Believe or not to Believe in Conspiracy Claims? That is a Question for Signal Detection Theory

Maude Tagand[1], Dominique Muller[15], Cécile Nurra[2], Olivier Klein[3], Benjamin Auber-Teillaud[4], and Kenzo Nera[3]

[1]Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, LIP/PC2S, 38000 Grenoble, France

[2]Univ. Grenoble Alpes, LaRAC, 38000 Grenoble, France

[3]Université Libre de Bruxelles, CeSCuP, Belgique

[4]Université Paris Cité, Laboratoire de Psychologie sociale, France

[5]Institut Universitaire de France (IUF)

IN PRESS: PSYCHOLOGICAL SCIENCE

Author Note

Maude Tagand https://orcid.org/0009-0006-9478-5647

Dominique Muller https://orcid.org/0000-0001-9544-5317

Cécile Nurra https://orcid.org/0000-0001-6830-8851

Olivier Klein https://orcid.org/0000-0003-2737-8049

Benjamin Auber-Teillaud https://orcid.org/0000-0003-4470-1856

Kenzo Nera https://orcid.org/0000-0001-8518-386X

Correspondence concerning this article should be addressed to Maude Tagand, Université Grenoble Alpes, BP47, 38040 Grenoble Cedex 9, France, email: maude.tagand@gmail.com or Dominique Muller, Université Grenoble Alpes, BP47, 38040 Grenoble Cedex 9, France, email: dominique.muller@univ-grenoble-alpes.fr.

Abstract

Conspiracy mentality is conceptualized as a continuum. Research on this topic has focused on unwarranted conspiracy claims and the upper end of the conspiracy mentality continuum—people seeing conspiracies everywhere. This focus neglects warranted conspiracy claims and the lower end of the continuum. To better understand conspiracy mentality, we aimed to clarify both ends of the continuum using Signal Detection Theory. We examined how people evaluate warranted and unwarranted conspiracy claims across levels of conspiracy mentality in two studies with 331 French-speaking participants (Study 1) and 576 English-speaking participants (Study 2). Compared with participants high in conspiracy mentality, those low in conspiracy mentality not only believed less in conspiracies but also underestimated their prevalence. However, participants low in conspiracy mentality were more accurate at distinguishing warranted from unwarranted conspiracy claims. These results provide a better understanding of conspiracy mentality and its relationship with perceived truthfulness of conspiracies.

Keywords: Conspiracy claims, conspiracy theory, conspiracy mentality, signal detection theory, response bias, sensitivity

Research Transparency Statement

General Disclosures

Conflicts of interest: All authors declare no conflicts of interest. Funding: The research was supported by a grant allocated by ANR and a grant allocated by the fonds de la recherche scientifique (FRS-FNRS, Belgium). Artificial intelligence: ChatGPT (version 3.5, OpenIA, 2022) was employed to generate stimulus ideas, provide corrections to the manuscript (i.e., grammar and fluency). No other artificial intelligence assisted technologies were used in this research or the creation of this article. Ethics: This research was conducted under French jurisdiction and therefore does not require ethical approval for an experiment aimed to understand cognitive processes. Computational reproducibility: The computational reproducibility of the results in the main article (but not the supplementary materials) has been independently confirmed by the journal's STAR team. Open Science Framework (OSF): To ensure long-term preservation, all OSF files were registered at https://osf.io/kjgca/overview.

Study One

Preregistration: The hypotheses, methods, and analysis plan were preregistered (https://osf.io/k73x5) prior to the data collection. Minor deviations are detailed in the supplementary material (Table 1). Additionally, non-preregistered exploratory analyses were conducted and are identified as such in the main text. Materials: All study materials including instructions to participants, stimuli, and scripts are publicly available (https://osf.io/bk6up). Data: All raw data and cleaned data are publicly available (https://osf.io/tqyw4). Analysis scripts: All analysis scripts are publicly available (https://osf.io/fa76h/overview).

Study Two

Preregistration: The hypotheses, methods, and analysis plan were preregistered (https://osf.io/emrt3) prior to the data collection. There were no deviations from the

preregistration. Non-preregistered exploratory analyses were also conducted and are identified as such in the main text. Materials: All study materials including instructions to participants, stimuli, and scripts are publicly available (https://osf.io/fduqg). Data: All raw data and cleaned data are publicly available (https://osf.io/9c2rb). Analysis scripts: All analysis scripts are publicly available (https://osf.io/tmby2/overview).

