Published in final edited form as:
Q J Exp Psychol (Hove). 2022 July 15; 76(6): 1275–1297. doi:10.1177/17470218221111750.

Combining Refutations and Social Norms Increases Belief Change

Ullrich K. H. Ecker[1], **Jasmyne A. Sanderson**[1], **Paul McIlhiney**[1], **Jessica J. Rowsell**[1], **Hayley L. Quekett**[1], **Gordon D. A. Brown**[2], **Stephan Lewandowsky**[3,1]

[1]School of Psychological Science, University of Western Australia, 35 Stirling Hwy, Perth 6009, Australia

[2]Department of Psychology, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, United Kingdom

[3]School of Psychological Science, University of Bristol, 12a Priory Road, Bristol BS8 1TU, United Kingdom

Abstract

Misinformed beliefs are difficult to change. Refutations that target false claims typically reduce false beliefs, but tend to be only partially effective. In this study, a social norming approach was explored to test whether provision of peer norms could provide an alternative or complementary approach to refutation. Three experiments investigated whether a descriptive norm—by itself or in combination with a refutation—could reduce the endorsement of worldview-congruent claims. Experiment 1 found that using a single point estimate to communicate a norm affected belief but had less impact than a refutation. Experiment 2 used a verbally-presented distribution of four values to communicate a norm, which was largely ineffective. Experiment 3 used a graphically-presented social norm with 25 values, which was found to be as effective at reducing claim belief as a refutation, with the combination of both interventions being most impactful. These results provide a proof of concept that normative information can aid in the debunking of false or equivocal claims, and suggests that theories of misinformation processing should take social factors into account.

Keywords

misinformation; false beliefs; social norms; debunking; continued influence effect; belief change

Communicators and fact-checkers often face the challenge of reducing misconceptions in the contemporary media landscape (Graves & Amazeen, 2019; Hameleers & van der Meer, 2019). The primary method that has been used to reduce false beliefs is provision of factual corrections and counterarguments (e.g., Bode & Vraga, 2018; Ecker, Lewandowsky,

[3]Visual inspection suggested larger norm effects in participants with either low social assertiveness and low need for authenticity or high social assertiveness and high need for authenticity.

Declaration of conflicting interests
The Authors declare that there is no conflict of interest.

Europe PMC Funders Author Manuscripts

Jayawardana, & Mladenovic, 2019; Ecker, O' Reilly, Reid, & Chang, 2020; van der Meer & Jin, 2020). Research into the efficacy of corrections has found that, by-and-large, corrections reduce misconceptions but have limited effectiveness. That is, if people are exposed to an equivocal argument or outright falsehood, a subsequent rebuttal will typically reduce, but not reliably eliminate, the impact of the initially-provided misleading information (Chan, Jones, Hall Jamieson, & Albarracín, 2017; Ecker et al., 2022; Lewandowsky, Ecker, Seifert, Schwarz, & Cook, 2012; Walter & Tukachinsky, 2020).

Theoretical accounts of the impact of corrections have focussed mainly on cognitive factors, specifically breakdowns in information integration and memory updating processes, or selective retrieval (Ecker, Hogan, & Lewandowsky, 2017; Ecker, Lewandowsky, & Chadwick, 2020; Gordon et al., 2019; Kendeou, Butterfuss, Kim, & van Boekel, 2019; Kendeou, Walsh, Smith, & O' Brien, 2014; Swire, Ecker, & Lewandowsky, 2017; Walter & Tukachinsky, 2020). However, outside the laboratory information is rarely received without contextual social influences, and social context can thus be an important determinant of belief in contested claims and of the efficacy of corrections (Amin et al., 2017; Bode & Vraga, 2018; Hornsey & Fielding, 2017; Margolin, Hannak, & Weber, 2018; Mosleh, Martel, Eckles, & Rand, 2021; Trevors, 2021; Vlasceanu & Coman, 2021; Vlasceanu, Morais, Duker, & Coman, 2020). Some researchers have even argued that assessment of evidence, belief formation, and belief change are largely driven by social influences and are thus *primarily* functions of social cognition (Cohen, 2003; Kahan, Jenkins-Smith, & Braman, 2011).

To illustrate, some work assessing the efficacy of corrections has investigated the role of source credibility. This research has found that initially-provided (mis)information has a stronger influence if it originates from a credible source (e.g., Swire, Berinsky, Lewandowsky, & Ecker, 2017; Traberg & van der Linden, 2022; Walter & Tukachinsky, 2020; also see Pornpitakpan, 2004), although source credibility effects seem to depend on people paying attention to the source (e.g., Sparks & Rapp, 2011; also see Albarracín, Kumkale, & Poyner-Del Vento, 2017) and can be absent when the information comes from a news outlet and not a person or organization (Dias, Pennycook, & Rand, 2020; Pennycook & Rand, 2020). If a piece of misleading information is challenged by a correction, the credibility of the correction source also plays a role but seems to be less influential (e.g., Swire, Berinsky et al., 2017; Walter & Tukachinsky, 2020). More specifically, correction effectiveness seems to be influenced by the perceived trustworthiness of the correction source but not by its perceived expertise (Ecker & Antonio, 2021; Guillory & Geraci, 2013; also see Connor Desai, Pilditch, & Madsen, 2020; O'Rear & Radvansky, 2020; Vraga & Bode, 2017).

Other work has investigated the impact of worldview (i.e., a person's fundamental "ideology" and values) and group identity and membership. Those studies focus on whether a false belief aligns with a person's worldview, or whether a correction comes from an in-group or out-group source. Some studies have reported a strong impact of worldview, reporting that worldview-threatening corrections (i.e., corrections of worldview-consistent misinformation through provision of worldview-inconsistent correct information) can be ineffective or even backfire (Ecker & Ang, 2019; Nyhan & Reifler, 2010). This

effect has been interpreted as a cognitive bias in line with motivated reasoning accounts (Edwards & Smith, 1996; Sleegers, Proulx, & van Beest, 2019; also see Kunda, 1990). However, it has been difficult to replicate these findings (Wood & Porter, 2019), and other studies have found that corrections can be equally effective irrespective of the source of the misinformation or its worldview congruence (Ecker, Sze, & Andreotta, 2021; Swire-Thompson, Ecker, Lewandowsky, & Berinsky, 2020; Weeks, 2015). Berinsky (2017) found that corrections from unlikely partisan sources (i.e., sources sympathetic to the misinformation) can be particularly effective, whereas others have found that corrective messages are most persuasive when they come from (online) friends (Hannak, Margolin, Keegan, & Weber, 2014; Margolin et al., 2018) or from groups rather than individuals (Vlasceanu & Coman, 2021). To the best of our knowledge, no study has directly compared the efficacy of in-group versus out-group corrections, although it has been reported that corrections can be ineffective even when they come from an in-group source (Prasad et al., 2009). Overall, while the evidence is still inconclusive, the emerging consensus position seems to be that worldview and group membership affect consumption of and reliance on information generally, but do not have a specific and reliable impact on the effectiveness of corrections (Swire, Berinsky et al., 2017; Swire-Thompson, DeGutis, & Lazer, 2020; Swire-Thompson, Ecker et al., 2020; Swire-Thompson, Miklaucic, Wihbey, Lazer, & DeGutis, 2021; Vlasceanu & Coman, 2021).

One pertinent socio-cognitive factor that has so far not been systematically investigated in the area of reducing belief in contested information is the role of social norms. In general, people are fundamentally motivated to behave in ways that enhance the likelihood of creating positive social relationships. People are thus known to regulate expressed beliefs and behaviours in order to achieve social acceptance or avoid social exclusion (e.g., Asch, 1956; Brown, Lewandowsky, & Huang, 2021; Cialdini & Goldstein, 2004; Hornsey & Fielding, 2017; Kahan, 2013). If an individual is exposed to normative information that contradicts their current belief (e.g., information that the vast majority of people does not share the person's belief), this is likely to result in cognitive dissonance that individuals are motivated to reduce (Festinger, 1957; Moscovici, 1980). Individuals may adjust their own position to conform for two main reasons (e.g., see Kaplan & Miller, 1987): (1) information obtained from others can be accepted as "evidence about reality" based on the assumption that a consensus is usually correct (i.e., *informational* influence); (2) people prefer social conformity and tend to show normatively acceptable belief expressions and behaviours because they desire social inclusion (i.e., *normative* influence).

Consistent with this, there is evidence that social peer norms can change attitudes, stereotypes, and behaviours, and enhance public conformity (e.g., Bolsen, Leeper, & Shapiro, 2014; Cialdini & Goldstein, 2004; Cialdini, Reno, & Kallgren, 1990; Cohen & Prinstein, 2006; Ferraro & Price, 2013; Goldberg, Linden, Leiserowitz, & Maibach, 2020; Lewandowsky, Cook, Fay, & Gignac, 2019; Moussaïd, Kämmer, Analytis, & Neth, 2013; Nolan, Schultz, Cialdini, Goldstein, & Griskevicius, 2008; Puhl, Schwartz, & Brownell, 2005; Schmiege, Klein, & Bryan, 2010; Schultz, Nolan, Cialdini, Goldstein, & Griskevicius, 2007; Sechrist & Milford, 2007; Stangor, Sechrist, & Jost, 2001; Tankard & Paluck, 2016). It is assumed that such normative influence is stronger if behaviours—including belief expressions—are public rather than private (Cialdini & Trost, 1998; Deutsch & Gerard,

1955; Lapinski & Rimal, 2005; MacDonald & Nail, 2005; but see Wood, 2000); this is in line with findings that participants take significantly longer to submit claim-belief ratings when they assume responses to be public rather than private (Quekett, 2016). It is also known that communicating an expert consensus is a powerful means by which to counter misconceptions (e.g., Lewandowsky, Gignac, & Vaughn, 2013: van der Linden, Leiserowitz, Feinberg, & Maibach, 2015; van der Linden, Leiserowitz, & Maibach, 2019). Finally, providing information about others' opinions can also be considered a social good per se, because it avoids distorted perception of public opinion and its potential impact on public discourse (Prentice & Miller, 1993; Sargent & Newman, 2021; Shamir & Shamir, 1997).

With regards to false beliefs, recent work has shown that perceived in-group norms can foster a person's conspiracy beliefs (Cookson et al., 2021), that a norming approach can decrease the sharing of online misinformation (Andı & Akesson, 2021) and increase the reporting of fake news (Gimpel, Heger, Olenberger, & Utz, 2021), and that evidence endorsed by others can induce stronger belief change (Vlasceanu & Coman, 2021). Here, in three experiments, we examined the potential of social norms to reduce belief in a contested claim. All experiments used claims that were worldview-congruent for most participants, as determined by a pilot rating. In Experiment 1, participants were provided with a claim refutation, as well as one or two (fictitious) descriptive point norms suggesting that a majority of peers did not endorse the claim. Claim belief was measured multiple times, with the belief expression being either private or public. A norm that was presented on its own was found to reduce claim belief somewhat. A subsequent refutation reduced claim belief further; the efficacy of that refutation was not significantly enhanced by a second, post-refutation norm, and the confidentiality manipulation had no substantial impact. Experiment 2 used only a public expression of belief, and a norm comprising four data points that were either narrowly or widely distributed (i.e., reflecting varying degrees of peer consensus). Norms were found to be ineffective. Experiment 3 replicated Experiment 2 but used graphically-displayed distributed norms comprising 25 data points. The norms were as effective as refutations at eliciting belief change, and there was some indication that the combination of a refutation and a norm intervention was most effective.

Pilot Study

Worldview-dissonant information is subject to enhanced scrutiny (Kahan et al., 2011), so belief change can be more difficult to achieve if a misconception supports an existing worldview (even though the evidence for this is mixed; for a recent discussion, see Ecker et al., 2021). Thus, norm effects may be most valuable with worldview-congruent (but false) beliefs. A claim was therefore selected that was generally endorsed in a pilot study but still amenable to belief reduction following a refutation. In the pilot study, $N = 29$ undergraduate students from the University of Western Australia (UWA) were presented with four claims; each was embedded in a short text that provided some background information, before an argument was made in support of the claim. Participants rated how confidently they agreed with each claim on a 1-10 scale, before being presented with refutations counterarguing each claim, and subsequently re-rating their agreement with the original claims. All pilot claims and refutations, as well as the associated descriptive statistics, are provided in

the Supplement (available at https://osf.io/ekxzy/). The claim selected for the main study was "*Accepting a slightly increased uptake of refugees will have a positive impact on the Australian economy within 20 years*." The claim itself was presented by an apparent expert and supported by an argument that most refugees integrate into society quickly. The refutation of this claim was provided by another apparent expert, referred to data from overseas, and stated that in reality "*each additional refugee comes at a certain long-term cost*." The claim's mean pre-refutation belief rating was $M = 6.62$ ($SD = 2.53$); its mean post-refutation belief rating was $M = 4.97$ ($SD = 2.63$). This was a significant reduction, $F(1, 28) = 21.69$, $MSE = 1.83$, $p < .001$, $\eta p^2 = .44$.

Experiment 1

Experiment 1 tested whether social norms can reduce belief in a contested claim, and whether such an effect is dependent on the public nature of belief expressions. Participants were tested in small groups, and were led to believe that their belief expressions would either remain confidential (private), or be shared with other participants in a group discussion at the end of the experiment (public). Participants' endorsement of the claim was measured at three time-points: (1) after initial reading of the claim (baseline); (2) after receiving a descriptive social norm indicating low peer endorsement of the claim; and (3) after receiving a refutation that was or was not combined with an additional post-refutation norm indicating even lower peer endorsement. Participants were told that the norm originated from a sample of their peers (i.e., students at the same university), because the extent of normative influence may depend on the degree to which the individual views the comparison group as relevant to their social identity (Abrams, Wetherell, Cochrane, Hogg, & Turner, 1990; Terry & Hogg, 1996; Terry, Hogg, & White, 1999). Finally, and following precedent (e.g., Ecker, O' Reilly et al., 2020), a series of inference questions was given to participants as a more indirect measure of their claim belief. Measures of social assertiveness and need for authenticity were taken as covariates, as public belief expressions may be influenced by these traits (MacDonald & Nail, 2005; Santee & Maslach, 1982; Williams & Warchal, 1981; Wood, Linley, Maltby, Baliousis, & Joseph, 2008). It was hypothesized that both a pre-refutation descriptive norm and a refutation would reduce claim endorsement, and that the combination of the refutation with a post-refutation norm would be most effective. Further, it was hypothesized that norm effects would be greater in the public condition.[1]

Method

All experiments were approved by the Human Research Ethics Office of the University of Western Australia (approval number RA/4/1/8104). Experiment 1 used a 2 × 2 between-participants design with factors confidentiality (private vs. public) and post-refutation norm (absent vs. present).

[1]We note that assessing whether the interventions offset the impact of initial exposure to the claim text completely (i.e., assessing continued influence) would require an additional control condition in which participants are not exposed to the claim text. As our focus was on the question if normative information can reduce claim belief at all, such a control condition was deemed unnecessary.

Participants—An a-priori power analysis (using G*Power; Faul, Erdfelder, Lang & Buchner, 2007) suggested that detecting an interaction effect of medium size (f = .25; α = .05; 1 – β = .80) would require a minimum sample size of 128 participants. We recruited N = 143 UWA undergraduate students, who participated for course credit (99 females, 44 males; mean age M = 19.92 years, SD = 4.41). Participants were randomly assigned to one of the four conditions, subject to a constraint of keeping the number of participants in each condition roughly equal. Participants signed up individually using an online sign-up site, but were tested in a lab room in groups of up to five participants.

