "overreaction"]	

Note. The manipulations of risk (2a/b), intent (3a/b), and outcome (5a-d) were all between-subjects. Bolded text was not presented to participants.

Please cite this article as: Kominsky, J.F., Reardon, D., & Bonawitz, E. Intuitive Judgments of "Overreaction" and Their Relationship to Compliance with Public Health Measures, *Journal of Applied Research in Memory and Cognition* (2021), https://doi.org/10.1016/j.jarmac.2021.11.001

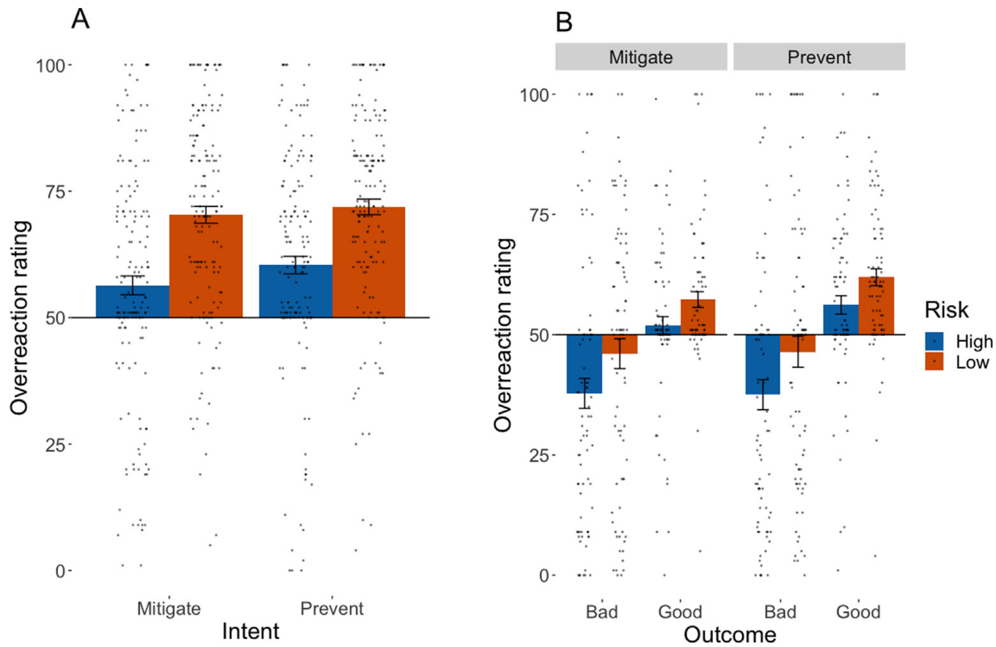

Figure 2. Overreaction ratings in Study 2. Bars show mean ratings divided by condition for prospective (A) and retrospective (B) ratings. Error bars represent ±1 SEM, dots are individual ratings. 50 on this scale was marked "appropriate response," indicated by the location of the x-axis.

$F(1, 317) = 2.00$, $p = .16$, and no interaction, $F(1, 317) = .39$, $p = .53$. These ratings are in line with what we observed in Study 1, and follow the same pattern. The average ratings significantly were above 50 in both risk conditions (Low: $t(159) = 15.30$, $p < .001$; High: $t(160) = 6.02$, $p < .001$), indicating that participants generally regarded all interventions as overreaction even when the risk of a bad outcome was high.

Retrospective Ratings

We conducted a 2 (risk) × 2 (intent) × 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 was no significant main effect of intent, $F(1, 313) = 1.31$, $p = .25$, and no significant interactions, *ps* > .25.

We examined whether ratings in each of the Risk × 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 (High risk: $M = 54.08$, $SD = 13.27$, $t(80) = 2.77$, $p = .007$; Low risk: $M = 59.69$, $SD = 12.08$, $t(79) =$, $p < .001$), the mean rating in the high-risk bad-outcome condition was significantly below 50 ($M = 37.65$, $SD = 23.28$), $t(79) = 4.75$, $p < .001$, and the mean rating in the low-risk bad-outcome condition was not significantly different from 50 ($M = 46.26$, $SD = 23.13$), $t(79) = 1.45$, $p = .15$. 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.