To Believe or not to Believe in Conspiracy Claims? That is a Question for Signal Detection Theory

Because conspiracy theory researchers explore ways to address the challenges posed by conspiracy theories, they are often criticized for disregarding the existence of conspiracies (Pigden, 2006). But, of course, conspiracies do exist and failing to detect them when they happen can have deleterious consequences. Interestingly, the literature has mainly focused its narrative on people seeing conspiracies everywhere—"conspiracists"—but has neglected the other end of the continuum. We argue that clarifying both ends is necessary for a complete understanding. To that end, we mobilize Signal Detection Theory (SDT; Green & Swets, 1966), because it provides a framework to disentangle two (non-exclusive) tendencies: seeing conspiracies everywhere vs. nowhere, reflecting a response bias, and distinguishing between unwarranted conspiracy claims (referred to as "conspiracy theories") and warranted ones, reflecting sensitivity.

Conspiracy Claims and Conspiracy Mentality

The conspiracy theory literature mostly emerged from concerns about the spread of dubious conspiracy theories linked to societal issues (reduced vaccine coverage, voter abstention, hate crimes; Jolley et al., 2020). This might explain why research on the topic has mostly focused on such conspiracy claims (Frenken et al., 2024; Imhoff & Bertlich, 2024; see Van Prooijen & Van Vugt, 2018 and Wagner-Egger, 2023 for exceptions). This focus, however, created a blind spot: the possibility that some people tend to disbelieve actual conspiracies. Addressing this question requires distinguishing between warranted and unwarranted conspiracy claims.[1] We define unwarranted conspiracy claims as conspiracy

[1] Other authors have investigated differences between *plausible* and *implausible* conspiracy claims (Frenken et al., 2024; Hattersley et al., 2022). While related to our research question, these works do not mobilize SDT, nor do they seek to characterize participants low in conspiracy mentality.

claims (conspiracy theories) that are rejected by relevant epistemic authorities and based on poor evidence (Brotherton, 2013). We define warranted conspiracy claims (e.g., the Watergate scandal) as conspiracy claims whose truthfulness is attested by relevant epistemic authorities (experts, scientific institutions, and governments). While this distinction does not strictly equate to *true* versus *false* conspiracy claims—since epistemic authorities are fallible and can only provide the best available guesses about the truth (Constantin & Grundmann, 2020)—we assume that the position of epistemic authorities provides a reasonable proxy for the truth or falsity of these claims.

The conspiracy theory literature has shown that people can endorse specific conspiracy theories, but also that individuals differ in their stable propensity to endorse such theories (Sutton & Douglas, 2014). This is illustrated by studies showing that beliefs in conspiracy theories are strongly and positively correlated (Swami et al., 2010, 2011), suggesting that such beliefs stem from a general disposition. Imhoff and Bruder (2014) call this disposition "*conspiracy mentality"*, describing it as a stable ideological belief system "conceptualized as a continuum ranging from naïve trust in the canonical version of contemporary history to extremely paranoid conspiracy thinking" (Imhoff & Bruder, 2014, p. 40).

Other definitions focus exclusively on the core belief associated with scoring high on conspiracy mentality (e.g., Sutton & Douglas, 2014; Wood et al., 2012). This might have led researchers to focus not only on unwarranted conspiracy claims, but also on individuals scoring high in conspiracy mentality (while still comparing them to participants scoring lower). This dual focus revealed that, consistent with these definitions, participants scoring high in conspiracy mentality were more likely to endorse unwarranted conspiracy claims than those scoring lower (Imhoff & Bruder, 2014). However, as we argued, paying attention to the

whole continuum *and* testing how participants react, not only to unwarranted conspiracy claims but also to warranted ones, raises important questions.

First, based on current definitions of conspiracy mentality, we can predict that people scoring high should endorse both warranted and unwarranted conspiracy claims more than those scoring low. In addition, Imhoff and Bruder's (2014) definition implies an over-rejection of conspiracy claims by people scoring low on conspiracy mentality, creating a different type of bias. Second, considering both types of conspiracy claims raises another question: Are people low (vs. high) in conspiracy mentality better at distinguishing warranted from unwarranted conspiracy claims? Signal Detection Theory (SDT; Green & Swets, 1966) provides an adequate framework to investigate these questions and formalize our predictions.

Applying a Signal Detection Theory to the Study of Conspiracy Claims

Originally developed to explain how people differentiate between signal and noise in perceptual tasks (Green & Swets, 1966), SDT allows us the computation of two key indices: *response bias* and *sensitivity*. The first index, response bias, refers to the overall tendency toward one of two possible responses. In a task in which participants judge the truthfulness of a set of conspiracy claims (choosing between 'true' and 'false'), response bias reflects a preference for the answer 'true', indicating a tendency to endorse conspiracy claims (a *liberal bia*s), or a preference for the answer 'false', indicating a tendency to reject conspiracy claims (a *conservative bias*). We will use these response biases in two ways. First, we will compare response bias across different levels of conspiracy mentality. Second, we will test the response bias at a specific level of conspiracy mentality against an absence of bias (i.e., the actual prevalence of warranted conspiracy claims in the task). We will label this an *objective bias*. The second index, sensitivity, captures participants' accuracy at distinguishing between warranted and unwarranted conspiracy claims.