Materials—The experimental survey used Qualtrics software (Qualtrics, Provo, UT). It presented a brief fictional text on the subject of refugees in Australia: *"The great debate about how many refugees Australia should take in continues. The main aim is to keep refugees safe, and while there is some immediate cost, Jack Western from the Australian Refugee Resource Centre argued today that there was no reason to be concerned about economic consequences. 'Within a short period of time, most refugees integrate into society, and a moderate increase in refugees will have a positive impact on the Australian economy within two decades.'"* Following this was a specific claim that *"Accepting a slightly increased uptake of refugees will have a positive impact on the Australian economy within 20 years."*

A second text contained the refutation: "*Later today, Dr Michael Smith from the Centre for Independent Studies strongly refuted earlier statements regarding an economic benefit associated with asylum seekers. According to Smith, recent modelling by the Centre shows that there is 'no economic gain to be expected in the foreseeable future by increasing the intake of asylum seekers in Australia. Evidence also comes from Germany where the intake of asylum seekers over the years has been an economic drain. Unfortunately, the reality is that each additional refugee comes at a certain long-term cost.'"*

The first fictional norm of peer endorsement read: *"37 out of 100 UWA students endorsed this claim."* The second, post-refutation norm read *"After receiving the refutation, 19 out of 100 UWA students endorsed this claim."*

Claim endorsement was measured on an 11-point Likert scale ranging from "confidently disagree" (0) to "confidently agree" (10); in the following, these will be referred to as "belief ratings." Additionally, six inference items were used to indirectly measure participants' endorsement of the claim. An example item is *"Australia is facing many costly issues and record-level debt; therefore we simply cannot afford the additional cost of taking on more refugees"* with agreement rated on a 10-point Likert scale ranging from "confidently disagree" to "confidently agree". Inference scores were calculated as the mean of the inference-item responses and transformed onto a quasi-continuous scale ranging from 0 to 1. A full list of the inference items can be found in the Supplement.

Social assertiveness (SA) was measured using 11 relevant items from the 50-item college self-expression scale (Galassi, DeLo, Galassi & Bastien, 1974). An example item is *"Do you freely volunteer information or opinions in class discussions?"* Responses were recorded on

a five-point Likert scale, ranging from “rarely/never” (1) to “always/almost always” (5). A full list of the SA items can be found in the Supplement.

Need for authenticity (NFA) was measured using the need for authenticity scale (Wood et al., 2008). An example item is *“I am strongly influenced by the opinions of others.”*The scale comprises 12 items, and responses were recorded on a seven-point Likert scale ranging from “does not describe me at all” (1) to “describes me very well” (7). A full list of the NFA items can be found in the Supplement.

Procedure—Participants completed the experiment in a lab environment on individual computers, separated by privacy blinds. All participants provided written informed consent after reading an ethics-approved information sheet. Participants were informed that their beliefs regarding a topic of general interest would be measured. In the private condition, participants were told that all of their responses would be kept confidential. Conversely, participants in the public condition were told that their responses would be potentially disclosed to the other participants and the experimenter in a group discussion at the end of the experiment. Both groups received multiple reminders of this throughout the experiment.

The first text was then presented, including the critical claim. Participants then completed the first belief rating (time-point 1). Following this, the first norm was presented, after which participants completed the second belief rating (time-point 2). Subsequently, the second text containing the refutation was presented, followed by the second, post-refutation norm in the relevant conditions. Participants then completed an unrelated 1-minute distractor task (a word puzzle), followed by the six inference items, and the third and final belief rating (time-point 3). Finally, participants completed the social assertiveness and need for authenticity questionnaires. At the end of the experiment, participants in the public condition were engaged in a brief discussion; this was conducted simply to avoid deception of participants. All participants were then fully debriefed. The experiment lasted approx. 20 minutes; it was preceded by an unrelated experiment of the same approx. duration.

Results

Belief change—Belief change across all three time-points is illustrated in Figure 1. Baseline claim belief was $M = 6.14$ ($SD = 2.49$) and thus, as expected, above the midpoint. Regarding the impact of the pre-refutation norm, mean belief-change scores (belief rating 2 – belief rating 1) differed significantly from zero in both the private condition, $M = -0.29$ ($SD = 0.93$), $t(71) = -2.67$, $d = 0.31$, $p = .009$, and the public condition, $M = -0.37$ ($SD = 0.83$), $t(70) = -3.71$, $d = 0.44$, $p < .001$. This demonstrated a small belief-reducing effect of the initial, pre-refutation norm. A one-way ANCOVA with confidentiality (private, public) as the between-subjects factor and social-assertiveness and need-for-authenticity z-scores as covariates (in a full-factorial model) returned no other significant effects, all Fs$(1, 135) < 2.37$, $p > .126$.

Regarding belief change from time-point 2 to time-point 3, there were four conditions defined by confidentiality (private, public) and post-refutation norm (absent, present) factors. Mean belief-change scores (belief rating 3 – belief rating 2) differed significantly from zero in all four conditions (private, norm-absent: $M = -1.14$ [$SD = 1.13$], $t[35] = -6.07$, $d = 1.01$,

$p < .001$; private, norm-present: $M = -0.97$ [$SD = 1.56$], $t[35] = -3.74$, $d = 0.62$, $p < .001$; public, norm-absent: $M = -0.94$ [$SD = 1.28$], $t[34] = -4.35$, $d = 0.74$, $p < .001$; public, norm-present: $M = -1.56$ [$SD = 1.50$], $t[35] = -6.22$, $d = 1.04$, $p < .001$), indicating that claim belief was reduced in response to a refutation.

A two-way ANCOVA additionally yielded an interaction between confidentiality and post-refutation norm, $F(1, 127) = 5.70$, $MSE = 1.83$, $p = .018$, $\eta p^2 = .04$; however, individual contrasts between norm-absent and norm-present conditions were nonsignificant (private: $F[1, 127] = 3.42$, $p = .067$, $\eta p^2 = .03$; public: $F[1, 127] = 2.30$, $p = .132$, $\eta p^2 = .02$), and the interaction was no longer significant when covariates were removed from the model, $F(1, 139) = 2.86$, $MSE = 1.90$, $p = .093$, $\eta p^2 = .02$.[2] There was also a three-way interaction of post-refutation norm and both covariates, $F(1, 127) = 10.68$, $p = .001$, $\eta p^2 = .08$, which was difficult to interpret.[3] All other effects were nonsignificant, all $F(1, 127) \leq 2.47$, $p \geq .118$.

Inference scores—A confidentiality × post-refutation norm ANCOVA including both covariates revealed no significant effects, all $F(1, 127) \leq 3.79$; $p \geq .054$).

Discussion

The results from Experiment 1 demonstrated that descriptive norms can reduce belief in a claim. This provides support for the general hypothesis that social norms can play a role in belief change, and more specifically, that norms can be used—like a refutation—to directly reduce people's reliance on previously endorsed information. This is in line with other studies investigating the effects of social consensus (e.g., Bolsen et al., 2014; Lewandowsky et al., 2019; Puhl et al., 2005; Stangor et al., 2001). However, the effect did not differ substantially between private and public conditions, and there was no meaningful difference in belief change between participants who were highly assertive with high need for authenticity, and participants who were unassertive with low authenticity needs. This suggests that the norm effect was largely driven by informational influence rather than normative influence per se.

In line with this, it could be argued that participants may have been generally uncertain about their belief in the claim, which required them to predict economic impacts (of accepting more refugees). Predicting such impacts is a difficult task (e.g., Angner, 2006), and most participants arguably had only a surface level of relevant knowledge. Thus, if participants lacked confidence in their initial belief rating, receiving a norm that suggested their peers largely disagreed with the claim may have led participants to assume that their peers were better informed, prompting them to lower the subsequent rating. Indeed, social influence tends to increase as tasks get more difficult and an objectively correct response becomes harder to determine (Jaeger, Lauris, Selmeczy, & Dobbins, 2012; Koop, King, & Kauffman, 2021). If this account is correct, it could help explain why belief ratings decreased in response to the descriptive norm largely independently of confidentiality and participants' social assertiveness and need for authenticity. Future research could aim to implement a stronger confidentiality manipulation (e.g., individuals entering their

[2]Throughout the paper, results from analyses including vs. excluding the covariates were equivalent unless otherwise noted.

Europe PMC Funders Author Manuscripts

own responses in closed cubicles vs. group testing with verbal responses entered by the experimenter).

Consistent with previous research (see Chan et al., 2017; Lewandowsky et al., 2012; Walter & Tukachinsky, 2020), a refutation was found to additionally reduce claim belief, with the effect being greater than that of the descriptive norm. A key ingredient of a successful refutation is specification of reasons why a claim is wrong, alongside provision of an alternative account (e.g., Ecker, Lewandowsky & Apai, 2011; Paynter et al., 2019; Seifert, 2002; Swire, Ecker et al., 2017). Experiment 1 demonstrated that such an approach can also be successful when refuting a prediction, which by definition cannot utilize already-observed direct evidence regarding the specific outcome. Providing a supporting descriptive norm with a refutation did not significantly increase the refutation's overall effectiveness. This questions the general utility of a norming approach; however, it is worth keeping in mind that Experiment 1 presented participants with two successive peer norms, and it is possible that the first stand-alone norm weakened the effect of the second, post-refutation norm. In other words, the initial norm already communicated a perhaps unexpectedly low peer endorsement, and its post-refutation decrease (from 37 to 19 out of 100) may not have been perceived as particularly meaningful. In order to build a stronger evidence base and generalize to other claims, we ran Experiment 2.

Experiment 2

Experiment 2 again examined whether social norm information can reduce the endorsement of questionable worldview-congruent claims, either in isolation or in combination with a refutation. Moreover, Experiment 2 used a distribution rather than a point norm, and tested whether the shape of a norm distribution—namely whether it is narrow or wide, reflecting strong or weak consensus—would impact the effectiveness of the norm.

Much of the social norming literature has used mean-based point norms (e.g., average energy consumption in your neighbourhood is *x* kWh per day; Schultz et al., 2007). Similarly, Experiment 1 also used a point norm (*x* out of 100 peers endorsed the claim). However, individuals may be more concerned with their position in a distribution, rather than a comparison to a point norm (Brown et al., 2021). For example, a particular distance from the mean can have different implications depending on the shape of the underlying distribution— using 3 kWh more a day than the average may be concerning if that puts a household at the 95th percentile but less concerning if it puts them only at the 55th percentile. Theoretical models of judgement and decision making such as range-frequency theory (Parducci, 1965) and the decision-by-sampling model (Stewart, Chater, & Brown, 2006) also suggest that individuals often base their actions on a series of smaller-than and larger-than comparisons within a social context, and their estimation of where they fall within a perceived distribution. Thus, Aldrovandi and colleagues proposed a distribution-based approach to social norming (Aldrovandi, Brown, & Wood, 2015; Aldrovandi, Wood, Maltby, & Brown, 2015). Normative information that highlights the position within a distribution relative to others has been found to better predict and influence healthy eating behaviours (Aldrovandi, Brown et al., 2015), alcohol consumption (Moore et al., 2016), and perception

of indebtedness (Aldrovandi, Wood et al., 2015). Experiment 2 therefore used a distribution-based approach.

Participants were presented with predictive claims that were designed to be endorsed by the majority of participants. Unlike in Experiment 1, claims also included quantitative predictive estimates (e.g., specifying by how much an increased refugee uptake might boost the economy). Participants then received a refutation and/or fictional norming information ostensibly indicating that the last four participants had rejected the claim. Participants rated their endorsement of the claims both before and after the refutation/norming manipulation. For each claim, participants additionally provided their own predictive estimate and responded to a series of post-manipulation inference questions to indirectly measure claim support.[4] Participants were tested in small groups, and discussed their responses with the experimenter and other participants at the end of the experiment; this "public" context was chosen to maximize normative influence.

We hypothesized that receiving either norming information or a refutation would produce significant belief change, and that combining norm information and a refutation would be more effective in evoking belief change than either individually. We expected that the effect of the norm might depend on the norm's distribution: on the one hand, a narrow norm may provide objectively stronger evidence and should therefore be more convincing than a wide norm (Molleman et al., 2020); on the other hand, a wide norm may be more effective because it may foster belief change by making it easier (in comparison to a narrow norm) for individuals to find and express a compromise between their true belief and the norm (i.e., a response that is socially acceptable as well as sufficiently authentic; see Brown et al., 2021; Hornsey & Jetten, 2004; Santee & Maslach, 1982; Wood, 2000).

Method

Experiment 2 comprised two parallel sub-experiments that we will refer to as Experiments 2A and 2B, each with three within-subjects conditions. Experiment 2A had refutation, narrow-norm, and wide-norm conditions. Experiment 2B combined norms and refutations, and thus compared a refutation condition (which was identical to Experiment 2A) with refutation-plus-narrow-norm and refutation-plus-wide-norm conditions.[5] Thus, each participant received three claims that were then challenged by a refutation and/or norm. Both experiments additionally included a no-refutation filler claim to reduce the build-up of refutation expectations.

Participants—A total of $N = 144$ first-year psychology students from UWA participated in the current research in exchange for course credit ($n = 72$ in Experiment 2A; $n = 72$ in Experiment 2B). The sample comprised 95 females, 48 males, and one participant of undisclosed gender; mean age was $M = 20.63$ years ($SD = 5.15$), ranging from 17 to 45 years. Participants signed up individually but were tested in small groups with up to 5

[4]Given the equivocal covariate effects in Experiment 1, need for authenticity and social assertiveness measures were included in Experiments 2 and 3 only for the sake of consistency; in the remainder of the paper, following the a-priori analysis plan, we focus on analyses excluding covariates; analyses including the covariates are presented in the Supplement for the sake of completeness.
[5]It was decided not to run Experiment 2 as a fully-crossed 2 × 3 (or 1 × 5) design for pragmatic reasons to do with the number of required claims (including additional filler claims), and testing-time constraints.

participants. Groups were randomly assigned to Experiments 2A and 2B (with the constraint of roughly equal participant numbers). Sensitivity analyses suggested that sample size was sufficient to detect an effect of $f = 0.26$ in each sub-experiment, or $f = 0.24$ for a 2 × 2 within-between interaction ($\alpha = .05$; $1 - \beta = .80$).

Materials

Claims: Three experimental claims and one filler claim were used. As in Experiment 1, selected claims were predictive in nature, concerned currently debated topics, and could be assumed to be endorsed by the majority of participants. The four topics chosen were (i) refugee intake and its effect on the Australian economy (as in Experiment 1); (ii) a move towards renewable energy and its effect on the Australian job market; (iii) the introduction of medicinal cannabis and its effect on the Australian healthcare system; and (iv) the provision of government-funded shelter to reduce homelessness (filler claim). Claims included a specific prediction (e.g., that a slight increase in refugees would boost the economy by an estimated 2% of annual GDP). Assignment of claims to conditions was counterbalanced across participants, and condition order was randomized (with the constraint that the filler claim always came second).

As in Experiment 1, each claim was countered by a refutation that provided some background information, followed by an explicit statement refuting the respective claim. In the filler condition, a neutral statement was provided that merely suggested the issue was complex. All claims and refutations can be found in the Supplement.

Norms: For each claim, fictional norming information was generated in the form of a distribution of four values, allegedly coming from the last four participants. The four values were calculated by multiplying the claim value (e.g., economic boost of 2%) by -1/6, -1/12, 1/12, and 1/6 (narrow distribution) or -2/3, -1/3, 1/3, and 2/3 (wide distribution), respectively. A small jitter of 1/20th of the claim value was added to the distribution values such that the distribution was roughly centred on zero, without pairs of distribution values having the same absolute value. Finally, values were rounded. For example, the norming information for the refugee topic read "*The original claim was: 'A slightly increased uptake of refugees will have a positive impact on the Australian economy. The estimated effect of a moderate increase in refugee intake on the Australian economy will be +2%.' The true effect on the Australian economy was estimated by the last four participants as: -0.2% | -0.1% | +0.3% | +0.4%"* (in the narrow-norm conditions) or "*-1.2% | -0.6% | +0.8% | +1.4%*" (in the wide-norm conditions).