Study 3

Studies 1 and 2 converged on three clear findings. First, people prospectively judge costly interventions to be overreactions, even when the risk of a bad outcome is high. Second, these prospective judgments are still sensitive to the risk of the bad outcome, and overreaction ratings are higher when the risk is low. Third, when the intervention is successful, people tend to judge it to be an overreaction. What is shared across these three findings is that things that lead people to consider possibilities in which the bad event does not occur, even without the intervention, seem to lead to judgments of overreaction.

However, there is a potential deflationary explanation for the fact that these ratings are so consistently on the "overreaction" side of our scale. We find an effect of risk, but even when the risk is intended to be high, people rate the intervention to be an overreaction. This could indicate that our "high" risk really is not that high at all, and the actual likelihood that people ascribe to the bad outcome is very low, perhaps even lower than 50% likely to occur. Another explanation is that the interventions we provided were obviously excessive in some way. In other words, participants may have considered possibilities in which a less costly intervention was effective.

Study 3 was designed to address these two points by asking participants to rate the likelihood of the bad outcome in each of the risk conditions from Study 1 and asking them if they could readily think of a less costly intervention.

Methods

Participants

We recruited 40 US participants from Amazon Mechanical Turk using CloudResearch to filter out low-quality participants and bots. Four additional participants failed to pass basic manipulation checks and were excluded from analyses.

Materials and Procedure

We took the risk manipulation and intervention sections from the vignettes used in Study 1 (i.e., the parts that were presented prior to the prospective rating). We first presented just the risk manipulation and asked participants to rate what percent chance they thought there was of the bad outcome happening on a scale from 0 to 100.

After making this rating, we presented them with the intervention text from the vignettes in Study 1 and asked them whether they could think of less costly solutions to the problem, and if so, what that solution would be. They were presented with two options: "no" and a "yes" option that included a text field the participant could fill in with the idea.

Following this, participants completed the risk manipulation check from Study 1. Any participant who answered either manipulation check incorrectly was excluded and replaced. Furthermore, after completing both scenarios, participants were asked to rate how engaged they had been with the task on a 0-100 scale with 0 labeled "not paying attention at all" and 100 labeled "completely engaged." We planned to exclude any participant who provided an answer of less than 60, but no participant who passed the other manipulation checks did so, so no exclusions were made on this basis.

Results

We analyzed the likelihood responses in a 2 (risk: high vs. low) x 2 (scenario) mixed-model ANOVA. This revealed significant main effects of risk, $F(1, 38) = 145.25$, $p < .001$, scenario, $F(1, 38) = 16.06$, $p < .001$, and a significant interaction, $F(1, 38) = 7.45$, $p = .010$. Subsequent t-tests analyzing the effect of risk in each scenario revealed that likelihood estimates were higher in the high-risk condition (Dam: $M = 54.35$, $SD = 23.70$; Fire: $M = 73.05$, $SD = 24.19$) than the low-risk condition (Dam: $M = 3.80$, $SD = 4.81$; Fire: $M = 7.35$, $SD = 8.29$): Dam $t(38) = 9.35$, $p < .001$, $d = 2.95$; Fire $t(38) = 11.49$, $p < .001$, $d = 3.63$. The interaction indicates that the effect was stronger in the Fire scenario, but in both cases the difference between the low- and high-risk estimates was clear, and the high-risk estimates were greater than 50%.

Across 80 opportunities to provide alternatives (once for each of two scenarios * 40 participants), participants did so 34 times. This suggests that some, but by no means all, overreaction ratings may have been influenced by participants considering less costly ways of addressing the problem. Participants had no direct incentive to provide an alternative by design. If anything, they were disincentivized to do so because it required more work than answering "no." Thus, when participants provided an alternative, it suggests that the alternative was highly salient and came to mind easily. If we had incentivized participants to provide an alternative, they would likely have done so at a much higher rate, but this might not reflect whether they would consider such an alternative when making overreaction judgments. This is not contrary to our broader point that participants make these judgments by considering alternative possibilities, but it highlights a potential factor that could impact the possibilities participants consider. Future work will need to also incorporate the cost of the intervention and whether a less costly alternative might be effective into their considerations.