Applying SDT allows us to formulate three hypotheses. The first two hypotheses both concern the response bias, which ranges from an extreme conservative bias for lower values (corresponding to a probability of 'true' answer of 0) to extreme liberal bias for higher values (corresponding to a probability of 'true' answer of 1; a probability of .5 neither bias). The first hypothesis, consistent with all the definitions of conspiracy mentality, predicts that conspiracy mentality will be positively associated with response bias. Specifically, lower values of conspiracy mentality should lean toward a conservative bias and higher values toward a liberal bias (Hyp. 1a). The second hypothesis focuses on people low in conspiracy mentality. Do they exhibit no bias or an objective conservative bias (a probability value significantly lower than .5)? By qualifying the lower end of the continuum as a "naïve trust in the canonical version of contemporary history", Imhoff and Bruder's (2014) definition implies that participants low in conspiracy mentality might reject the existence of conspiracies in general, including warranted ones. Hence, people at the lower end of the continuum of conspiracy mentality could go as far as exhibiting an objective conservative bias in conspiracy detection (Hyp. 1b).

The third hypothesis concerns sensitivity. Research has shown that low conspiracy mentality is associated with better critical thinking skills (Lantian et al., 2021) and analytic reasoning (Yelbuz et al., 2022). Accordingly, we predict that conspiracy mentality will be negatively related to sensitivity: higher conspiracy mentality should be associated with lower accuracy at distinguishing unwarranted from warranted conspiracy claims (Hyp. 2).

We tested these hypotheses in French (Study 1) and English (Study 2) speaking samples. We relied on one conspiracy mentality measure in Study 1 and three in Study 2.

Study 1

Method

Participants

To approximate our required sample size, we conducted a power analysis for a multiple regression assessing how true responses differ between unwarranted and warranted conspiracy claims depending on conspiracy mentality. Because no previous study has tested this effect, we used the closest available proxy: the relationship between conspiracy mentality and analytic reasoning ($r = -.189$; Yelbuz et al., 2022). Based on this effect size and assuming $\alpha = .05$ and power = .80, we estimated a required sample size of 213. Because this effect size was only a proxy and mixed models often require larger samples, we planned to recruit 300 participants.

Three hundred thirty-one French-speaking participants from Belgium, France, and Switzerland were recruited on the Prolific crowdsourcing platform. We applied the preregistered exclusion criteria by removing 5 participants who reported conducting internet searches during the task, none who failed the attention check, and 16 who responded unrealistically quickly (less than 7 minutes). The final sample included 310 participants (147 men, 153 women, 10 others), aged between 20 and 68 years (M = 31, SD = 10).

Procedure

After providing informed consent, participants completed the three parts of the study in the following order: performing the classification task, completing the conspiracy mentality scale and demographic questions. Participants were then free to write an open-ended comment. An attention check was included after the classification task, and an Internet research check followed the demographic and some control variables (i.e., interest about relevant topics). The task was implemented on jsPsych (version 7; de Leeuw et al., 2023).

Classification task. Participants received instructions stating that they would read short descriptions of events. For each event, participants judged whether it actually took place. Participants indicated whether they believed the event was ‘true’ or ‘false’. They read 52 events inspired by Béna et al. (2023) or created using chatGPT (version 3.5, OpenAI,

2022). Among these, 26 referred to epistemically warranted conspiracy claims (i.e., conspiracies whose existence is acknowledged by epistemic authorities, such as the Watergate scandal or the MK-Ultra CIA project) and 26 referred to epistemically unwarranted conspiracy claims (i.e., conspiracy claims that are rejected by epistemic authorities, such as the 9/11 or vaccine conspiracy theories). The order of the events and the anchor placement (i.e., 'true' on the right vs. 'true' on the left) were randomized for each event.

Scale. Participants completed the Conspiracy Mentality Questionnaire (CMQ; Bruder et al., 2013), translated into French by Lantian et al. (2016; $\alpha = .86$). For each of the five items (e.g., "I think that many very important things happen in the world, which the public is never informed about"), participants answered on an 11-point Likert scale ranging from *0% - certainly not* to *100% - certainly*.

Demographic and control variables. Participants indicated their gender, age, and reported their agreement with the statements "I am interested in political and geopolitical news" and "I am interested in history" using a 7-point Likert scale from *1 - Completely disagree* to *7 - Completely agree*. These measures were included for exploratory purposes, to test whether controlling for them impacted the observed results. Participants also reported their political orientation on a scale from *1 - Radical left* to *7 - Radical right*.