Belief ratings: Participants indicated their claim belief on a 0 (confidently disagree) to 10 (confidently agree) Likert scale. Belief rating 1 followed initial presentation of the claim; belief rating 2 was given after presentation of the refutation and/or norm.

Predictive estimates: For each topic, participants provided a post-manipulation estimate of the effect specified in the claim. Responses were given on 16-point or 21-point scales; the scale range included the claim value. For example, for the refugee claim, participants read "*If Australia were to slightly increase its refugee uptake, what do you estimate the effect to*

be on the Australian economy over the next 20 years (in percentage terms)?" with response options ranging from -3% to +3% in 0.3% increments. All estimation scales are provided in the Supplement. For analysis, estimate values were divided by the respective claim value to make them comparable across the different claim scales, such that a predictive-estimate score of 1 reflected full endorsement of the claim value, and a score of 0 reflected a fully effective refutation/norm manipulation, with scores < 0 reflecting hypercorrection (scores > 1 would reflect *increased* belief).

Inference questions: Participants completed a set of four inference questions per claim to indirectly measure claim belief. These used 10 or 11-point Likert scales and either measured agreement with a statement (e.g., "*Within a few decades, refugees contribute more to the economy than they take out*" – *confidently disagree* [1] to *confidently agree* [10]) or asked for a policy-related decision (e.g., *How would you change the rate of refugee intake?* [-50% to +50%, in 10% increments]). Inference scores were again calculated as the mean of the inference-question responses and transformed onto a quasi-continuous scale ranging from 0 to 1. All inference questions are provided in the Supplement.

Procedure—Participants read an ethics-approved information sheet and provided written informed consent. For each topic, participants first read the claim statement and provided their initial endorsement (belief rating 1). Then, depending on experimental condition, they received a refutation and/or norm information (or the neutral statement for the filler claim; in the refutation-plus-norm conditions of Experiment 2B, the refutation preceded the norm information). This was followed by belief rating 2, and participants' estimation of the true effect. Participants then responded to the four inference questions. This sequence was repeated for all four topics. Finally, participants completed the authenticity and social-assertiveness questionnaires, partook in a brief discussion, and were then fully debriefed. The experiment lasted approx. 20 minutes; it was preceded by an unrelated experiment of the same approx. duration.

Results

Belief change—Belief-change scores (belief rating 2 – belief rating 1) are shown in Figure 2.[6] Change scores were significantly different from zero in the refutation condition of Experiment 2A, $M = -1.53$, $SD = 1.76$, $d = 0.87$, and all conditions of Experiment 2B (refutation: $M = -1.75$, $SD = 1.98$, $d = 0.89$; refutation-plus-narrow-norm: $M = -1.61$, $SD = 2.14$, $d = 0.75$; refutation-plus-wide-norm: $M = -1.15$, $SD = 1.77$, $d = 0.65$), all $t(72) \leq -5.54$, all $p < .001$. There was no significant belief change in the norm-only conditions of Experiment 2A (narrow: $M = -0.21$, $SD = 1.15$, $d = 0.18$; wide: $M = -0.10$, $SD = 1.15$, $d = 0.08$), $t(72) \geq -1.54$, $p \geq .129$. This established that claim belief was reduced significantly by a refutation (either with or without an additional norm) but not a stand-alone norm.

In Experiment 2A, a repeated measures ANOVA on belief-change scores yielded a significant main effect of condition, $F(2, 142) = 21.57$, $MSE = 2.11$, $p < .001$, $\eta p^2 = .23$. Planned contrasts confirmed significant differences between the refutation condition

[6]Note that for both Experiment 2 and Experiment 3, additional figures showing dependent measures across conditions by topic are provided in the Supplement.

and both the narrow-norm, $F(1, 71) = 28.56$, $MSE = 2.19$, $p < .001$, $\eta p^2 = .29$, and the wide-norm condition, $F(1, 71) = 29.92$, $MSE = 2.46$, $p < .001$, $\eta p^2 = .30$. The effect of norm width was nonsignificant, $F < 1$. The analogous ANOVA in Experiment 2B showed no significant effect of condition, $F(2, 142) = 2.14$, $MSE = 3.28$, $p = .121$, $\eta p^2 = .03$.

Predictive estimate scores—Predictive estimate scores are shown in Figure 3. In Experiment 2A, a repeated measures ANOVA yielded a significant main effect of condition, $F(2, 142) = 16.77$, $MSE = 0.20$, $p < .001$, $\eta p^2 = .19$. Planned contrasts revealed significant differences between the refutation condition ($M = -0.03$, $SD = 0.52$) and both the narrow-norm condition ($M = 0.20$, $SD = 0.40$), $F(1, 71) = 9.18$, $MSE = 0.21$, $p = .003$, $\eta p^2 = .11$, and the wide-norm condition ($M = 0.40$, $SD = 0.49$), $F(1, 71) = 27.29$, $MSE = 0.25$, $p < .001$, $\eta p^2 = .28$. Mirroring the belief change analysis, this indicated that a refutation reduced beliefs more than either norm alone. There was also a significant difference between narrow-norm and wide-norm conditions, $F(1, 71) = 9.91$, $MSE = 0.15$, $p = .002$, $\eta p^2 = .12$, indicating that the narrow norm was more effective than the wide norm at reducing predictive estimates.

In Experiment 2B, there was a significant main effect of condition, $F(2, 142) = 5.44$, $MSE = 0.19$, $p = .005$, $\eta p^2 = .07$. Planned contrasts revealed significant differences between the refutation condition ($M = -0.03$, $SD = 0.48$) and the narrow-norm condition ($M = 0.17$, $SD = 0.40$), $F(1, 71) = 10.47$, $MSE = 0.14$, $p = .002$, $\eta p^2 = .13$. There was no difference between the refutation and the wide-norm condition ($M = -0.03$, $SD = 0.49$), $F < 1$. There was, however, a significant difference between narrow and wide-norm conditions, $F(1, 71) = 8.30$, $MSE = 0.18$, $p = .005$, $\eta p^2 = .10$. This indicated that the refutation and refutation-plus-wide-norm conditions had similar impacts on predictive estimates, and that the wide norm was more effective than the narrow norm when combined with a refutation.

A 2 (norm width; within-subjects) × 2 (refutation; between-subjects) ANOVA across experiments returned no significant main effect of norm width, $F < 1$, but a significant main effect of refutation, $F(1, 142) = 16.02$, $MSE = 0.23$, $p < .001$, $\eta p^2 = .10$, as well as a significant interaction, $F(1, 142) = 18.05$, $MSE = 0.17$, $p < .001$, $\eta p^2 = .11$, indicating that the refutation effect was stronger in the wide-norm condition. A post-hoc comparison confirmed a significant effect of a refutation in the wide-norm condition (i.e., lower estimates in the refutation-plus-wide-norm condition than the stand-alone wide-norm condition), $F(1, 142) = 27.85$, $MSE = 0.24$; $p < .001$, $\eta p^2 = .16$; this was absent in the narrow-norm condition, $F < 1$.

Inference scores—Inference scores are shown in Figure 4. In Experiment 2A, there was a significant main effect of condition, $F(2, 142) = 10.46$, $MSE = 0.02$, $p < .001$, $\eta p^2 = .13$. Planned contrasts revealed significant differences between the refutation condition ($M = 0.54$, $SD = 0.19$) and both the narrow-norm condition ($M = 0.61$, $SD = 0.18$), $F(1, 71) = 10.08$, $MSE = 0.02$, $p = .002$, $\eta p^2 = .12$, and the wide-norm condition ($M = 0.64$, $SD = 0.16$), $F(1, 71) = 19.02$, $MSE = 0.02$, $p < .001$, $\eta p^2 = .21$. There was no significant difference between narrow and wide-norm conditions, $F(1, 71) = 1.51$, $MSE = 0.02$, $p = .224$, $\eta p^2 = .02$. Mirroring the belief-change and predictive-estimates analyses, this indicated

that the refutation condition resulted in lower inference scores than the stand-alone norm conditions.

In Experiment 2B, there was no significant main effect of condition, $F(2, 142) = 2.68$, $MSE = 0.02$, $p = .072$, $\eta p^2 = .04$. Planned contrasts showed that the narrow-norm condition ($M = 0.60$, $SD = 0.14$) differed marginally from the refutation condition ($M = 0.55$, $SD = 0.18$), $F(1, 71) = 4.05$, $MSE = 0.02$, $p = .048$, $\eta p^2 = .05$, as well as the wide-norm condition ($M = 0.55$, $SD = 0.19$), $F(1, 71) = 4.67$, $MSE = 0.02$, $p = .034$, $\eta p^2 = .06$. There was no significant difference between refutation and wide-norm conditions, $F < 1$. This mirrors the predictive-estimates analysis, suggesting that refutation and refutation-plus-wide-norm conditions had slightly greater impact than the refutation-plus-narrow-norm condition. However, these marginal effects were nonsignificant in the analysis including covariates (see Supplement).

A 2 (norm width; within-subjects) × 2 (refutation; between-subjects) ANOVA across experiments showed a nonsignificant main effect of norm width, $F < 1$. However, the main effect of refutation was significant, $F(1, 142) = 5.37$, $MSE = 0.04$, $p = .022$, $\eta p^2 = .04$, and so was the interaction, $F(1, 142) = 5.90$, $MSE = 0.02$, $p = .016$, $\eta p^2 = .04$, indicating again that the refutation effect was stronger with a wide norm than a narrow norm. A post-hoc comparison confirmed a significant effect of a refutation in the wide-norm condition, $F(1, 142) = 9.88$, $MSE = 0.03$, $p = .002$, $\eta p^2 = .07$.

Discussion

Experiment 2 investigated whether social norm information would reduce the acceptance of equivocal, worldview-congruent claims, either in isolation or in combination with a refutation. Moreover, it investigated whether the shape of the norm distribution— narrow or wide—would impact the effectiveness of the norm. As in Experiment 1, a refutation effected belief change, whether measured directly, via more indirect inferential reasoning questions, or predictive estimates. Again, this result is consistent with previous findings of refutations reducing misconceptions (e.g., Ecker et al., 2011; Paynter et al., 2019; Seifert, 2002; Swire, Ecker et al., 2017).

Contrary to predictions, however, Experiment 2 found that social norms, both individually and when combined with a refutation, did not significantly reduce participants' claim beliefs. While stand-alone norms produced predictive-estimate scores that were substantially lower than the claim values, this cannot be interpreted as a corrective effect, as there was no no-intervention control condition. Regarding the norm's distribution width, there was some indication that a stand-alone norm more effectively reduced predictive estimates when it was narrow rather than wide; this is consistent with the prediction that a narrow distribution should provide objectively stronger evidence and may therefore be more convincing than a wide distribution, especially in the case of a stand-alone norm. However, a wide norm was found more impactful than a narrow norm when combined with a refutation, which is consistent with the prediction that a wide norm could make it is easier to find a compromise between one's true belief and the norm when one receives strong counterevidence from a refutation. This could suggest that in isolation, the stronger consensus signal associated with a narrow norm is particularly persuasive, but that a wider norm becomes more influential

when paired with a refutation that highlights reasons for disagreement about a matter, perhaps because a wider norm then appears more plausible, thus providing a more credible signal. Without replication, however, this interpretation remains speculative. Overall, while these findings lend some preliminary support to the notion that the effect of a social norm on an individual's beliefs may be influenced by the shape of its underlying distribution (Aldrovandi, Brown et al., 2015; Stewart et al., 2006), the effects should be interpreted with caution given their inconsistent nature.

The relative ineffectiveness of social norms we observed is somewhat inconsistent with the norming literature (e.g., Cialdini et al., 1990; Deutsch & Gerard, 1955; Schultz et al., 2007), especially research using a distribution-based approach to social norming (e.g., Aldrovandi, Brown et al., 2015; Aldrovandi, Wood et al., 2015). However, it is important to point out that these studies largely focused on behaviour change, not belief change, and that not all research in this domain has found norms to be effective. For example, Silva and John (2017) found a descriptive norm about students' payments of tuition fees ineffective in improving payment rates (also see Thombs, Dotterer, Olds, Sharp, & Raub, 2004).

Given that social norming relies on the reference group being relevant to a person's social identity (Abrams et al., 1990; Terry & Hogg, 1996; Terry et al., 1999), one reason for their observed ineffectiveness in Experiment 2 may be that participants only weakly identified with the reference group—a plausible notion given the heterogeneous nature of the student body enrolled in first-year psychology[7] (Silva & John, 2017). Alternatively, the lack of effect may be related to the norms' specific implementation in Experiment 2. Norms may have lacked credibility because they were based on only four values, and participants may have assumed that the last four participants lacked relevant knowledge. Perceived credibility is an important determinant of persuasive communication (Aronson, Turner, & Carlsmith, 1963; Jaccard, 1981; Thombs et al., 2004), and research on the wisdom-of-crowds heuristic shows that people consider judgements of larger groups more accurate than judgements of smaller groups (Darke et al., 1998; Mannes, 2009). Thus, the norms may have not provided the strength of evidence required to change public belief expressions (Jellison & Mills, 1969). Moreover, the norms used in Experiment 2 focused on the predicted estimates associated with the claims (e.g., the estimated size of the effect of an additional refugee uptake on future GDP), rather than directly on peer endorsement of the claims, which may be a more persuasive approach.

For the next step, it was therefore decided to focus on norm persuasiveness. To this end, Experiment 3 presented both claim-endorsement and predicted-estimates norms. Both norms again used a distribution-based approach but with a larger number of data points; to facilitate this, a graphical presentation format was employed (Ancker, Senathirajah, Kukafka, & Starren, 2006; Eberhard, 2021; Gelman, Pasarica, & Dodhia, 2002; Meyer, Shamo, & Gopher, 1999).

[7]First-year psychology units are taken as broadening units by UWA students from multiple degrees.

Europe PMC Funders Author Manuscripts

Experiment 3

Experiment 3 was a conceptual replication of Experiment 2, and thus again involved two parallel sub-experiments (3A and 3B). However, instead of using a norm comprising four verbally-communicated values, ostensibly obtained from the previous 4 participants, Experiment 3 presented two norms, each using a graphical representation of 25 values, ostensibly obtained from a representative peer sample.

Method

Participants—A total of $N = 154$ first-year psychology students from UWA participated in exchange for course credit. The sample comprised 101 females, 52 males, and 1 participant of undisclosed gender; mean age was $M = 22$ years ($SD = 7.40$), with age range 17 to 55 years. Participants signed up individually, but were tested in groups of up to 5 participants. Participants were randomly assigned to one of the two experimental groups, with the constraint of achieving approximately equal numbers across Experiments 3A ($n = 76$) and 3B ($n = 78$). Sensitivity analyses suggested that sample size was sufficient to detect an effect of $f = 0.26$ in each sub-experiment, or $f = 0.23$ for a 2 × 2 within-between interaction ($\alpha = .05$; $1 - \beta = .80$).

Materials

Claims: Claims were largely identical to Experiment 2; for some claims, additional information on specific predicted outcomes was added, to allow for more concrete estimates (e.g., instead of stating merely that refugee intake would boost the Australian economy by an estimated 2% of GDP, additionally stating that this was approximately $1,300 per person in per-capita GDP terms).

Norms: Distributed claim-endorsement norms were presented as a graphical representation of 25 individual ratings, on a 0-100 scale, ostensibly collected from a representative sample of 25 UWA students. In reality, the distributions were obtained by drawing a random sample of 25 values from a normal distribution with pre-determined mean and variance (truncated at 0). For each claim, the distribution mean was set at 0.25 × the mean claim endorsement obtained from a pilot rating (approx. 60/100). The standard deviations used for the narrow and wide norms were set at $SD = 5$ and $SD = 15$, respectively. The samples thus formed either a narrow cluster or a wide cluster around the mean. If necessary, sampling was repeated until the means of the narrow and wide sample distributions approximated each other (actual sample means were approximately 17). Sampling and plotting was performed using the *rtruncnorm* and *plot* functions of the R programming language (R Core Team, 2013); the resulting plot was a simple graph of semi-transparent blue circles, which allowed for a visualisation of sample density (i.e., overlapping samples resulted in a more saturated colour; see Figure 5). The graphical norms were presented together with the corresponding original claims and the statement "*We surveyed a representative sample of 25 UWA students […] this is how your peers rated their belief in the claim*."