Study 4

For practical implications, the most alarming finding from Studies 1-2 is that when an intervention is successful, it is *more* likely to be judged an overreaction, and in general, any costly intervention is likely to be judged an overreaction. Study 3 does not find convincing evidence that these patterns are an artifact of our fictional stimuli. It is therefore even more critical to determine both whether these factors are related to judgments of overreaction in real-world cases and whether judgments of overreaction are related to compliance with the interventions being judged.

Therefore, in Study 4, we asked 450 US participants to rate real-world measures that had been taken against the SARS-CoV-2 pandemic. This survey was conducted between January 14 and 16, 2021, shortly before the approval of the first SARS-CoV-2 vaccine in the US.

Methods

Participants

We recruited a new set of 450 participants from Prolific Academic. An additional 30 participants were excluded from analyses under preregistered exclusion criteria. Participants were compensated at the same rate as in Study 1.

Materials

Unlike Studies 1-3, this study did not involve any manipulations but instead asked participants about several opinions related to the COVID-19 pandemic and public health measures that have been put in place to address it. All participants, therefore, filled out the exact same survey.

We first provided participants with a definition of "COVID-19 regulations" as "rules or directives used by authorities to prevent people from getting COVID and prevent the spread of the disease" and then asked them to provide a rating of COVID-19 regulations "as a whole" on our 100-point overreaction scale. This was our primary measure of interest. Following this, we asked them to rate, on new 100-point scales, how much of a threat COVID-19 was to the general public and themselves personally (with 0 marked "not a threat" and 100 marked "large threat"), how bad the COVID-19 pandemic has been in terms of illnesses and deaths (with 0 marked "not bad at all" and 100 marked "very bad"), and whether the pandemic would have been better, worse, or the same without COVID-19 regulations (with 0 marked "much worse," 50 marked "the same," and 100 marked "much better"). All scales except the better/worse scale started at 0, the better/worse scale started at 50.

Participants were also asked two binary choice questions at this point in the survey. First, whether the goal of COVID-19 regulations was to "prevent the disease from spreading altogether or limiting its spread" (i.e., prevent/mitigate), and second, whether they lived in the USA. The latter question was a check question—anyone who answered “no” was excluded from analyses (because we wanted to restrict the sample to people in the USA).

Following this, we asked them for separate ratings of nine *specific* COVID-19 regulations on the 100-point overreaction scale. These regulations were: closure of non-essential businesses, curfew mandating businesses close at a specific time, schools and workplaces moving to remote operations, a mask mandate for all people over six, recommendation to maintain six feet of distance from people around you, mandatory quarantine following interstate or international travel, halting interstate and international travel including cruises, businesses reopening with limited capacity, and disinfecting heavily used equipment (e.g., shopping carts). In addition, there was a tenth item that asked participants to move the scale as close to 50 as possible, which also served as an attention check and exclusion criterion: Participants who provided a rating greater than 55 or less than 45 (and the exact number was visible while they were making their rating) were excluded and replaced.

Next, we asked participants about their behavior during the pandemic, and specifically the degree to which they complied with various COVID-19 regulations. We asked how often people had engaged in seven different behaviors in the last three months, with five levels of frequency: “not once in the last 3 months,” “once or twice,” “3-5 times a month,” “6-12 times a month,” or “more than 12 times.” The seven behaviors were: going to indoor gyms, restaurants, or casinos; travel between states; travel between countries; ordered food for delivery (reverse-coded); gone into an indoor public space WITHOUT a mask; worn a mask in a public space (reverse-coded); left house or property for reasons other than work or essential shopping. For analysis, these were combined into a composite compliance score (see below).

Finally, we asked participants what state they lived in, whether someone close to them had gotten seriously ill or died of COVID-19, whether they themselves had gotten COVID-19, and if they had, how severe their illness had been. We also asked demographic questions about ethnicity and age.