Checks. As an attention check, participants were instructed to type "baguette" when asked for their favorite color. This ensured that participants who did not read the instruction carefully would fail the check. Participants were also asked whether they performed an internet search during the task, with assurances that their response (i.e., yes or no) would not affect their remuneration. We excluded participants who did not answer "baguette" to the first question and those who answered "yes" to the second question.

Results

Analytic Approach

Traditionally, SDT relies on two indices calculated for each participant using the formula $\frac{z(H) + z(FA)}{2}$ for the response bias and $z(H) - z(FA)$ for the sensitivity (Green & Swets, 1966). We also conducted this analysis for both studies. The analysis strategy, results, and corresponding figures are presented in the Supplementary Material (see Section 'Traditional SDT,' p. 7). However, the sensitivity formula assumes that both noise and signal distributions are normal and have equal standard deviations. These assumptions cannot be empirically tested. Lacking theoretical justification for them, we used a more robust alternative approach that remains valid when such assumptions may be violated (Stanislaw & Todorov, 1999). This alternative involves calculating these indices using a logistic regression for each participant (De Carlo, 1998).

These logistic regressions model the probability of responding "true" as a function of the warrantedness of the conspiracy claims (warranted vs. unwarranted). In this approach, each participant's response bias corresponds to the intercept of the logistic regression. Specifically, the intercept reflects the probability of a 'true' response, independently of the warrantedness of the claims. An absence of response bias corresponds to a probability of .5, indicating an equal number of 'true' and 'false' responses (i.e., log-odds of 0). Sensitivity corresponds to the slope of the regression and captures the difference in the probability of responding 'true' to warranted conspiracy claims (hits) versus unwarranted conspiracy claims (false alarms). A positive slope indicates that participants are more likely to answer 'true' for warranted than for unwarranted conspiracy claims, reflecting good sensitivity. Conversely, a negative slope indicates that participants go as far as being more likely to answer 'true' for unwarranted than for warranted conspiracy claims. We implemented this logistic regression using a mixed-effects model (De Carlo, 2010) to allow generalization across both participants and stimuli (Judd et al., 2012).

We conducted a mixed multinomial logistic regression using R (version 4.5.0; R Core Team, 2022) and the lme4 package (version 1.1.37; Bates et al., 2015) on participants' answers (coded 0 for false and 1 for true). As fixed effects, we specified the type of conspiracy claim (coded -0.5 for unwarranted and 0.5 for warranted), the conspiracy mentality, centered at the 16th percentile (participants low in conspiracy mentality), and the interaction between the two variables. The 16th percentile corresponds to one standard deviation below the mean which is a standard reference point for simple effects. Using this percentile helps avoid issues related to non-normal distribution, where the standard deviation might be biased. As random effects, we specified intercepts for participants and events (i.e., descriptions of conspiracy claims), as well as by-participant slopes for the type of conspiracy claim and by-events slopes for conspiracy mentality. These random effects were allowed to be correlated. When necessary, these models were then simplified at the level of the random effects following the recommendations of Bates et al. (2015). As there were only a few simplifications, we report only the cases in which they were applied.

This mixed multinomial logistic regression allows us to examine response bias in the target population (i.e., the intercept; Hyp. 1b) and how this bias varies as a function of conspiracy mentality (i.e., the effect of conspiracy mentality; Hyp. 1a). It also allows us to assess sensitivity in the target population (i.e., the effect of conspiracy claims) and how sensitivity depends on conspiracy mentality (i.e., the interaction between conspiracy mentality and conspiracy claims; Hyp. 2).

Confirmatory Analysis

In line with Hyp. 1a, the effect of conspiracy mentality was significant, $b = 0.28$, $z = 9.24$, $p < .001$, $OR = 1.32$, 95% CI [1.24, 1.40], suggesting that the response bias increases with conspiracy mentality. The analysis also revealed a significant negative intercept, $b = -0.96$, $z = -5.18$, $p < .001$, $OR = 0.38$, 95% CI [0.27, 0.55]. Because conspiracy mentality is

centered on the 16th percentile, this significant negative intercept indicates, in line with Hyp. 1b, an objective conservative bias for participants with a low conspiracy mentality. As can be seen in Fig. 1, taken together, these effects show that response bias starts from a significant conservative bias at the low end of conspiracy mentality and shifts toward a more liberal bias as conspiracy mentality rises, approaching or potentially exceeding the point of no bias (i.e., a probability of .5 for 'true' responses; see Section 'Exploratory Analysis'). Because we relied on a specific percentile to center conspiracy mentality—a somewhat arbitrary choice (Spiller et al., 2013)—we also conducted floodlight analyses to identify regions along the conspiracy mentality continuum where response bias is significant and regions where it is not (see Section 'Floodlight Analysis,' p. 1 of the Supplementary Material).