Predictive-estimate norms were also presented as a graphical representation of 25 individual predictions. These were again ostensibly collected from a representative sample of UWA

students, but in actual fact represented a random draw from a pre-specified normal distribution. For each claim, the mean of the normal distribution was set at 0.10 × the respective claim value (i.e., the estimate mentioned in the claim; e.g., a predicted $1300 boost to the economy). The standard deviations were set at *SD* = 0.05 × the claim value (narrow norm) and *SD* = 0.15 × the claim value (wide norm), respectively. The presented scale was claim-specific; each scale used the range from 0 to the claim value, with a buffer on either side spanning approximately 15-20 percent of the claim value. Figure 5 illustrates this using the example of the refugees claim, where the claim value was $1300, and the range was set as -$200 to $1600. Samples were presented in red to differentiate the prediction norms from the endorsement norms. The norms were presented alongside the statement "*Let us look at what the representative sample of your peers thought. […] The true effect […] was estimated by your peers to be…*" Prior to commencing the task, participants were provided with examples and detailed explanations regarding how to interpret the graphical norm representations. All norms are provided in the Supplement.

Belief ratings: Unlike Experiment 2, participants indicated their claim belief on 0-100 scales, to match the scale used for presenting claim-endorsement norms. An interactive slider rating scale was used with the slider centred on neutral (50) to minimize potential anchoring effects. Again, belief rating 1 followed initial presentation of the claim; belief rating 2 was given after presentation of the refutation and/or claim-endorsement norm.

Predictive estimates: Predictive estimates were again obtained post-manipulation. Participants indicated their "true effect" estimates on claim-specific scales identical to the scales used for the predictive-estimate norms (e.g., "*If Australia were to slightly increase its refugee uptake now, what do you estimate the effect to be on the Australian economy over the next 20 years, in $-per-resident terms?*" [scale -200 to 1600]). Again, this used a slider anchored at the scale mid-point.

Inference questions: Inference questions were identical to Experiment 2.

Procedure—Procedure was identical to Experiment 2. The only meaningful difference regarded the presentation of the norms: the relevant claim-endorsement norm was presented before the second belief rating (i.e., in the same position as the norm in Experiment 2); the relevant predictive-estimate norm immediately preceded the estimation of the true effect. The experiment lasted approx. 30 minutes; it was run as a stand-alone experiment.

Results

Belief change—Belief-change scores (belief rating 2 – belief rating 1) are shown in Figure 6. Belief change was significantly different from zero in all conditions of Experiment 3A (refutation: $M = -8.93$, $SD = 9.71$, $d = 0.92$; narrow-norm: $M = -8.11$, $SD = 10.54$, $d = 0.77$; wide-norm: $M = -6.55$, $SD = 10.15$, $d = 0.65$), as well as all conditions of Experiment 3B (refutation: $M = -10.00$, $SD = 18.45$, $d = 0.54$; refutation-plus-narrow-norm: $M = -18.66$, $SD = 17.29$, $d = 1.08$; refutation-plus-wide-norm: $M = -15.04$, $SD = 14.36$, $d = 1.05$), all $t(75/77) \leq -4.79$, all $p < .001$, establishing that claim belief was reduced significantly by a refutation or either type of norm.

In Experiment 3A, a repeated measures ANOVA on belief-change scores yielded no significant main effect of condition, $F(2, 150) = 1.16$, $MSE = 95.98$, $p = .317$, $\eta p^2 = .02$, suggesting comparable belief-reducing effects of the refutation and the norms. In Experiment 3B, the ANOVA returned a main effect of condition, $F(2, 154) = 7.19$, $MSE = 205.14$, $p = .001$, $\eta p^2 = .09$. Planned comparisons showed that compared to the refutation condition, there was greater belief change in both the refutation-plus-narrow-norm condition, $F(1, 77) = 12.54$, $MSE = 233.27$, $p < .001$, $\eta p^2 = .14$, and the refutation-plus-wide-norm condition, $F(1, 77) = 4.72$, $MSE = 209.45$, $p = .033$, $\eta p^2 = .06$. The two norm conditions did not differ significantly, $F(1, 77) = 2.97$, $MSE = 172.70$, $p = .089$, $\eta p^2 = .04$. This provides some evidence that a refutation and a social norm can have additive effects.[8]

A between-experiment 2 × 2 within-between ANOVA with the within-subjects factor norm width (narrow, wide) and the between-subjects factor refutation (no refutation [Exp. A], refutation [Exp. B]) returned a significant main effect of refutation, $F(1, 152) = 31.28$, $MSE = 223.07$, $p < .001$, $\eta p^2 = .17$. There was no significant main effect of norm width, $F(1, 152) = 3.73$, $MSE = 138.42$, $p = .055$, $\eta p^2 = .02$,[9] nor an interaction, $F < 1$. This demonstrates that addition of a refutation boosted the belief change associated with a norm.

Predictive estimate scores—Mean predictive estimate scores are shown in Figure 7. In Experiment 3A, a repeated measures ANOVA returned a significant main effect of condition, $F(2, 150) = 3.38$, $MSE = 0.07$, $p = .037$, $\eta p^2 = .04$.[10] Planned contrasts showed lower estimates (i.e., greater belief reduction) in the wide-norm compared to the narrow-norm condition, $F(1, 75) = 7.61$, $MSE = 0.07$, $p = .007$, $\eta p^2 = .09$. Neither norm condition differed from the refutation condition, both $F(1, 75) \leq 2.15$, $p \geq .147$. Overall, this suggested that provision of either a refutation or a norm had comparable impact on predictive estimates, with some benefit of a wide compared to a narrow norm.

In Experiment 3B, the ANOVA returned a significant main effect of condition, $F(2, 154) = 6.06$, $MSE = 0.07$, $p = .003$, $\eta p^2 = .07$. Planned comparisons showed that relative to the refutation condition, estimates were lower (i.e., belief reduction greater) in both the refutation-plus-narrow-norm condition, $F(1, 77) = 10.15$, $MSE = 0.08$, $p = .002$, $\eta p^2 = .12$, and the refutation-plus-wide-norm condition, $F(1, 77) = 4.76$, $MSE = 0.08$, $p = .032$, $\eta p^2 = .06$.[11] The two norm conditions did not differ from each other, $F(1, 77) = 1.71$, $MSE = 0.05$, $p = .194$, $\eta p^2 = .02$. This suggested that provision of a norm enhanced the impact of a refutation.

An across-experiments 2 × 2 within-between ANOVA was conducted on estimate scores with factors norm width (narrow, wide) and refutation (no refutation, refutation). There was no main effect of norm width, $F(1, 152) = 1.45$, $MSE = 0.06$, $p = .230$, $\eta p^2 = .01$, but a marginal main effect of refutation, $F(1, 152) = 4.02$, $MSE = 0.10$, $p = .047$, $\eta p^2 = .03$,[12] and a significant interaction between norm width and refutation, $F(1, 152) = 8.64$, $p = .004$,

[8]The significance levels for the latter two contrasts were reversed in an analysis including covariates, suggesting a benefit for the combination of a refutation only with a narrow norm; see Supplement.
[9]This effect was significant in an analysis with covariates, see Supplement.
[10]This effect was nonsignificant in an analysis with covariates; see Supplement.
[11]This contrast was nonsignificant in an analysis with covariates; see Supplement.
[12]This effect was nonsignificant in an analysis with covariates; see Supplement.

ηp^2 = .05. This indicated that the effect of a narrow norm on estimates was weakened by the absence of a refutation, unlike the effect of a wide norm.

Inference scores—Inference scores are provided in Figure 8. In Experiment 3A, there was a significant main effect of condition, $F(2, 150) = 3.84$, $MSE = 0.02$, $p = .024$, $\eta p^2 = .05$. Planned contrasts revealed significant differences between the narrow-norm condition ($M = 0.62$, $SD = 0.18$) and both the refutation condition ($M = 0.56$, $SD = 0.18$), $F(1, 75) = 5.54$, $MSE = 0.03$, $p = .021$, $\eta p^2 = .07$, and the wide-norm condition ($M = 0.57$, $SD = 0.19$), $F(1, 75) = 5.17$, $MSE = 0.02$, $p = .026$, $\eta p^2 = .06$. There was no significant difference between refutation and wide-norm conditions, $F < 1$. This pattern suggests that a refutation or a wide norm led to lower inference scores and thus less reliance on the disputed claim compared to a narrow norm. In Experiment 3B, there was no significant main effect of condition, $F(2, 154) = 2.39$, $MSE = 0.02$, $p = .095$, $\eta p^2 = .03$.

A 2 (norm width; within-subjects) × 2 (refutation; between-subjects) ANOVA across experiments showed nonsignificant main effects of norm width, $F(1, 152) = 2.44$, $MSE = 0.02$, $p = .121$, $\eta p^2 = .02$,[13] and refutation, $F < 1$, as well as a nonsignificant interaction, $F(1, 152) = 2.64$, $p = .106$, $\eta p^2 = .02$.

Discussion

Experiment 3 was a conceptual replication of Experiment 2, with a stronger norm manipulation. Specifically, Experiment 3 used two norms—one specific to claim endorsement, one specific to the predictive estimate—that were expressed through a graphical representation of 25 values, ostensibly obtained from a representative peer sample. With this operationalization, it was found that belief in a predictive claim was reduced significantly not only by a refutation but also by a (either narrow or wide) norm, and that the combination of both interventions was most impactful. Providing either a refutation or an (additional) norm also had comparable impact on participants' predictive estimates, while there was again some evidence that the combination of norm and refutation achieved the greatest impact. Both in terms of predictive estimates and inference scores, there seemed to be slightly greater influence of wide norms compared to narrow ones.

Thus, Experiment 3 established that descriptive, distributed peer norms, when sufficiently strong and communicated in a salient manner, can significantly reduce belief in potentially false claims. The experiment provided some evidence that the combination of a refutation and a peer norm can achieve the strongest belief reduction, even though it should be noted that the apparent additive nature of refutation and norm effects observed in the direct belief-change scores was much weaker in predictive-estimate and inference scores, and should thus be interpreted and applied with caution. The data regarding norm width suggests that wider norms may be more effective than narrow norms; this may reflect greater ease finding a compromise between one's authentic belief and the norm (Brown et al., 2021), although it could also be that participants simply perceived the wider norms to be more plausible than the narrow ones. We hasten to add that the evidence for an effect of norm

[13]This effect was significant in an analysis with covariates; see Supplement.

width was weak, and that we only compared a very limited set of distribution shapes, which of course vary much more widely in the real world. Future research could use a more diverse set of distribution shapes, to explore whether the size of the norm group or the shape of the distribution is more influential, and also whether norm effects depend on the visual presentation modality.

General Discussion

The present study set out to explore the question of whether descriptive social-norm information may aid in reducing beliefs in potentially false claims, either in isolation or in combination with a refutational intervention. It was hypothesized that social norms should be able to reduce beliefs in equivocal claims, based both on the wider norming literature (e.g., Cialdini & Goldstein, 2004) and the proposition that social factors are influential when it comes to the continued influence that false claims can have after they have been corrected (see Ecker et al., 2022).

Overall, results demonstrated that descriptive peer norms can reduce potentially false attitude-congruent beliefs. Norms can also affect down-stream reasoning (i.e., predictive estimates, responses to inferential reasoning items). This suggests that individuals will express reduced false-claim belief if they perceive the claim to be rejected by their peers (Cohen, 2003; Kahan et al., 2011). In terms of post-correction misinformation reliance more specifically, it suggests that continued influence effects will be smaller if a correction aligns with a perceived social norm. However, results from Experiment 2 also highlighted that such norm influence can be small or even entirely absent if the normative information is perceived to be weak or is poorly communicated (also see Silva & John, 2017). In this context, one limitation of the research is that norm information was fictitious and arbitrary, such that norms may not have been perceived as plausible; moreover, no manipulation checks were implemented to ascertain that norm interventions had the intended effect on perceived consensus.

This study adds to the evidence that refutations that present relevant counterevidence, rather than just stating that a claim is false, cause substantive belief reductions (e.g., Ecker, O' Reilly et al., 2020; Swire, Ecker et al., 2017). One might argue that the 15-20% belief reduction achieved was not sufficient to eliminate the influence of the initially presented equivocal information entirely—in other words, there may have been continued influence of the initial information—however, due to the absence of a baseline from a control group not exposed to the initial claim, no strong conclusions regarding continued influence can be drawn.

The implications of this research are straightforward. At a theoretical level, results suggest that models of post-correction misinformation reliance should take social context factors into account, similar to theories of persuasion and belief change more generally (e.g., Burns & Gomoliñska, 2001; Hornsey & Fielding, 2017; Jaccard, 1981; Kahan et al., 2011;

Sherif & Hovland, 1961). For example, researchers have begun to incorporate factors such as the threat that corrections pose to a person's identity into their theories of belief updating

and continued influence (Trevors, 2021). A particularly promising approach has been the development of Bayesian network models that explicitly account for the impact of trust and perceived source reliability (Connor Desai et al., 2020; Cook & Lewandowsky, 2016).

At the practical level, our study suggests that social norming interventions should be explored more systematically, especially in the field. In this regard, it was encouraging that wider norms seemed to be somewhat more influential, as real-world norm distributions will not tend to be as narrow as the narrow distribution implemented in Experiment 3. Replication with realistic norms would also be important given the norms in the current study were fictional and arbitrary. In general, exploration of social norming interventions is important because false beliefs that deviate from a social consensus are not only potential drivers of misinformed and thus potentially harmful behaviours (e.g., Bauman & Geher, 2002; Botvin, Botvin, Baker, Dusenbury, & Goldberg, 1992), but they can also affect public discourse. Specifically, overestimation of the social acceptance of one's view (i.e., a false-consensus effect; Marks & Miller, 1987) can be detrimental if the view is ill-informed: For example, a person assuming that many others share the misconception that anthropogenic climate change is a hoax will be motivated to promote their false view and actively oppose mitigative action (e.g., Santos, Levin, & Vasconcelos, 2021). On the flipside, if a vocal minority keeps a false belief in the public spotlight, this can lead people who hold a factual majority belief to underestimate the social acceptance of their view and falsely assume they are in the minority (i.e., a false-uniqueness effect; Bosveld, Koomen, van der Pligt, Plaisier, 1995). The resulting pluralistic ignorance (see Lewandowsky, Facer, & Ecker, 2021; Prentice & Miller, 1993; Sargent & Newman, 2021; Shamir & Shamir, 1997) can lead to hesitation to express concerns about an issue if people incorrectly assume the concern is not shared by others (e.g., Geiger & Swim, 2016; Santos et al., 2021), and can ultimately affect policy decisions in a manner that is fundamentally at odds with democratic ideals (Todorov & Mandisodza, 2004). People's attitudes also often move towards the perceived majority opinion (e.g., Prentice & Miller, 1993), meaning that others' views can drive attitude adjustment at the expense of people's actual preferences, and thus unduly shape public opinion. It is therefore advisable to harness peer consensus information where it is available in order to avoid such biases.