Results

There were two key questions we sought to answer in this study. First, do the factors that we identified as causally affecting judgments of overreaction in Studies 1-3 correlate with those judgments in a real-life case? Second, do judgments of overreaction predict compliance with public health measures? There are other questions that could be explored within these data, but as those were the focus of this project, we report analyses focusing on those issues. The full dataset is available at our OSF repository at https://osf.io/k4cbq.

Correlates of Overreaction Judgments

Our primary interest in this analysis was to look at the relationships between measures that corresponded to manipulations in Studies 1-3 and judgments of overreaction. To that end, we focused on five specific items in addition to the overall overreaction rating. Corresponding to "risk" in earlier studies, we examined "threat to public" and "threat to self." Corresponding to "outcome" in earlier studies, we had "how bad the pandemic had been." Corresponding to "intent" in Study 2, we had "prevent/mitigate." In order to examine whether counterfactual thoughts about what would have happened without the interventions, we had "better or worse without regulations." In addition, to examine whether having had COVID affected these judgments, we created a COVID-19 severity score based on the severity question (with the lowest score being those who responded they had not had COVID-19 at all). It is worth noting that only ∼10% of our sample (43 participants) reported having a case of COVID-19 at any point.

To examine how these factors might be related to each other and the overall overreaction rating, we conducted a full set of partial Pearson correlations, looking at each two-way correlation while controlling for all of the other factors. The results of this analysis can be found in Table 4. (We excluded one additional participant from this analysis because they failed to respond to one of these measures.)

This analysis revealed that ratings of the threat to the public (corresponding to risk) and how bad the pandemic had been (corresponding to outcome) significantly negatively correlated with overreaction ratings. This is in line with Studies 1-3, as higher threat ratings correspond to higher risk (thus lower overreaction ratings), and higher “how bad” ratings similarly correspond to worse outcomes (thus lower overreaction ratings). The counterfactual better-or-worse question significantly positively correlated with overreaction ratings, which corresponds with our overall account of how these judgments are made: Participants who thought things would have been the same or better without the regulations tended to give higher overreaction ratings. Notably, these ratings were also significantly correlated with each other, and with the “threat to self” rating,

Table 4
Partial Correlation Matrix for Study 4

	ThreatPublic	ThreatSelf	HowBad	BetterWorse	Intent	Severity	Overall Overreaction
ThreatPublic		0.48***	0.53***	−0.25***	−0.07	−0.01	**−0.15****
ThreatSelf			0.13***	0.14***	−0.07	−0.08	**0.01**
HowBad				−0.13***	0.04	−0.05	**−0.2*****
BetterWorse					−0.07	−0.01	**0.17*****
Intent						−0.04	**−0.21*****
Severity							**−0.05**

Note. Values reported are Pearson’s r, with each two-way correlation controlled for every other factor. Correlations with the key overreaction ratings are highlighted in bold.
*p < .05. **p < .01. ***p < .001.

though the "threat to self" partial correlation with overreaction was not significant (likely because it was fully explained by the high correlation with "threat to public"). There were no significant correlations between COVID-19 severity (including not having had it at all) and any of the other measures.

Surprisingly, the prevent/mitigate significantly correlated with overreaction ratings as well, which was not what we predicted from the null effect of the intent manipulation in Study 2. We confirmed that there was a significant effect of intent with a post-hoc *t*-test comparing overreaction ratings of those who said the goal of these regulations was to *limit* the spread of COVID-19 (406/450) gave lower overreaction ratings (M = 37.3, SD = 23.9) than those who said the goal was to *prevent* the spread of COVID-19 (43/450) (M = 57.2, SD = 30.9), $t(47.46) = 4.10$, $p < .001$, $d = 0.81$. In other words, those who thought the goal of these regulations was mitigation rather than prevention judged the regulations to be appropriate or (more often) insufficient, which does fit what we originally hypothesized the intent manipulation would do in Study 2.

In short, this analysis confirms that the factors we causally manipulated in hypothetical scenarios in Studies 1-3 do influence real-world overreaction ratings about issues with direct public health relevance. The only notably unexpected finding was that the judged intent of the interventions, mitigation versus prevention, correlated with overreaction judgments here, but not in the hypothetical scenarios in Study 2.