The conspiracy claim type by conspiracy mentality interaction effect was also significant, $b = -0.41$, $z = -7.53$, $p < .001$, $OR = 0.66$, 95% CI [0.59, 0.74], providing evidence that the accuracy at distinguishing between warranted and unwarranted conspiracy claims decreases as the conspiracy mentality increases (Hyp. 2). To further ensure that this effect was not driven by participants' interest in history, politics, or geopolitics, we conducted additional analyses controlling for these variables and their interactions with scenario (Yzerbyt et al., 2004). The predicted interaction remained significant after controlling for these variables. Detailed information on the analytical strategy and the full results of these analyses can be found in the Supplementary Material (see Section 'Controlled Variables,' p. 3). Finally, although we had no explicit prediction regarding this effect, our analysis also showed that the effect of conspiracy type (warranted vs. unwarranted) was significant, $b = 3.58$, $z = 9.82$, $p < .001$, $OR = 35.80$, 95% CI [17.53, 73.10]. Because we centered conspiracy mentality on the 16th percentile, this significant effect indicates that participants low in conspiracy mentality were accurate at distinguishing between warranted and unwarranted conspiracy claims (Fig. 1).

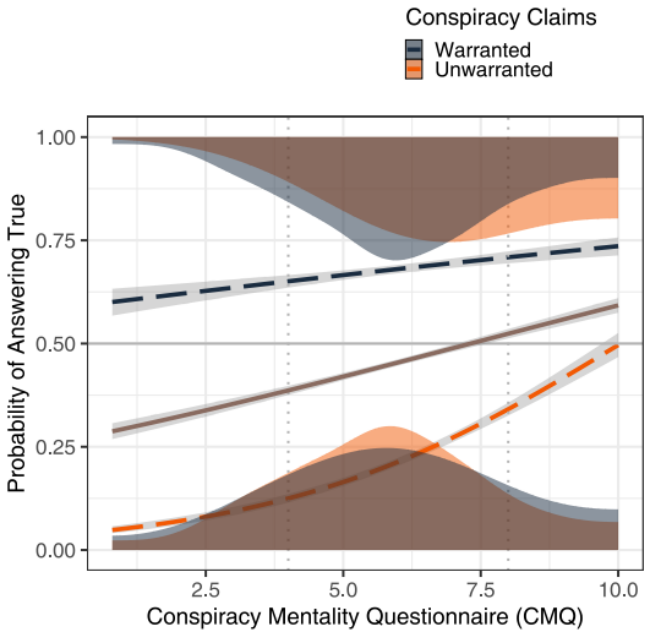

Fig. 1. Probability of answering true as a function of conspiracy mentality (CMQ) and type of conspiracy claim (warranted vs. unwarranted). The dashed lines represent predictions for warranted or unwarranted conspiracy claims, while the solid brown line shows the overall trend. In SDT terms, this overall trend amounts to the response bias (where 0 corresponds to an extreme conservative bias and 1 corresponds to an extreme liberal bias) and the difference between the dashed lines amounts to discrimination sensitivity. Shaded areas illustrate the distribution of 'true' (top) and 'false' (bottom) responses at each CMQ level and for warranted or unwarranted conspiracy claims. Error bars represent 95% confidence intervals for the predictions. A solid gray horizontal line at a probability of .5 indicates the actual prevalence of warranted conspiracy claims, while the two dotted vertical lines represent the CMQ scores at the 16th and 84th percentiles.

Exploratory Analysis

Our confirmatory analyses showed that the response bias increases with conspiracy mentality (Hyp. 1a) and that participants low in conspiracy mentality underestimated the actual prevalence of conspiracies (Hyp. 1b). Although this was not central to our main

research question, we also examined whether participants high in conspiracy mentality would overestimate the actual prevalence of conspiracies. In other words, we tested whether these participants exhibited a liberal bias. For this reason, we reran the confirmatory model focusing on participants high in conspiracy mentality (i.e., on the 84th percentile, corresponding to one standard deviation above the mean in a normal distribution). Unexpectedly, the intercept was not significant, $b = 0.15$, $z = 1.07$, $p = .28$, $OR = 1.16$, 95% CI [0.88, 1.53]. Therefore, although the positive intercept descriptively in the direction of an objective liberal bias, this effect was not statistically reliable.

Discussion

These results support our three hypotheses. However, they are limited to a French-speaking sample and a single conspiracy mentality measure. In Study 2, we aimed to extend our results to an English-speaking sample and to two other scales of conspiracy mentality. For each scale, we tested the same hypotheses.