The present results provide support for the continued investigation of social norms in the context of misinformation processing. Future research may consider testing the impact of injunctive norms (i.e., norms that refer to the social acceptability of a belief or behaviour; Cialdini et al., 1990), as previous research has shown that presenting aligned descriptive and injunctive norms can result in larger attitudinal and behavioural changes than presenting either type of norm in isolation (Cialdini, 2003; Hamann, Reese, Seewald, & Loeschinger, 2015; Smith & Louis, 2008). Future work could also use claims that are not predictive in nature, but are instead selected to be objectively false. The speculative nature of predictive claims may be associated with less-firmly held views, and thus participants may not have been overly motivated to maintain their beliefs and defend them against change. Such research would be especially valuable if it made use of formal models of decision making to derive and then test quantitative predictions in a principled manner. Future work could also explore the role of pre-existing attitudes and perceived subject-matter knowledge, which could influence the extent to which people are motivated to defend potentially entrenched

misconceptions (e.g., see Motta, Callaghan, & Sylvester, 2018). In this vein, it may be useful to replicate this work using non-student samples and a larger and more diverse set of claims, for better generalizability. Finally, future studies could investigate if dynamic norms (e.g., "more and more people dispute this claim") may be especially powerful due to bandwagon effects.

In sum, while much remains to be explored, this study has presented a proof-of-concept that social factors can affect revision of potentially false beliefs, and that social norm information may be a useful tool in the arsenal of the debunking practitioner.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was facilitated by the Australian Research Council (grants DP160103596 to UKHE and SL, FT190100708 to UKHE) and the European Research Council under the European Union's Horizon 2020 research and innovation programme (grants 788826 RRTJDM to GDAB, 101020961 PRODEMINFO to SL). SL also acknowledges support from the Alexander von Humboldt Foundation, the Volkswagen Foundation (large grant "Reclaiming individual autonomy and democratic discourse online"), the Economic and Social Research Council through a Knowledge Exchange Fellowship, and the European Commission (Horizon 2020 grant 964728 JITSUVAX). We thank Charles Hanich for research assistance.

Data availability statement

Data are available at https://osf.io/ekxzy/.

References

Abrams D, Wetherell M, Cochrane S, Hogg MA, Turner JC. Knowing what to think by knowing who you are: Self-categorization and the nature of norm formation, conformity and group polarization. British Journal of Social Psychology. 1990; 29 (2) 97–119. DOI: 10.1111/j.2044-8309.1990.tb00892.x [PubMed: 2372667]

Albarracín D, Kumkale GT, Poyner-Del Vento P. How people can become persuaded by weak messages presented by credible communicators: Not all sleeper effects are created equal. Journal of Experimental Social Psychology. 2017; 68: 171–180. DOI: 10.1016/j.jesp.2016.06.009 [PubMed: 34054141]

Aldrovandi S, Brown GDA, Wood AM. Social norms and rank-based nudging: Changing willingness to pay for healthy food. Journal of Experimental Psychology: Applied. 2015; 21 (3) 242–254. DOI: 10.1037/xap0000048 [PubMed: 26010301]

Aldrovandi S, Wood AM, Maltby J, Brown GDA. Students' concern about indebtedness: a rank based social norms account. Studies in Higher Education. 2015; 40 (7) 1307–1327. DOI: 10.1080/03075079.2014.881349

Amin AB, Bednarczyk RA, Ray CE, Melchiori KJ, Graham J, Huntsinger JR, Omer SB. Association of moral values with vaccine hesitancy. Nature Human Behaviour. 2017; 1: 873–880. DOI: 10.1038/s41562-017-0256-5

Ancker J, Senathirajah Y, Kukafka R, Starren J. Design features of graphs in health risk communication: a systematic review. Journal of the American Medical Informatics Association. 2006; 13 (6) 608–618. DOI: 10.1197/jamia.M2115 [PubMed: 16929039]

Andı S, Akesson J. Nudging away false news: Evidence from a social norms experiment. Digital Journalism. 2021; 9: 106–125. DOI: 10.1080/21670811.2020.1847674

Europe PMC Funders Author Manuscripts

Angner E. Economists as experts: Overconfidence in theory and practice. Journal of Economic Methodology. 2006; 13 (1) 1–24. DOI: 10.1080/13501780600566271

Aronson E, Turner JA, Carlsmith JM. Communicator credibility and communication discrepancy as determinants of opinion change. The Journal of Abnormal and Social Psychology. 1963; 67 (1) 31–36. DOI: 10.1037/h0045513

Asch SE. Studies of independence and conformity: A minority of one against a unanimous majority. Psychological Monographs: General and Applied. 1956; 70: 1–70. DOI: 10.1037/h0093718

Bauman KP, Geher G. We think you agree: The detrimental impact of the false consensus effect on behavior. Current Psychology. 2002; 21 (4) 293–318. DOI: 10.1007/s12144-002-1020-0

Berinsky AJ. Rumors and health care reform: Experiments in political misinformation. British Journal of Political Science. 2017; 47: 241–262. DOI: 10.1017/s0007123415000186

Bode L, Vraga EK. See something, say something: Correction of global health misinformation on social media. Health Communication. 2018; 33 (9) 1131–1140. DOI: 10.1080/10410236.2017.1331312 [PubMed: 28622038]

Bolsen T, Leeper TJ, Shapiro MA. Doing what others do: Norms, science, and collective action on global warming. American Politics Research. 2014; 42 (1) 65–89. DOI: 10.1177/1532673X13484173

Bosveld W, Koomen W, van der Pligt J, Plaisier JW. Differential construal as an explanation for false consensus and false uniqueness effects. Journal of Experimental Social Psychology. 1995; 31 (6) 518–532. DOI: 10.1006/jesp.1995.1023

Botvin GJ, Botvin EM, Baker E, Dusenbury L, Goldberg CJ. The false consensus effect: Predicting adolescents' tobacco use from normative expectations. Psychological Reports. 1992; 70 (1) 171–178. DOI: 10.2466/PR0.70.1.171-178 [PubMed: 1565717]

Brown GDA, Lewandowsky S, Huang Z. Social sampling and expressed attitudes: Authenticity preference and social extremeness aversion lead to social norm effects and polarization. Psychological Review. 2021.

Burns TR, Gomoliñska A. Socio-cognitive mechanisms of belief change: Applications of generalized game theory to belief revision, social fabrication, and self-fulfilling prophesy. Cognitive Systems Research. 2001; 2 (1) 39–54. DOI: 10.1016/S1389-0417(01)00014-6

Chan M-PS, Jones CR, Hall Jamieson K, Albarracín D. Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation. Psychological Science. 2017; 28: 1531–1546. DOI: 10.1177/0956797617714579 [PubMed: 28895452]

Cialdini RB. Crafting normative messages to protect the environment. Current Directions in Psychological Science. 2003; 12 (4) 105–109. DOI: 10.1111/1467-8721.01242

Cialdini RB, Goldstein NJ. Social influence: Compliance and conformity. Annual Review of Psychology. 2004; 55: 591–621. DOI: 10.1146/annurev.psych.55.090902.142015

Cialdini RB, Reno RR, Kallgren CA. A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. Journal of Personality and Social Psychology. 1990; 58 (6) 1015–1026. DOI: 10.1037/0022-3514.58.6.1015

Cialdini, RB, Trost, MR. The Handbook of Social Psychology. 4th ed. Vol. 1-2. McGraw-Hill; New York, NY, US: 1998. 151–192.

Cohen GL. Party over policy: The dominating impact of group influence on political beliefs. Journal of Personality and Social Psychology. 2003; 85 (5) 808–22. DOI: 10.1037/0022-3514.85.5.808 [PubMed: 14599246]

Cohen GL, Prinstein MJ. Peer contagion of aggression and health risk behavior among adolescent males: An experimental investigation of effects on public conduct and private attitudes. Child Development. 2006; 77: 967–983. DOI: 10.1111/j.1467-8624.2006.00913.x [PubMed: 16942500]

Connor Desai SA, Pilditch TD, Madsen JK. The rational continued influence of misinformation. Cognition. 2020; 205 104453 doi: 10.1016/j.cognition.2020.104453 [PubMed: 33011527]

Cook J, Lewandowsky S. Rational irrationality: Modeling climate change belief polarization using Bayesian networks. Topics in Cognitive Science. 2016; 8: 160–179. DOI: 10.1111/tops.12186 [PubMed: 26749179]

Darke PR, Chaiken S, Bohner G, Einwiller S, Erb HP, Hazlewood JD. Accuracy motivation, consensus information, and the law of large numbers: Effects on attitude judgement in the absence of

argumentation. Personality and Social Psychology Bulletin. 1998; 24 (11) 1205–1215. DOI: 10.1177/01461672982411007

Deutsch M, Gerard HB. A study of normative and informational social influences upon individual judgment. Journal of Abnormal and Social Psychology. 1955; 51 (3) 629–636. DOI: 10.1037/h0046408

Dias N, Pennycook G, Rand DG. Emphasizing publishers does not effectively reduce susceptibility to misinformation on social media. Harvard Kennedy School Misinformation Review. 2020; doi: 10.37016/mr-2020-001

Eberhard K. The effects of visualization on judgment and decision-making: A systematic literature review. Management Review Quarterly. 2021; doi: 10.1007/s11301-021-00235-8

Ecker UKH, Ang LC. Political attitudes and the processing of misinformation corrections. Political Psychology. 2019; 40: 241–260. DOI: 10.1111/pops.12494

Ecker UKH, Antonio LM. Can you believe it? An investigation into the impact of retraction source credibility on the continued influence effect. Memory & Cognition. 2021; 49: 631–644. DOI: 10.3758/s13421-020-01129-y [PubMed: 33452666]

Ecker UKH, Hogan JL, Lewandowsky S. Reminders and repetition of misinformation: Helping or hindering its retraction? Journal of Applied Research in Memory and Cognition. 2017; 6: 185–192. DOI: 10.1016/j.jarmac.2017.01.014

Ecker UKH, Lewandowsky S, Apai J. Terrorists brought down the plane!— No, actually it was a technical fault: Processing corrections of emotive information. Quarterly Journal of Experimental Psychology. 2011; 64: 283–310. DOI: 10.1080/17470218.2010.497927

Ecker UKH, Lewandowsky S, Chadwick M. Can corrections spread misinformation to new audiences? Testing for the elusive familiarity backfire effect. Cognitive Research: Principles and Implications. 2020; 5: 41. doi: 10.1186/s41235-020-00241-6 [PubMed: 32844338]

Ecker UKH, Lewandowsky S, Cook J, Schmid P, Fazio LK, Brashier N, Kendeou P, Vraga EK, Amazeen MA. The psychological drivers of misinformation belief and its resistance to correction. Nature Reviews Psychology. 2022.

Ecker UKH, Lewandowsky S, Jayawardana K, Mladenovic A. Refutations of equivocal claims: No evidence for an ironic effect of counterargument number. Journal of Applied Research in Memory and Cognition. 2019; 8: 98–107. DOI: 10.1016/j.jarmac.2018.07.005

Ecker UKH, O'Reilly Z, Reid JS, Chang EP. The effectiveness of shortformat refutational fact-checks. British Journal of Psychology. 2020; 111: 36–54. DOI: 10.1111/bjop.12383 [PubMed: 30825195]

Ecker UKH, Sze BKN, Andreotta M. Corrections of political misinformation: No evidence for an effect of partisan worldview in a US convenience sample. Philosophical Transactions of the Royal Society B. 2021; 376 (1822) 20200145 doi: 10.1098/rstb.2020.0145

Edwards K, Smith EE. A disconfirmation bias in the evaluation of arguments. Journal of Personality and Social Psychology. 1996; 71: 5–24. DOI: 10.1037/0022-3514.71.1.5

Ferraro PJ, Price MK. Using nonpecuniary strategies to influence behavior: Evidence from a large-scale field experiment. Review of Economics and Statistics. 2013; 95 (1) 64–73. DOI: 10.1162/REST_a_00344

Festinger, L. A theory of cognitive dissonance. Stanford University Press; 1957.

Galassi JP, DeLo JS, Galassi MD, Bastien S. The college self-expression scale: A measure of assertiveness. Behavior Therapy. 1974; 5 (2) 165–171. DOI: 10.1016/S0005-7894(74)80131-0

Geiger N, Swim JK. Climate of silence: Pluralistic ignorance as a barrier to climate change discussion. Journal of Environmental Psychology. 2016; 47: 79–90. DOI: 10.1016/j.jenvp.2016.05.002

Gelman A, Pasarica C, Dodhia R. Let's practice what we preach: Turning tables into graphs. American Statistician. 2002; 56 (2) 121–130. DOI: 10.1198/000313002317572790

Gimpel H, Heger S, Olenberger C, Utz L. The effectiveness of social norms in fighting fake news on social media. Journal of Management Information Systems. 2021; 38 (1) 196–221. DOI: 10.1080/07421222.2021.1870389

Goldberg MH, van der Linden S, Leiserowitz A, Maibach E. Perceived social consensus can reduce ideological biases on climate change. Environment and Behavior. 2020; 52 (5) 495–517. DOI: 10.1177/0013916519853302

Gordon A, Quadflieg S, Brooks JCW, Ecker UKH, Lewandowsky S. Keeping track of 'alternative facts': The neural correlates of processing misinformation corrections. NeuroImage. 2019; 193: 46–56. DOI: 10.1016/j.neuroimage.2019.03.014 [PubMed: 30872047]

Graves L, Amazeen M. Fact-checking as idea and practice in journalism. Oxford Research Encyclopedia of Communication. 2019; doi: 10.1093/acrefore/9780190228613.013.808

Guillory JJ, Geraci L. Correcting erroneous inferences in memory: The role of source credibility. Journal of Applied Research in Memory and Cognition. 2013; 2: 201–209. DOI: 10.1016/j.jarmac.2013.10.001

Hamann KRS, Reese G, Seewald D, Loeschinger DC. Affixing the theory of normative conduct (to your mailbox): Injunctive and descriptive norms as predictors of anti-ads sticker use. Journal of Environmental Psychology. 2015; 44: 1–9. DOI: 10.1016/j.jenvp.2015.08.003

Hameleers M, van der Meer TGLA. Misinformation and polarization in a high-choice media environment: How effective are political fact-checkers? Communication Research. 2020; 47 (2) 227–250. DOI: 10.1177/0093650218819671

Hannak, A; Margolin, DB; Keegan, B; Weber, I. Get back! You don't know me like that: The social mediation of fact checking interventions in Twitter conversations; Proceedings of the International AAAI Conference on Web and Social Media; 2014. 187–196.