Overreaction and Compliance

To examine whether these ratings of overreaction were connected to people's actual compliance with the regulations they were judging, we first created a composite "noncompliance score" from the seven behaviors we asked about, assigning each response a score of 0-4 by frequency (and reverse-coding items that indicated compliance, like wearing masks and ordering food) and then averaging all seven responses for each participant.

Our initial analysis was a straightforward linear regression of overreaction scores with this noncompliance composite score. This regression showed a significant positive relationship between overreaction ratings and noncompliance scores, such that higher overreaction ratings corresponded to more noncompliance with public health measures (b = .009, SE = .0008, R^2 = .24), $F(1,448) = 143.9$, $p < .001$. The fit of this regression can be seen in Figure 3.

However, as noted in the previous section, this overreaction rating was strongly correlated with many of our other measures. Therefore, it is possible that overreaction ratings hold no predictive power over and above these other measures. To test this, we conducted a backwards stepwise regression. A backwards stepwise regression takes a complete model and removes the weakest predictors from it until it arrives at a best-fitting model. In this case, we used AIC as the measure of model fit. We conducted two different backwards stepwise regressions. The first only involved main effects for every predictor of compliance (Overreaction plus the six predictors examined in the previous analysis, listed in Table 4). This stepwise regression was only able to remove one term: the "threat to the public" rating (AIC of base model = -818.23, AIC final model = -819.66). The second backwards stepwise regression started with a fully crossed model, including every possible interaction between these seven factors. The final model removed 36 of the interaction terms (of 128 terms in the fully crossed model), including the full seven-way interaction and most of the six-, five-, and four-way interactions (AIC of base model = -895.46, AIC of final model = -907.72). Critically, in both cases, the main effect of overreaction ratings persisted,

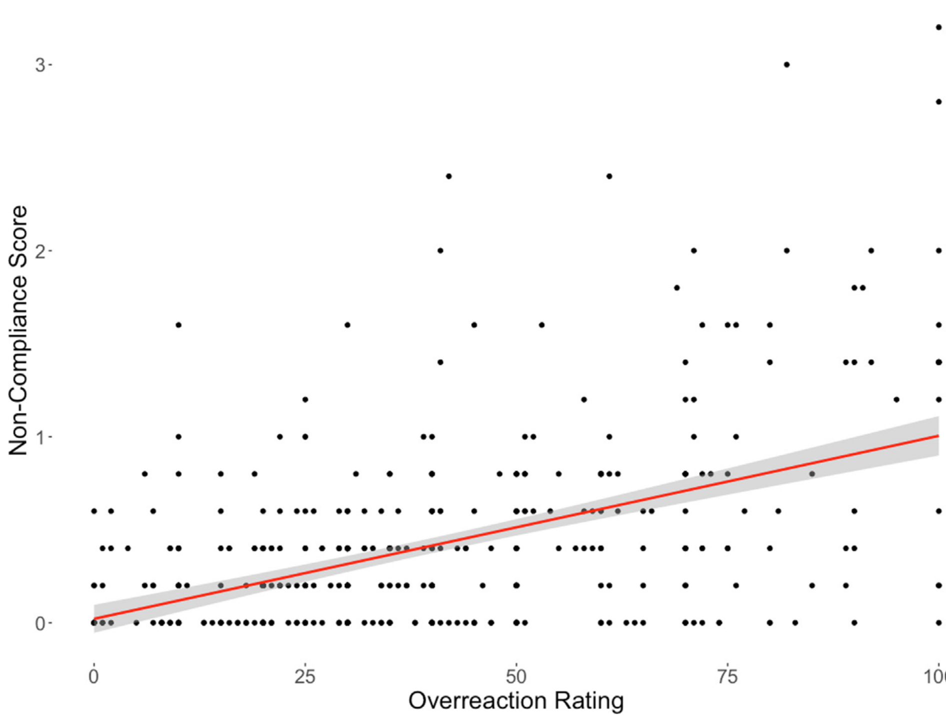

Figure 3. Relationship of overreaction ratings to noncompliance score.