Study 2

Method

Participants

Because we mistakenly believed[2] that, in Study 1, the effect of conspiracy mentality on distinguishing warranted from unwarranted conspiracy claims was not significant, we based our Study 2 power analysis on an equivalence test (TOSTER, Caldwell, 2022; Lakens, 2017). This analysis indicated that 854 participants would be needed. Due to resource constraints, we aimed for an integrative analysis, combining the 310 participants from Study 1 with approximately 540 participants recruited for Study 2. To gather 540 valid participants,

[2] A recording error in the Conspiracy Mentality Questionnaire led us to this erroneous conclusion. We only discovered this recording error during the analyses of Study 2. Accordingly, our power analysis was conducted with the goal to learn something from a potentially non-significant interaction. Nevertheless, the hypothesis of an interaction effect was still also preregistered.

we recruited 570 participants. Although this rationale was preregistered for the power analysis, it did not determine the analytical approach: Study 2 was first analyzed independently and the integrative analysis was conducted only afterward.

Five hundred and seventy-six English-speaking participants from the US and the UK were recruited through the Prolific crowdsourcing platform. To increase the likelihood of including individuals with a high conspiracy mentality, we applied a filter to ensure that 50% of participants self-reported as not vaccinated against COVID-19 (Bierwiaczonek et al., 2022). We applied the preregistered exclusion criteria by removing 16 participants who reported conducting internet searches during the task, 1 who failed the attention check, and 40 who responded unrealistically quickly (less than 7 minutes)—519 participants (254 men, 259 women, 6 others), aged between 18 and 77 years (M = 40, SD = 13) remained.

Procedure

The general procedure was the same as Study 1, with the following three deviations. First, we translated and adapted the material for an English-speakers sample from British or American culture. Second, in Study 1, several participants reported making mistakes because of the randomization of the anchor placement (i.e., 'true' on the right vs. 'true' on the left) across events. Instead, we randomized the anchor placement across participants, with half of the participants having the 'true' anchor on the right and the other half having 'true' anchor on the left. Third, we added the conspiracy mentality scale (CMS; Imhoff & Bruder, 2014) and the single item-conspiracy belief scale (SICBS; Lantian et al., 2016). The conspiracy mentality scale (CMS; Imhoff & Bruder, 2014; $\alpha = .93$) consists of 12 items (e.g., "Those at the top do whatever they want") to which participants responded on a 7-point Likert scale ranging from *1 - Strongly disagree* to *7 - Strongly agree*. The single item-conspiracy belief scale (SICBS; Lantian et al., 2016) includes an introductory text stating that the "official version" of some events may conceal the truth and that powerful individuals or organizations

could have secretly orchestrated them. This text is followed by the item "I think that the official version of the events given by the authorities very often hides the truth", rated on a 9-point Likert scale ranging from *1 - Completely false* to *9 - Completely true*. The presentation of these two scales was randomized, and they were presented immediately after the conspiracy mentality questionnaire (CMQ; Bruder et al., 2014; α = .87), ensuring that the CMQ was the first conspiracy beliefs scale completed to maximize similarity between Studies 1 and 2.

Results

Confirmatory Analysis

We ran the same model as in Study 1 for each of the three conspiracy mentality scales (i.e., Conspiracy Mentality Questionnaire, Conspiracy Mentality Scale, and Single Item Conspiracy Beliefs Scale; Fig. 2). In line with Hyp. 1a, this analysis revealed a significant effect of conspiracy mentality for the three scales (CMQ: $b = 0.31$, $z = 10.05$, $p < .001$, $OR = 1.36$, 95% CI [1.28, 1.44]; CMS: $b = 0.51$, $z = 10.28$, $p < .001$, $OR = 1.66$, 95% CI [1.51, 1.83]; SICBS: $b = 0.25$, $z = 9.69$, $p < .001$, $OR = 1.29$, 95% CI [1.22, 1.36]), suggesting that the response bias increases with conspiracy mentality. We also observed a significant negative intercept (CMQ: $b = -0.49$, $z = -3.43$, $p < .001$, $OR = 0.61$, 95% CI [0.47, 0.81]; CMS: $b = -0.56$, $z = -3.82$, $p < .001$, $OR = 0.57$, 95% CI [0.43, 0.76]; SICBS: $b = -0.68$, $z = -4.28$, $p < .001$, $OR = 0.51$, 95% CI [0.37, 0.69]). Because conspiracy mentality is centered on the 16th percentile, this significant negative intercept reveals, in line with Hyp. 1b across all three scales, an objective conservative bias for participants low in conspiracy mentality. As can be seen in Fig. 2, taken together, these effects show that response bias starts from a significant conservative bias at the low end of conspiracy mentality and shifts toward a more liberal bias as conspiracy mentality rises, approaching or potentially exceeding the point of no bias (see Section 'Exploratory Analysis'). Again, we also conducted floodlight analyses to

identify regions along the conspiracy mentality continuum where response bias is significant and regions where it is not (see Section 'Floodlight Analysis,' p. 1 of the Supplementary Material).