Hornsey MJ, Fielding KS. Attitude roots and jiu jitsu persuasion: Understanding and overcoming the motivated rejection of science. American Psychologist. 2017; 72 (5) 459–473. DOI: 10.1037/a0040437 [PubMed: 28726454]

Hornsey MJ, Jetten J. The individual within the group: Balancing the need to belong with the need to be different. Personality and Social Psychology Review. 2004; 8 (3) 248–264. DOI: 10.1207/s15327957pspr0803_2 [PubMed: 15454348]

Jaccard J. Toward theories of persuasion and belief change. Journal of Personality and Social Psychology. 1981; 40 (2) 260–269. DOI: 10.1037/0022-3514.40.2.260

Jaeger A, Lauris P, Selmeczy D, Dobbins IG. The costs and benefits of memory conformity. Memory & Cognition. 2012; 40: 101–112. DOI: 10.3758/s13421-011-0130-z [PubMed: 21773846]

Jellison JM, Mills J. Effect of public commitment upon opinions. Journal of Experimental Social Psychology. 1969; 5 (3) 340–346. DOI: 10.1016/0022-1031(69)90058-4

Kahan D. Ideology, motivated reasoning, and cognitive reflection. Judgment and Decision Making. 2013; 8: 407–424. DOI: 10.2139/ssrn.2182588

Kahan DM, Jenkins-Smith H, Braman D. Cultural cognition of scientific consensus. Journal of Risk Research. 2011; 14 (2) 147–174. DOI: 10.1080/13669877.2010.511246

Kaplan MF, Miller CE. Group decision making and normative versus informational influence: Effects of type of issue and assigned decision rule. Journal of Personality and Social Psychology. 1987; 53 (2) 306–313. DOI: 10.1037/0022-3514.53.2.306

Kendeou P, Butterfuss R, Kim J, van Boekel M. Knowledge revision through the lenses of the three-pronged approach. Memory & Cognition. 2019; 47: 33–46. DOI: 10.3758/s13421-018-0848-y [PubMed: 30117115]

Kendeou P, Walsh EK, Smith ER, O'Brien EJ. Knowledge revision processes in refutation texts. Discourse Processes. 2014; 51: 374–397. DOI: 10.1080/0163853x.2014.913961

Koop GJ, King A, Kauffman KJ. Infrequent but adaptive outsourcing in recognition memory. Journal of Memory and Language. 2021; 118 104216 doi: 10.1016/j.jml.2020.104216

Kunda Z. The case for motivated reasoning. Psychological Bulletin. 1990; 108: 480–498. DOI: 10.1037/0033-2909.108.3.480 [PubMed: 2270237]

Lapinski MK, Rimal RN. An explication of social norms. Communication Theory. 2005; 15: 127–147. DOI: 10.1111/j.1468-2885.2005.tb00329.x

Lewandowsky S, Cook J, Fay N, Gignac G. Science by social media: Attitudes towards climate change are mediated by perceived social consensus. Memory & Cognition. 2019; 47: 1445–1456. DOI: 10.3758/s13421-019-00948-y [PubMed: 31228014]

Lewandowsky S, Ecker UKH, Seifert CM, Schwarz N, Cook J. Misinformation and its correction: Continued influence and successful debiasing. Psychological Science in the Public Interest. 2012; 13: 106–131. DOI: 10.1177/1529100612451018 [PubMed: 26173286]

Europe PMC Funders Author Manuscripts

Lewandowsky S, Facer K, Ecker UKH. Losses, hopes, and expectations for sustainable futures after COVID. Humanities & Social Sciences Communications. 2021; 8: 296. doi: 10.1057/s41599-021-00961-0

Lewandowsky S, Gignac G, Vaughan S. The pivotal role of perceived scientific consensus in acceptance of science. Nature Climate Change. 2013; 3: 399–404. DOI: 10.1038/nclimate1720

MacDonald G, Nail PR. Attitude change and the public–private attitude distinction. British Journal of Social Psychology. 2005; 44: 15–28. DOI: 10.1348/014466604X23437 [PubMed: 15901389]

Mannes AE. Are we wise about the wisdom of crowds? The use of group judgements in belief revision. Management Science. 2009; 55 (8) 1267–1279. DOI: 10.1287/mnsc.1090.1031

Margolin DB, Hannak A, Weber I. Political fact-checking on Twitter: When do corrections have an effect? Political Communication. 2018; 35 (2) 196–219. DOI: 10.1080/10584609.2017.1334018

Marks G, Miller N. Ten years of research on the false-consensus effect: An empirical and theoretical review. Psychological Bulletin. 1987; 102 (1) 72–90. DOI: 10.1037/0033-2909.102.1.72

Meyer J, Shamo MK, Gopher D. Information structure and the relative efficacy of tables and graphs. Human Factors. 1999; 41 (4) 570–587. DOI: 10.1518/001872099779656707 [PubMed: 10774128]

Molleman L, Tump AN, Gradassi A, Herzog S, Jayles B, Kurvers RHJM, van den Bos W. Strategies for integrating disparate social information. Proceedings of the Royal Society B: Biological Sciences. 2020; 287 (1939) 20202413 doi: 10.1098/rspb.2020.2413

Moore SC, Wood AM, Moore L, Shepherd J, Murphy S, Brown GDA. A rank based social norms model of how people judge their levels of drunkenness whilst intoxicated. BMC Public Health. 2016; 16 (1) 798. doi: 10.1186/s12889-016-3469-z [PubMed: 27619969]

Moscovici S. Toward a theory of conversion behavior. Advances in Experimental Social Psychology. 1980; 13: 209–239. DOI: 10.1016/S0065-2601(08)60133-1

Mosleh, M; Martel, C; Eckles, D; Rand, DG. Perverse downstream consequences of debunking: Being corrected by another user for posting false political news increases subsequent sharing of low quality, partisan, and toxic content in a Twitter field experiment; Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems; 2021. 1–13.

Motta M, Callaghan T, Sylvester S. Knowing less but presuming more: Dunning-Kruger effects and the endorsement of anti-vaccine policy attitudes. Social Science & Medicine. 2018; 211: 274–281. DOI: 10.1016/j.socscimed.2018.06.032 [PubMed: 29966822]

Moussaïd M, Kämmer JE, Analytis PP, Neth H. Social influence and the collective dynamics of opinion formation. PLOS ONE. 2013; 8 (11) 78433 doi: 10.1371/journal.pone.0078433

Nolan JM, Schultz PW, Cialdini RB, Goldstein NJ, Griskevicius V. Normative social influence is underdetected. Personality and Social Psychology Bulletin. 2008; 34: 913–923. DOI: 10.1177/0146167208316691 [PubMed: 18550863]

Nyhan B, Reifler J. When corrections fail: The persistence of political misperceptions. Political Behavior. 2010; 32: 303–330. DOI: 10.1007/s11109-010-9112-2

O'Rear AE, Radvansky GA. Failure to accept retractions: A contribution to the continued influence effect. Memory & Cognition. 2020; 48: 127–144. DOI: 10.3758/s13421-019-00967-9 [PubMed: 31317393]

Parducci A. Category judgment: A range-frequency model. Psychological Review. 1965; 72 (6) 407–418. DOI: 10.1037/h0022602 [PubMed: 5852241]

Paynter J, Luskin-Saxby S, Keen D, Fordyce K, Frost G, Imms C, Miller S, Trembath D, Tucker M, Ecker UKH. Evaluation of a template for countering misinformation—Real-world Autism treatment myth debunking. PLOS ONE. 2019; 14 (1) e0210746 doi: 10.1371/journal.pone.0210746 [PubMed: 30699155]

Pennycook G, Rand DG. Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. Journal of Personality. 2020; 88: 185–200. DOI: 10.1111/jopy.12476 [PubMed: 30929263]

Pornpitakpan C. The persuasiveness of source credibility: A critical review of five decades evidence. Journal of Applied Social Psychology. 2004; 34: 243–281. DOI: 10.1111/j.1559-1816.2004.tb02547.x

Prasad M, Perrin AJ, Bezila K, Hoffman SG, Kindleberger K, Manturuk K, Powers AS. “There must be a reason”: Osama, Saddam, and inferred justification. Sociological Inquiry. 2009; 79: 142–162. DOI: 10.1111/j.1475-682x.2009.00280.x

Prentice DA, Miller DT. Pluralistic ignorance and alcohol use on campus: Some consequences of misperceiving the social norm. Journal of Personality and Social Psychology. 1993; 64: 243–256. DOI: 10.1037/0022-3514.64.2.243 [PubMed: 8433272]

Puhl RM, Schwartz MB, Brownell KD. Impact of perceived consensus on stereotypes about obese people: A new approach for reducing bias. Health Psychology. 2005; 24 (5) 517–525. DOI: 10.1037/0278-6133.24.5.517 [PubMed: 16162046]

Quekett, H. Does social context affect belief change? – Private vs public expressions of belief and the continued influence of misinformation. University of Western Australia; 2016. Unpublished thesis

R Core Team. R: A language and environment for statistical computing. 2013. http://www.R-project.org/

Santee RT, Maslach C. To agree or not to agree: Personal dissent amid social pressure to conform. Journal of Personality and Social Psychology. 1982; 42 (4) 690–700. DOI: 10.1037/0022-3514.42.4.690

Santos FP, Levin SA, Vasconcelos VV. Biased perceptions explain collective action deadlocks and suggest new mechanisms to prompt cooperation. iScience. 2021; 24 (4) 102375 doi: 10.1016/j.isci.2021.102375 [PubMed: 33948558]

Sargent RH, Newman LS. Pluralistic ignorance research in psychology: A scoping review of topic and method variation and directions for future research. Review of General Psychology. 2021; 25 (2) 163–184. DOI: 10.1177/1089268021995168

Schmiege SJ, Klein WM, Bryan AD. The effect of peer comparison information in the context of expert recommendations on risk perceptions and subsequent behavior. European Journal of Social Psychology. 2010; 40: 746–759. DOI: 10.1002/ejsp.645

Schultz PW, Nolan JM, Cialdini RB, Goldstein NJ, Griskevicius V. The constructive, destructive, and reconstructive power of social norms. Psychological Science. 2007; 18 (5) 429–434. DOI: 10.1111/j.1467-9280.2007.01917.x [PubMed: 17576283]

Sechrist GB, Milford LR. The influence of social consensus information on intergroup helping behavior. Basic and Applied Social Psychology. 2007; 29 (4) 365–374. DOI: 10.1080/01973530701665199

Seifert CM. The continued influence of misinformation in memory: What makes a correction effective? Psychology of Learning and Motivation. 2002; 41: 265–292. DOI: 10.1016/S0079-7421(02)80009-3

Shamir J, Shamir M. Pluralistic ignorance across issues and over time: Information cues and biases. Public Opinion Quarterly. 1997; 61 (2) 227–260. DOI: 10.1086/297794

Sherif, M, Hovland, CI. Social judgment: Assimilation and contrast effects in communication and attitude change. Yale University Press; 1961.

Silva A, John P. Social norms don’t always work: An experiment to encourage more efficient fees collection for students. PLOS ONE. 2017; 12 (5) e0177354 doi: 10.1371/journal.pone.0177354 [PubMed: 28542164]

Sleegers WWA, Proulx T, van Beest I. Confirmation bias and misconceptions: Pupillometric evidence for a confirmation bias in misconceptions feedback. Biological Psychology. 2019; 145: 76–83. DOI: 10.1016/j.biopsycho.2019.03.018 [PubMed: 30965093]

Smith JR, Louis WR. Do as we say and as we do: the interplay of descriptive and injunctive group norms in the attitude-behaviour relationship. British Journal of Social Psychology. 2008; 47 (4) 647–666. DOI: 10.1348/014466607X269748 [PubMed: 18163950]

Sparks JR, Rapp DN. Readers’ reliance on source credibility in the service of comprehension. Journal of Experimental Psychology: Learning Memory and Cognition. 2011; 37: 230–247. DOI: 10.1037/a0021331 [PubMed: 21244116]

Stangor C, Sechrist GB, Jost JT. Changing racial beliefs by providing consensus information. Personality and Social Psychology Bulletin. 2001; 27: 486–496. DOI: 10.1177/0146167201274009

Stewart N, Chater N, Brown GDA. Decision by sampling. Cognitive Psychology. 2006; 53 (1) 1–26. DOI: 10.1016/j.cogpsych.2005.10.003 [PubMed: 16438947]

Swire B, Berinsky AJ, Lewandowsky S, Ecker UKH. Processing political misinformation—Comprehending the Trump phenomenon. Royal Society Open Science. 2017; 4 160802 doi: 10.1098/rsos.160802 [PubMed: 28405366]

Swire B, Ecker UKH, Lewandowsky S. The role of familiarity in correcting inaccurate information. Journal of Experimental Psychology: Learning, Memory, & Cognition. 2017; 43: 1948–1961. DOI: 10.1037/xlm0000422 [PubMed: 28504531]

Swire-Thompson B, De Gutis J, Lazer D. Searching for the backfire effect: Measurement and design considerations. Journal of Applied Research in Memory and Cognition. 2020; 9 (3) 286–299. DOI: 10.1016/j.jarmac.2020.06.006 [PubMed: 32905023]

Swire-Thompson B, Ecker UKH, Lewandowsky S, Berinsky A. They might be a liar but they're my liar: Source evaluation and the prevalence of misinformation. Political Psychology. 2020; 41: 21–34. DOI: 10.1111/pops.12586

Swire-Thompson B, Miklaucic N, Wihbey J, Lazer D, DeGutis J. Backfire effects after correcting misinformation are strongly associated with reliability. Journal of Experimental Psychology: General. 2021; doi: 10.31234/osf.io/e3pvx

Tankard ME, Paluck EL. Norm perception as a vehicle for social change. Social Issues and Policy Review. 2016; 10: 181–211. DOI: 10.1111/sipr.12022

Terry DJ, Hogg MA. Group norms and the attitude–behavior relationship: A role for group identification. Personality and Social Psychology Bulletin. 1996; 22 (8) 776–793. DOI: 10.1177/0146167296228002

Terry DJ, Hogg MA, White KM. The theory of planned behaviour: Selfidentity, social identity and group norms. British Journal of Social Psychology. 1999; 38 (3) 225–244. DOI: 10.1348/014466699164149 [PubMed: 10520477]

Thombs D, Dotterer S, Olds R, Sharp K, Raub C. A close look at why one social norms campaign did not reduce student drinking. Journal of American College Health. 2004; 53 (2) 61–68. DOI: 10.3200/JACH.53.2.61-70 [PubMed: 15495882]

Todorov A, Mandisodza AN. Public opinion on foreign policy: The multilateral public that perceives itself as unilateral. Public Opinion Quarterly. 2004; 68 (3) 323–348. DOI: 10.1093/poq/nfh036

Trevors GJ. The roles of identity conflict, emotion, and threat in learning from refutation texts on vaccination and immigration. Discourse Processes. 2021; doi: 10.1080/0163853X.2021.1917950

van der Linden SL, Leiserowitz AA, Feinberg GD, Maibach EW. The scientific consensus on climate change as a gateway belief: Experimental evidence. PLOS ONE. 2015; 10 e0118489 doi: 10.1371/journal.pone.0118489 [PubMed: 25714347]

van der Linden S, Leiserowitz A, Maibach E. The gateway belief model: A large-scale replication. Journal of Environmental Psychology. 2019; 62: 49–58. DOI: 10.1016/j.jenvp.2019.01.009

van der Meer TGLA, Jin Y. Seeking formula for misinformation treatment in public health crises: The effects of corrective information type and source. Health Communication. 2020; 35 (5) 560–575. DOI: 10.1080/10410236.2019.1573295 [PubMed: 30761917]

Vlasceanu M, Coman A. The impact of social norms on belief update. Applied Psychology: Health and Well-Being. 2021; doi: 10.1111/aphw.12313

Vlasceanu M, Morais MJ, Duker A, Coman A. The synchronization of collective beliefs: From dyadic interactions to network convergence. Journal of Experimental Psychology: Applied. 2020; 26: 453–464. DOI: 10.1037/xap0000265 [PubMed: 31999143]

Vraga EK, Bode L. Using expert sources to correct health misinformation in social media. Science Communication. 2017; 39: 621–645. DOI: 10.1177/1075547017731776

Walter N, Tukachinsky R. A meta-analytic examination of the continued influence of misinformation in the face of correction: How powerful is it, why does it happen, and how to stop it? Communication Research. 2020; 47: 155–177. DOI: 10.1177/0093650219854600

Weeks BE. Emotions, partisanship, and misperceptions: How anger and anxiety moderate the effect of partisan bias on susceptibility to political misinformation. Journal of Communication. 2015; 65: 699–719. DOI: 10.1111/jcom.12164

Williams JM, Warchal J. The relationship between assertiveness, internalexternal locus of control, and overt conformity. Journal of Psychology. 1981; 109 (1) 93–96. DOI: 10.1080/00223980.1981.9915291

Wood W. Attitude change: Persuasion and social influence. Annual Review of Psychology. 2000; 51 (1) 539–570. DOI: 10.1146/annurev.psych.51.1.539

Wood AM, Linley PA, Maltby J, Baliousis M, Joseph S. The authentic personality: A theoretical and empirical conceptualization and the development of the Authenticity Scale. Journal of Counseling Psychology. 2008; 55 (3) 385–399. DOI: 10.1037/0022-0167.55.3.385

Wood T, Porter E. The elusive backfire effect: Mass attitudes' steadfast factual adherence. Political Behavior. 2019; 41: 135–163. DOI: 10.1007/s11109-018-9443-y