Please cite this article as: Kominsky, J.F., Reardon, D., & Bonawitz, E. Intuitive Judgments of "Overreaction" and Their Relationship to Compliance with Public Health Measures, *Journal of Applied Research in Memory and Cognition* (2021), https://doi.org/10.1016/j.jarmac.2021.11.001

indicating that judgments of overreaction do make unique contributions to predicting non-compliance behavior, though many of the other things we measured do so as well.

General Discussion

Studies 1-2 find that prospective judgments of overreaction are strongly influenced by the risk of a bad outcome, and retrospective judgments are strongly influenced by whether the bad outcome actually occurs. However, even when the risk is high, prospective judgments of costly interventions tend to fall on the overreaction side of the scale, and when the bad outcome does not occur, retrospective judgments do as well. Study 3 rules out some deflationary explanations of these patterns. This would seem to vindicate Dr. Fauci's comment: If it works, people seem to think it is an overreaction. Study 4 finds convergent evidence for these patterns in judgments of real-world interventions taken against the COVID-19 pandemic. Critically, judgments of overreaction predict non-compliance with public health measures: When people think that an intervention is an overreaction, they are less likely to comply with it.

Intuitive Judgments of Overreaction

There are many potentially relevant factors that these studies were not able to investigate in-depth, such as the role of the cost of the intervention, political partisanship, the level of restriction and risk in the participants' current area, and more. Political partisanship is a particularly notable factor in the US: Opinions about the pandemic and public health measures used to address have broken sharply along political lines (e.g., Clinton, Cohen, Lapinski, & Trussler, 2021). However, while we believe it would be worth measuring political partisanship in future work related to the COVID-19 pandemic, there are two reasons we did not do so here. The first reason is that, while partisanship may have a strong influence on judgments about the pandemic and compliance with public health measures for this particular crisis among this particular US population, it may not generalize to future crises or even to other populations during the COVID-19 pandemic (e.g., Mækelæ et al., 2020; Lalot, Heering, Rullo, Travaglino, & Abrams, 2020). The second is that it is not clear how political partisanship might influence judgments of overreaction or compliance or both. Political partisanship is not a direct cause: It influences behavior through the information or direction provided by partisan sources and shared by others with the same political alignment. Partisanship could have a direct influence on compliance because individuals of a particular political alignment are instructed to behave in certain ways, or it could directly influence judgments of risk and overreaction based on information provided by partisan sources, and those judgments in turn influence compliance. It is evident that political partisanship has influenced behavior during the pandemic in the US, but we believe more work is required to understand how it has done so.

We also did not delve deeply into the cognitive mechanisms behind judgments of overreaction. We suggest that people make these judgments by considering possibilities, and our results indicate that, in line with some work in the causal reasoning literature, people tend to consider possibilities that are prescriptively good (Icard, Kominsky, & Knobe, 2017; Kominsky & Phillips, 2019; Phillips, Morris, & Cushman, 2019). However, we also found ample curiosities that will need to be unraveled in future work, such as the fact that the intent of an intervention (prevent vs. mitigate) had no effect in Study 2 but was a strong correlate of judgments of overreaction in the real-world cases of Study 4. It is interesting in and of itself that participants disagreed about the purpose of these interventions, but more work will be needed to understand what leads participants to think that the goal of these interventions was prevention versus mitigation. One possibility is that the prevention/mitigation judgment is influenced by some other factor (e.g., medical knowledge, primary news sources, etc.) that also influences judgments of overreaction. Alternatively, the prevent/mitigate manipulation in Study 2 may be different from what participants took prevent and mitigate to mean in Study 4 (e.g., it is already far too late to prevent the pandemic from happening in the first place). Both why people think a real-world intervention's goal is prevention or mitigation, and the influence of this distinction on judgments of overreaction, require further attention in future research.