In line with Hyp. 2 and as we found in Study 1, the conspiracy claim type by conspiracy mentality interaction was significant (CMQ: b = -0.35, z = -6.30, p < .001, OR = 0.71, 95% CI [0.63, 0.79]; CMS: b = -0.55, z = -6.06, p < .001, OR = 0.58, 95% CI [0.48, 0.69]; SICBS: b = -0.24, z = -5.05, p < .001, OR = 0.79, 95% CI [0.72, 0.86]), indicating that the accuracy at distinguishing between warranted and unwarranted conspiracy claims decreases with conspiracy mentality. Again, the predicted interaction remained significant even after controlling for participants' interest in history, politics, or geopolitics. See Supplementary Material (Section 'Controlled Variables,' p. 3) for details of these control analyses. Finally, although we had no explicit prediction regarding this effect, the effect of conspiracy claim type was also significant (CMQ: b = 2.05, z = 7.47, p < .001, OR = 7.81, 95% CI [4.55, 13.38]; CMS: b = 2.11, z = 7.41, p < .001, OR = 8.25, 95% CI [4.72, 14.41]; SICBS: b = 2.13, z = 6.92, p < .001, OR = 8.38, 95% CI [4.59, 15.30]). Because we centered conspiracy mentality on the 16th percentile, this significant effect indicates that participants low in conspiracy mentality were accurate at distinguishing between warranted and unwarranted conspiracy claims.