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

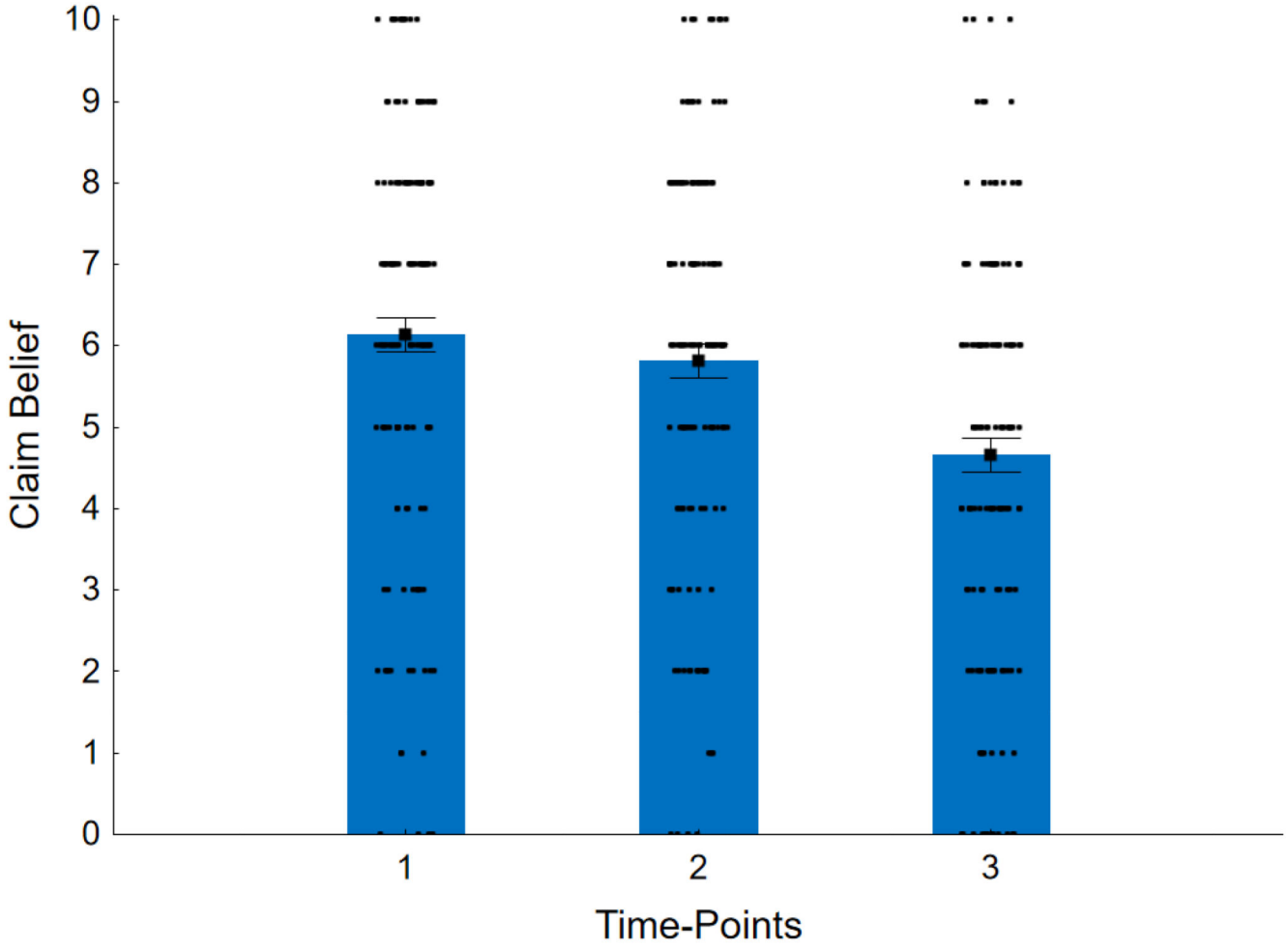

Figure 1. Mean Claim-Belief Ratings Across Time-Points in Experiment 1

Note. Belief change from time-point 1 to time-point 2 reflects the impact of a descriptive pre-refutation norm (collapsed across confidentiality conditions). Belief change from time-point 2 to time-point 3 reflects the impact of a refutation (collapsed across post-refutation norm and confidentiality conditions). Error bars show standard error of the mean. Dots show individual raw data points (jittered).

Europe PMC Funders Author Manuscripts

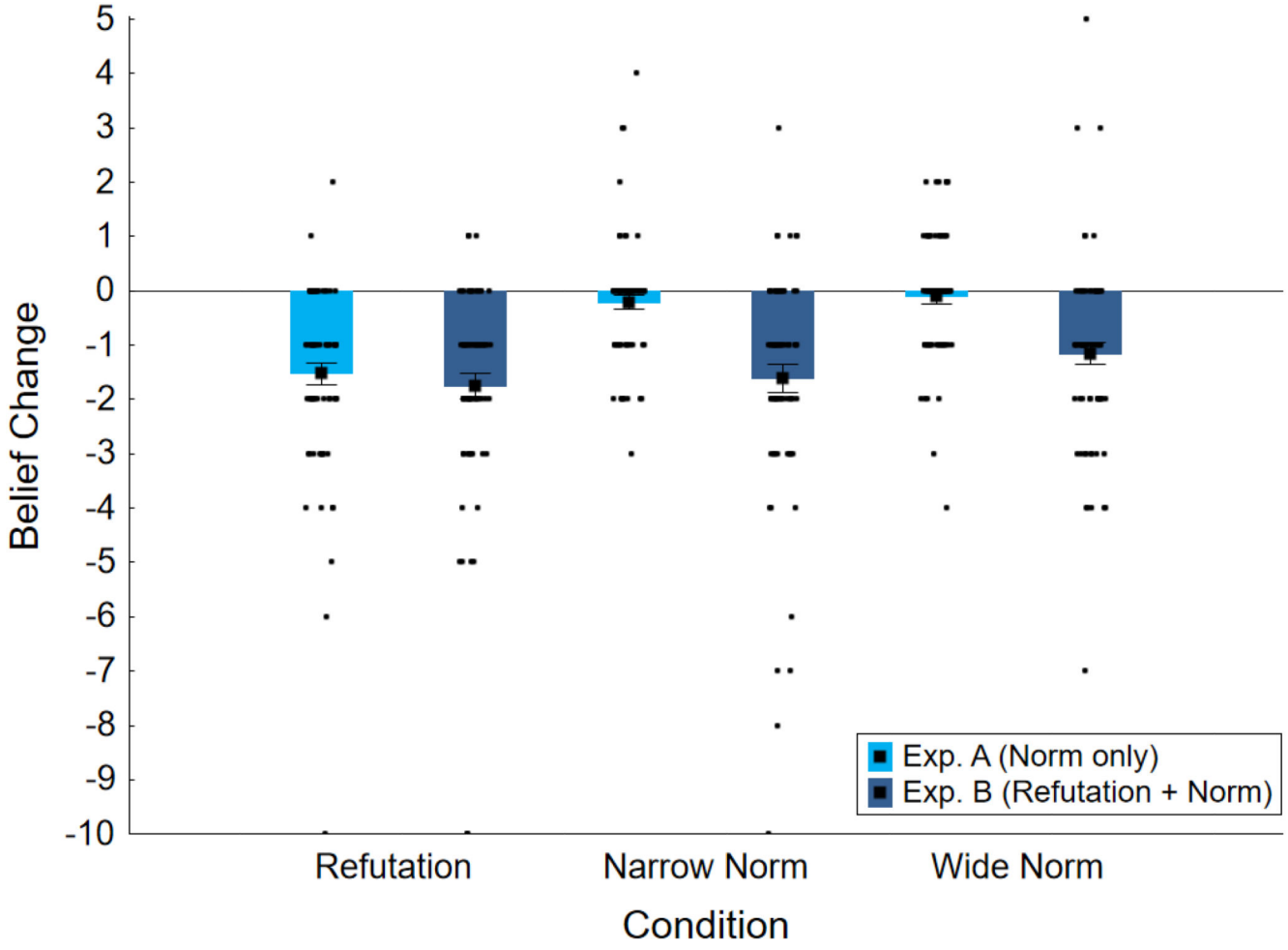

Figure 2. Mean Belief-Change Scores Across Conditions in Experiment 2

Note. Claim beliefs were measured on 0-10 scales. Norm information was only provided in narrow and wide norm conditions, either by itself (Experiment 2A) or together with the refutation (Experiment 2B). The two refutation conditions were identical. Error bars show standard error of the mean. Dots show individual raw data points (jittered).

Europe PMC Funders Author Manuscripts

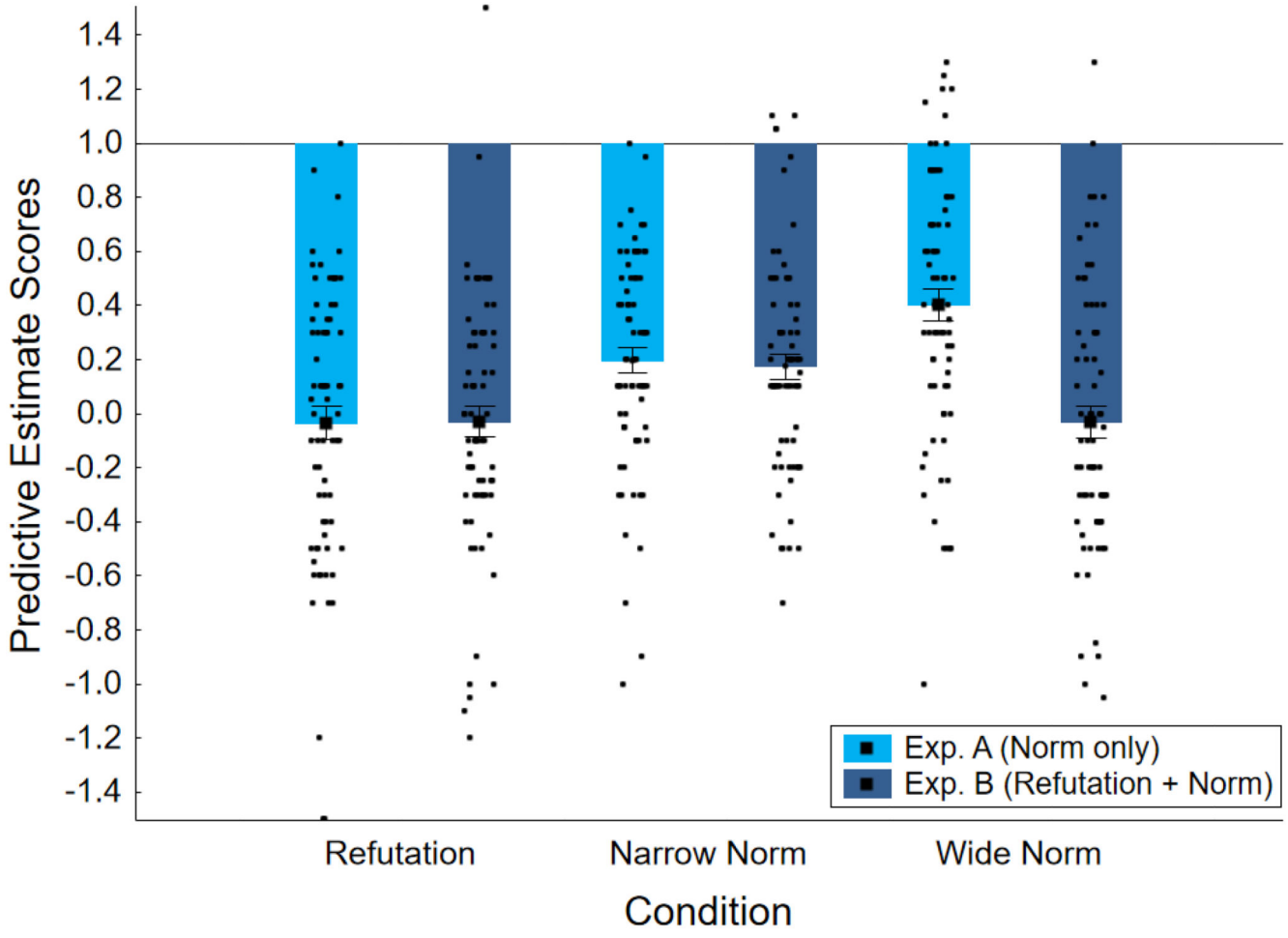

Figure 3. Mean Predictive Estimate Scores Across Conditions in Experiment 2

Note. Predictive estimate scores of 1 indicate full claim endorsement; scores of 0 reflect fully-effective intervention; scores < 0 indicate hypercorrection; scores > 1 reflect increased belief. Norm information was only provided in narrow and wide norm conditions, either by itself (Experiment 2A) or together with the refutation (Experiment 2B). The two refutation conditions were identical. Error bars show standard error of the mean. Dots show raw data points (jittered).

Europe PMC Funders Author Manuscripts

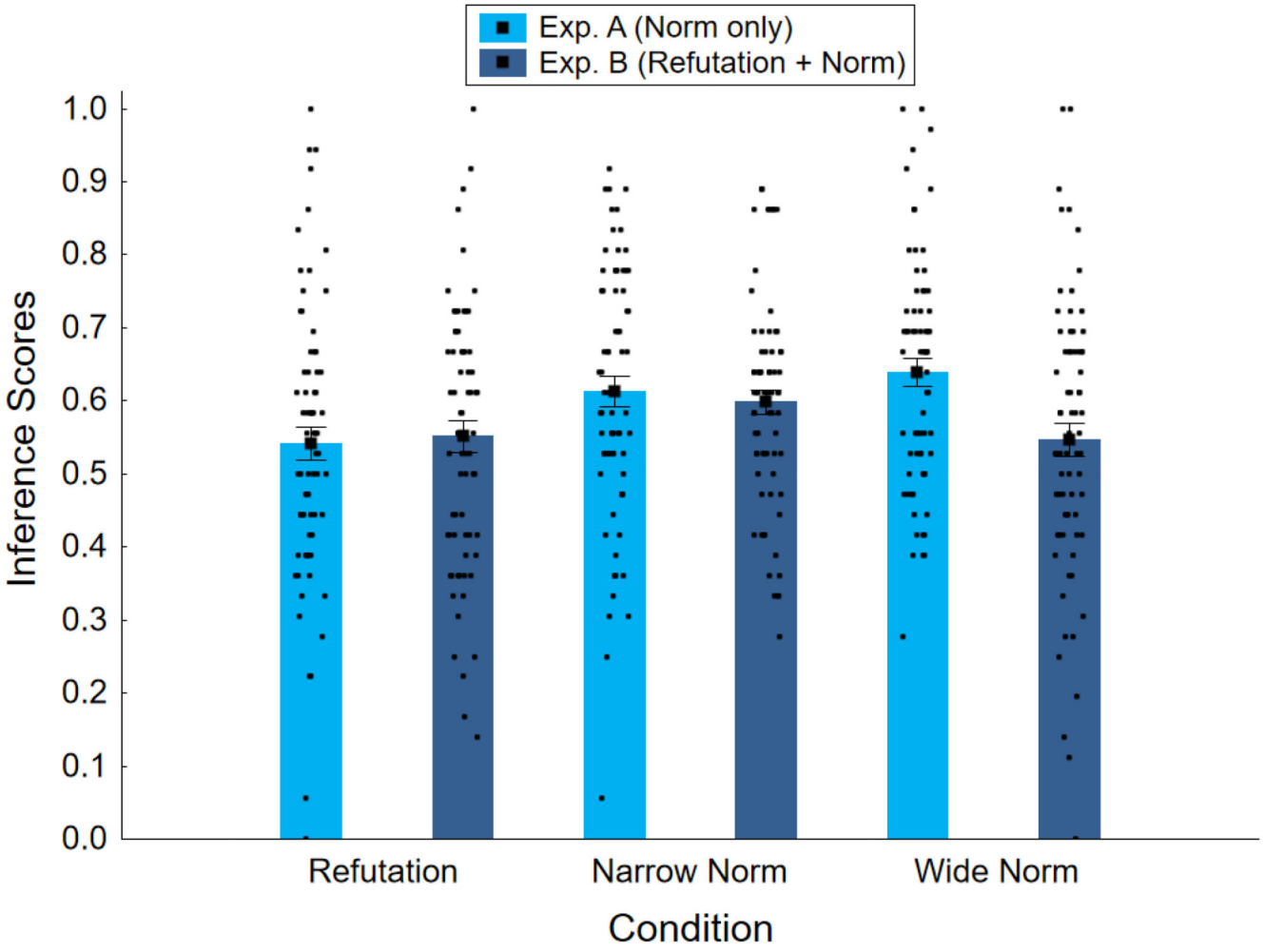

Figure 4. Mean Inference Scores Across Conditions in Experiment 2

Note. Greater inference scores indicate greater claim endorsement. Norm information was only provided in narrow and wide norm conditions, either by itself (Experiment 2A) or together with the refutation (Experiment 2B). The two refutation conditions were identical. Error bars show standard error of the mean. Dots show individual raw data points (jittered).