Finally, it is important to consider the circumstances under which Study 4 was conducted. Data collection occurred from January 14 to 16, 2021, and the absolute highest 7-day average number of new cases in the US at any point in the pandemic was January 10-11, according to CDC data (retrieved from https://covid.cdc.gov/covid-data-tracker/#trends_dailycases). In other words, we were asking for these judgments during the absolute peak of the crisis and when our participants faced the greatest risk of infection. It is difficult to determine the actual risk participants faced as we only collected information on the relatively coarse level of the state they lived in, and in any case, recent work has found that lay estimations of risk during the pandemic are somewhat more optimistic than objective calculations of risk (Sinclair, Hakimi, Stanley, Adcock, & Samanez-Larkin, 2021). As a result, it is unclear what kind of influence this timing had on our results. It is entirely possible that conducting the same survey now when the threat of the pandemic is less acute (in many regions of the US), and vaccines are available would yield different results. Past work has found that people tend to consider possibilities that preserve good outcomes in the past (e.g., if someone avoided infection at the peak of the pandemic, they would not consider the possibility that they could have gotten infected instead), but also consider counterfactuals that improve past bad outcomes (De Brigard & Giovanello, 2012). As a result, when people are facing a less acute threat, they might consider even more optimistic outcomes and produce higher overreaction ratings as a result. We look forward to testing this hypothesis at a later date, when the pandemic has reached a lower ebb, or better yet, ceased to be an acute crisis altogether (as of this revision, in October 2021, we have only recently passed a secondary peak in case rates in the US).

Please cite this article as: Kominsky, J.F., Reardon, D., & Bonawitz, E. Intuitive Judgments of "Overreaction" and Their Relationship to Compliance with Public Health Measures, *Journal of Applied Research in Memory and Cognition* (2021), https://doi.org/10.1016/j.jarmac.2021.11.001

Overreaction Judgments and Compliance with Public Health Measures

At first glance, these studies paint a somewhat pessimistic picture: Successful interventions are judged to be overreactions, and people are less likely to comply with policies they believe are overreactions. We argue that judgments of overreaction make a distinct contribution to compliance behavior based on the correlation we observed in Study 4, but even then, we must acknowledge it is one of many different factors, possibly including factors that were not even included in the study, such as political partisanship, or the specific restrictions that were applied wherever a given participant lives.

That said, there are also hints about how policy communication could be improved, both from our results and from other recent investigations. Recent work has suggested that making people consider bad outcomes (which we argue they do not consider otherwise) may make them more likely to comply with public health measures (Sinclair, Hakimi, Stanley, Adcock, & Samanez-Larkin, 2021). Our findings also suggest other potential measures that might reduce judgments of overreaction and potentially compliance, such as emphasizing the specific causal mechanisms by which the intervention will prevent or mitigate bad outcomes (Study 1). Further research in this area could yield critical new strategies for crisis messaging and public health communicators (Vermeulen, 2014).

Conclusion

These four studies are, as far as we know, the first empirical studies of intuitive judgments of overreaction. We do not claim to have provided a comprehensive account of these judgments, but rather an initial demonstration that these judgments have both theoretical relevance for the mechanisms of causal judgment and decision-making and practical relevance for facing future crises. We hope these findings highlight the critical need for further studies of these judgments to improve future crisis messaging and compliance with public health measures.

Open Practices Statement

All studies were preregistered at https://osf.io/k4cbq/. A supplemental study and associated stimulus generation study, neither reported in this manuscript, are also included in this repository.

Acknowledgements

This work was supported by a Social Science Research Council COVID-19 Rapid Response Grant to J.F.K. and E.B. The authors would like to thank the members of the Computational Cognitive Development Lab for their helpful feedback throughout the project.

References

Ahl, R. E., Amir, D., & Keil, F. C. (2020). The world within: Children are sensitive to internal complexity cues. *Journal of Experimental Child Psychology*, *200*. Article 104932. https://doi.org/10.1016/j.jecp.2020.104932.

Ahn, W.-K., & Kalish, C. W. (2000). The role of mechanism beliefs in causal reasoning. In R. Wilson and F. C. Keil (Eds.), Explanation and cognition, (pp. 199–225). Bradford Books.

B Research. (2020). *Fox News Poll: December 6–9 2020*. Retrieved June 22, 2021, from https://static.foxnews.com/foxnews.com/content/uploads/2020/12/Fox_December-6-9-2020_National_Topline_December-14-Release.pdf.

Buchanan, D. W., & Sobel, D. M. (2011). Mechanism-based causal reasoning in young children. *Child Development*, *82*(6), 2053–2066. https://doi.org/10.1111/j.1467-8624.2011.01646.x.