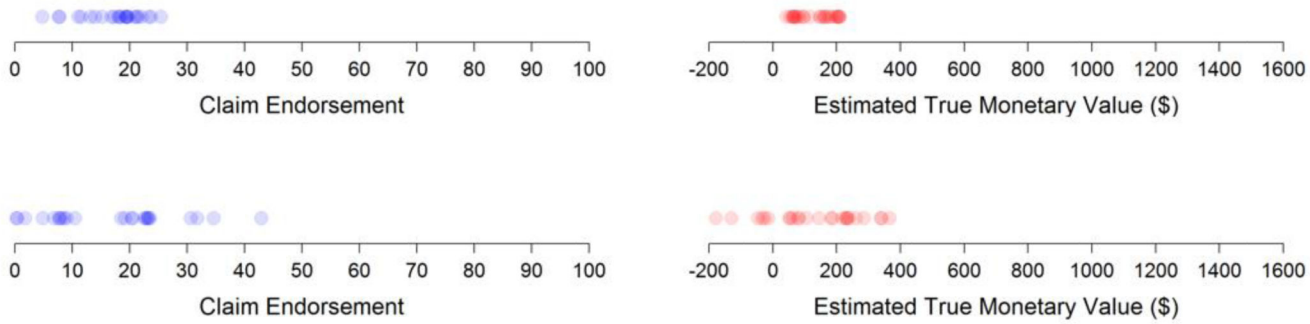

Figure 5. Graphical Norm Representations in Experiment 3

Note. Illustration of narrow (top panel) and wide (bottom panel) claim-endorsement (left panel) and predictive-estimate (right panel) norm distributions; each circle ostensibly represents one individual peer rating.

Europe PMC Funders Author Manuscripts

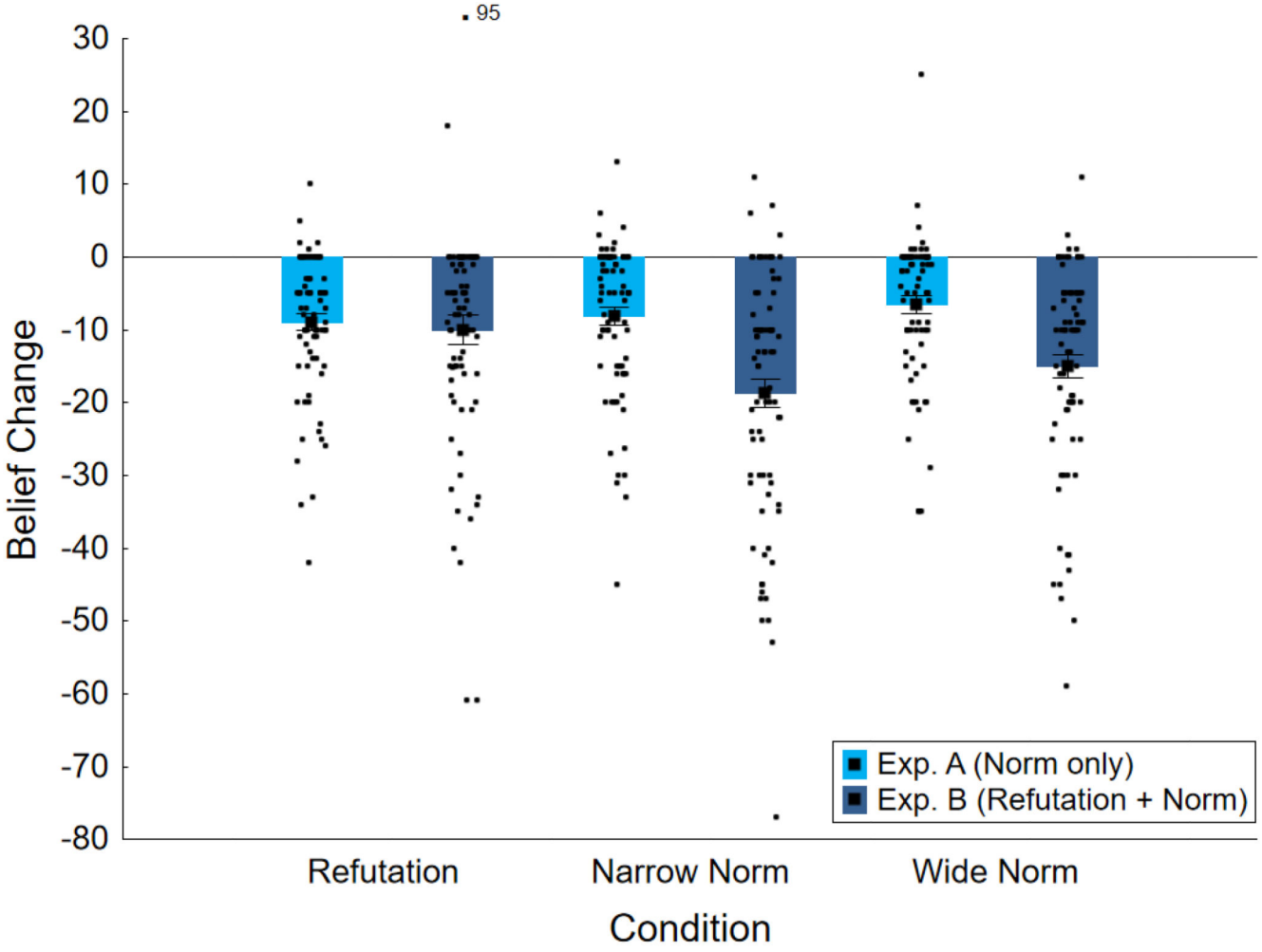

Figure 6. Mean Belief-Change Scores Across Conditions in Experiment 3

Note. Claim beliefs were measured on 0-100 scales. Norm information was only provided in narrow and wide norm conditions, either by itself (Experiment 3A) or together with the refutation (Experiment 3B). The two refutation conditions were identical. Error bars show standard error of the mean. Dots show individual raw data points (jittered; “95” labels a data point not shown in its actual y-axis position).

Europe PMC Funders Author Manuscripts

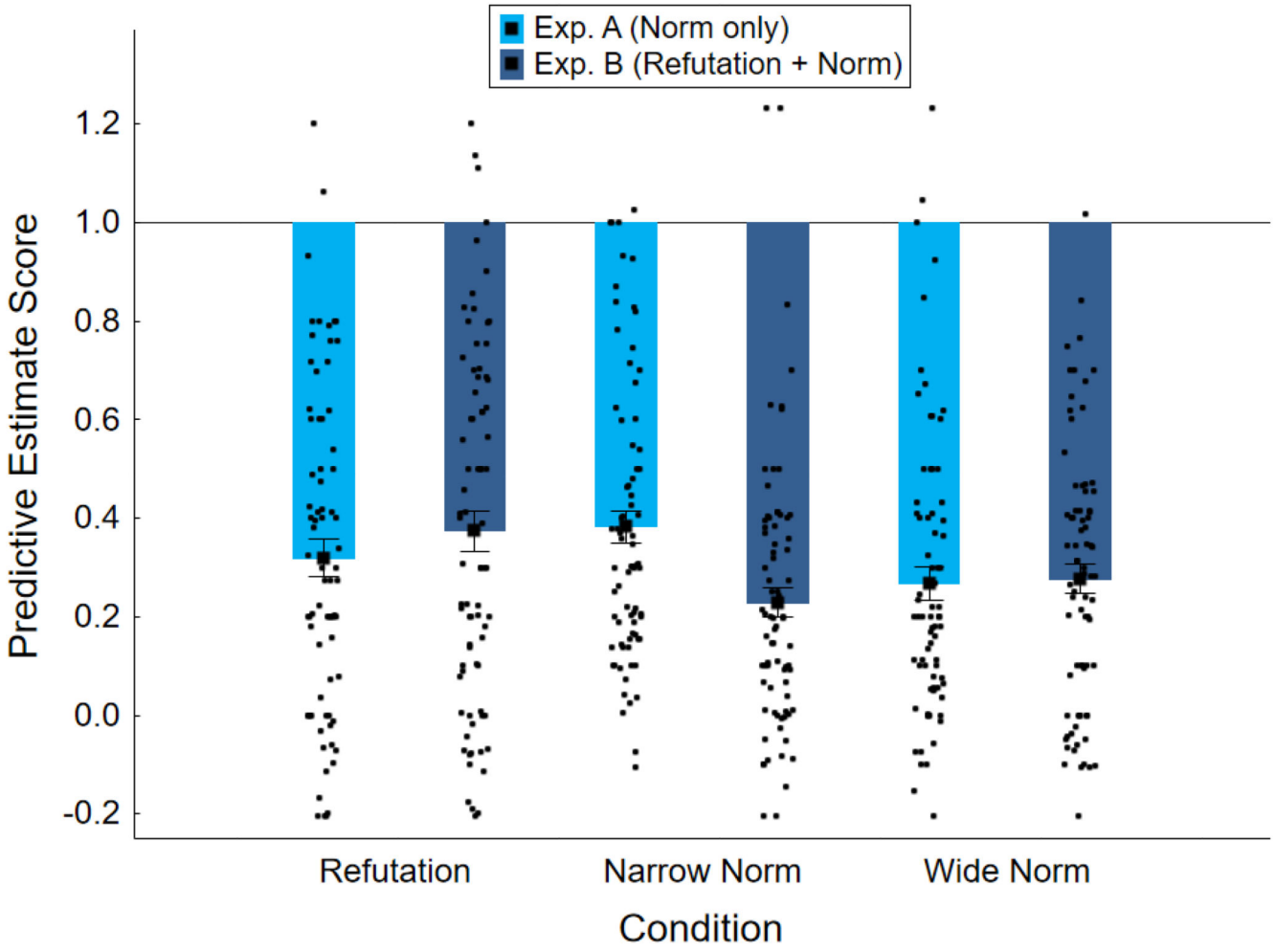

Figure 7. Mean Predictive Estimate Scores Across Conditions in Experiment 3

Note. Predictive estimate scores of 1 indicate full claim endorsement; scores of 0 reflect fully-effective intervention; scores < 0 indicate hypercorrection. Norm information was only provided in narrow and wide norm conditions, either by itself (Experiment 3A) or together with the refutation (Experiment 3B). The two refutation conditions were identical. Error bars show standard error of the mean. Dots show individual raw data points (jittered).

Europe PMC Funders Author Manuscripts

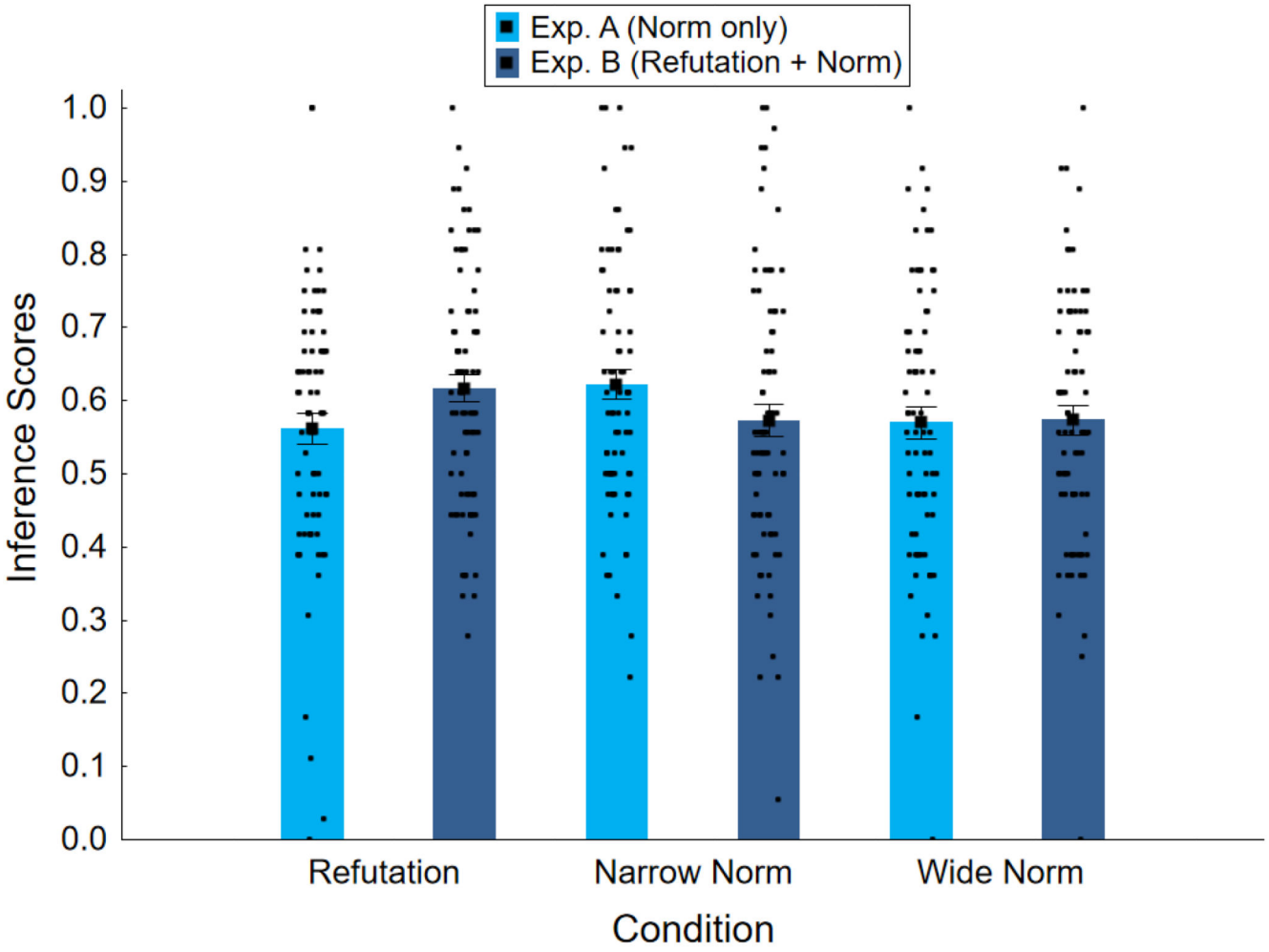

Figure 8. Mean Inference Scores Across Conditions in Experiment 3

Note. Greater inference scores indicate greater claim endorsement. Norm information was only provided in narrow and wide norm conditions, either by itself (Experiment 3A) or together with the refutation (Experiment 3B). The two refutation conditions were identical. Error bars show standard error of the mean. Dots show individual raw data points (jittered).