CBS News (2020). *Transcript: Dr. Anthony Fauci discusses coronavirus on "Face the Nation," March 15, 2020*. Retrieved May 5, 2021, from https://www.cbsnews.com/news/transcript-dr-anthony-fauci-discusses-coronavirus-on-face-the-nation-march-15-2020/.

Clinton, J., Cohen, J., Lapinski, J., & Trussler, M. (2021). Partisan pandemic: How partisanship and public health concerns affect individuals' social mobility during COVID-19. *Science Advances*, *7*(2). Article eabd7204. https://doi.org/10.1126/sciadv.abd7204.

De Brigard, F., & Giovanello, K. S. (2012). Influence of outcome valence in the subjective experience of episodic past, future, and counterfactual thinking. *Consciousness and Cognition*, *21*(3), 1085–1096. https://doi.org/10.1016/j.concog.2012.06.007.

Icard, T. F., Kominsky, J. F., & Knobe, J. (2017). Normality and actual causal strength. *Cognition*, *161*, 80–93. https://doi.org/10.1016/j.cognition.2017.01.010.

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review*, *93*(2), 136–153. https://doi.org/10.1037/0033-295X.93.2.136.

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

Karson, K. (2020). *Large majorities of Americans back coronavirus restrictions, slower return to normal: POLL*. ABC News. Retrieved May 28, 2020, 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*(11). Article e12792. https://doi.org/10.1111/cogs.12792.

Lalot, F., Heering, M. S., Rullo, M., Travaglino, G. A., & Abrams, D. (2020). The dangers of distrustful complacency: Low concern and low political trust combine to undermine compliance with governmental restrictions in the emerging Covid-19 pandemic. *Group Processes & Intergroup Relations*. Advance online publication. https://journals.sagepub.com/doi/pdf/10.1177/1368430220967986.

Lewis, D. (1973). Causation. *The Journal of Philosophy*, *70*(17), 556–567. https://doi.org/10.2307/2025310.

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. https://doi.org/10.1037/rev0000074.

Lombrozo, T. (2010). Causal-explanatory pluralism: How intentions, functions, and mechanisms influence causal ascriptions. *Cognitive Psychology*, *61*(4), 303–332. https://doi.org/10.1016/j.cogpsych.2010.05.002.

Mækelæ, M. J., Reggev, N., Dutra, N., Tamayo, R. M., Silva-Sobrinho, R. A., Klevjer, K., & Pfuhl, G. (2020). Perceived efficacy of COVID-19 restrictions, reactions and their impact on mental health during the early phase of the outbreak in six countries. *Royal Society Open Science*, *7*. https://doi.org/10.1098/rsos.200644. Article 200644.

Maor, M. (2014). Policy bubbles: Policy overreaction and positive feedback. *Governance*, *27*(3), 469–487. https://doi.org/10.1111/gove.12048.

O'Neill, K., Henne, P., Bello, P., Pearson, J., & De Brigard, F. (2021). Degrading causation. osf.io. https://osf.io/m5w9c/download.

Peters, B. G., Jordan, A., & Tosun, J. (2017). Over-reaction and under-reaction in climate policy: An institutional analysis. *Journal of Environmental Policy & Planning*, *19*(6), 612–624. https://doi.org/10.1080/1523908x.2017.1348225.

Phillips, J., Morris, A., & Cushman, F. (2019). How we know what not to think. *Trends in Cognitive Sciences*, *23*(12), 1026–1040. https://doi.org/10.1016/j.tics.2019.09.007.

Sinclair, A. H., Hakimi, S., Stanley, M. L., Adcock, R. A., & Samanez-Larkin, G. R. (2021). Pairing facts with imagined consequences improves pandemic-related risk perception. *Proceedings of the National Academy of Sciences*, *118*(32). Article e2100970118. https://doi.org/10.1073/pnas.2100970118.

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. https://doi.org/10.1515/jhsem-2014-0055.

Received August 3, 2021
Received in revised form October 11, 2021
Accepted November 1, 2021
Available Online: xxxx