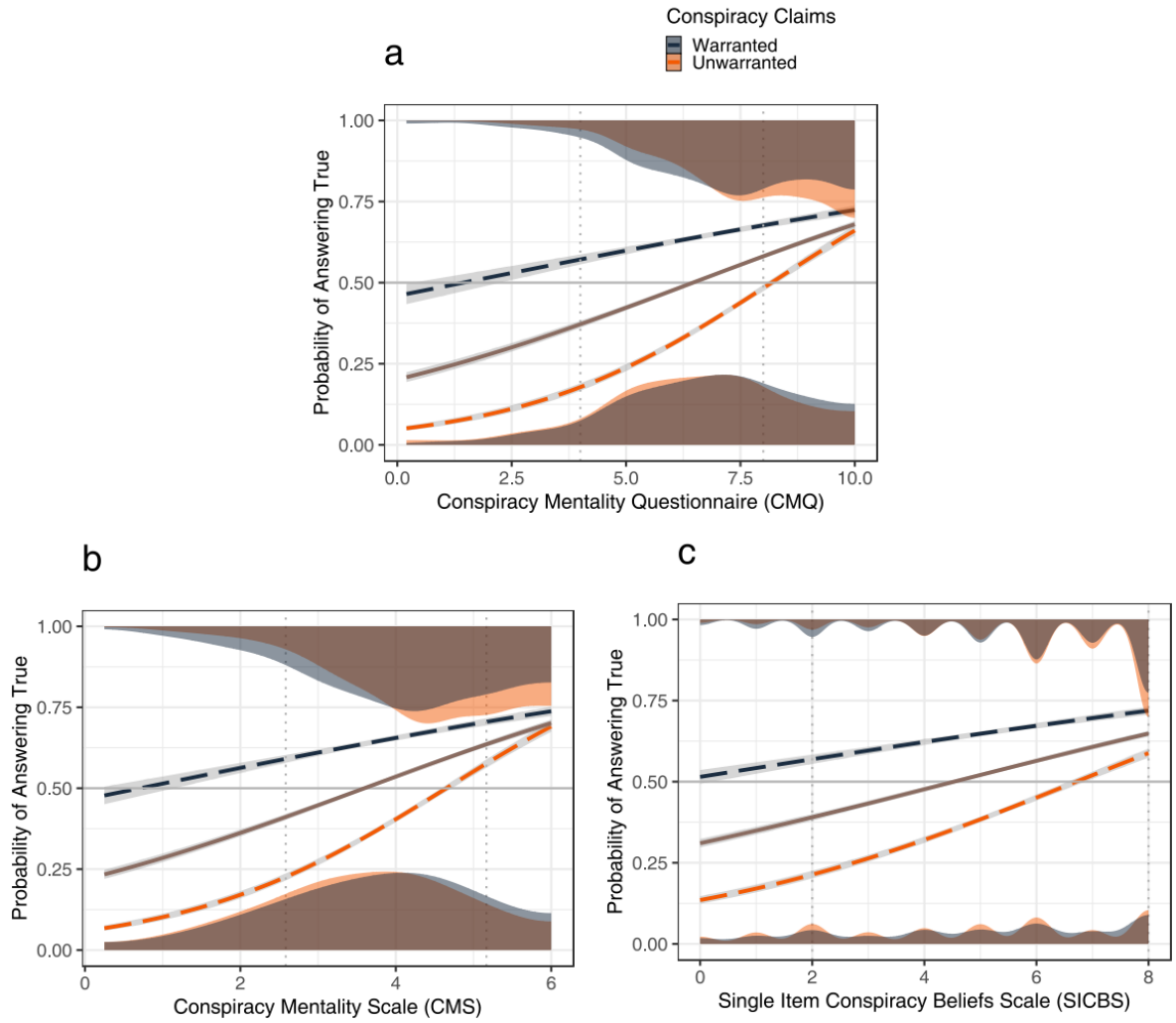

Fig. 2. Probability of answering true as a function of conspiracy mentality and conspiracy claim type (warranted vs. unwarranted) for (a) the CMQ, (b) the CMS and (c) the SICBS. The dashed lines represent predictions for warranted or unwarranted conspiracy claims, while the solid brown line shows the overall trend. In SDT terms, this overall trend amounts to the response bias (where 0 corresponds to an extreme conservative bias and 1 corresponds to an extreme liberal bias) and the difference between the dashed lines amounts to discrimination sensitivity. Shaded areas illustrate the distribution of 'true' (top) and 'false' (bottom) responses at each conspiracy mentality level and for warranted or unwarranted conspiracy claims. Error bars represent 95% confidence intervals for the predictions. A solid gray horizontal line at a probability of .50 indicates the actual

prevalence of warranted conspiracy claims, while the two dotted vertical lines represent the conspiracy mentality scores at the 16th and 84th percentiles.

Exploratory Analysis

As in Study 1, and although it was not central to our main research question, we aimed to test whether participants with a high conspiracy mentality exhibited an objective liberal bias. Following Bates et al. (2015), we removed the non-significant correlation between random intercepts and conspiracy-mentality slopes at the event level in all three models. Unlike in Study 1, the analysis showed a significant positive intercept for all three scales (CMQ: $b = 0.74$, $z = 6.75$, $p < .001$, $OR = 2.09$, 95% CI [1.69, 2.59]; CMS: $b = 0.76$, $z = 7.01$, $p < .001$, $OR = 2.13$, 95% CI [1.72, 2.63]; SICBS: $b = 0.84$, $z = 7.54$, $p < .001$, $OR = 2.31$, 95% CI [1.86, 2.87]). Contrary to Study 1, Study 2 revealed a significant objective liberal bias among participants high in conspiracy mentality, as reflected by these positive intercepts.

Integrative Analysis

To maximize statistical power and to improve effect size estimates, we ran an integrative analysis (Curran & Hussong, 2009) with the scale common to both studies, namely the CMQ. We used the same models as in each separate study.

Unsurprisingly, the effect of conspiracy mentality remained significant, $b = 0.30$, $z = 11.77$, $p < .001$, $OR = 1.35$, 95% CI [1.29, 1.42], confirming that the response bias increases with conspiracy mentality. We also found a significant negative intercept, $b = -0.67$, $z = -4.59$, $p < .001$, $OR = 0.51$, 95% CI [0.38, 0.68], again demonstrating an objective conservative bias for participants low in conspiracy mentality. In order to get a more complete picture of the general pattern for the response bias, and although this test was only exploratory, we also tested the objective bias for participants high in conspiracy mentality. Following Bates et al. (2015), in this model we removed the non-significant correlation

between random intercepts and conspiracy mentality slopes at the event level. Interestingly, the analysis showed a significant positive intercept, $b = 0.60$, $z = 5.66$, $p < .001$, $OR = 1.82$, 95% CI [1.48, 2.25], therefore revealing an objective liberal bias for participants high in conspiracy mentality. These effects interpreted together indicate that the positive relationship between response bias and conspiracy mentality starts from a significant conservative at the low end of conspiracy mentality and shifts toward a liberal bias at the high end. Going back to the main model, this integrative analysis also confirmed the conspiracy claim type by conspiracy mentality interaction, $b = -0.42$, $z = -8.58$, $p < .001$, $OR = 0.66$, 95% CI [0.60, 0.72]. This confirms that accuracy at distinguishing between warranted and unwarranted conspiracy claims decreases with conspiracy mentality. Finally, although we had no explicit prediction for this effect, the effect of conspiracy type (warranted vs. unwarranted) was significant, $b = 2.68$, $z = 9.31$, $p < .001$, $OR = 14.63$, 95% CI [8.31, 25.73], confirming that people low in conspiracy mentality are accurate at distinguishing warranted from unwarranted conspiracy claims.

Discussion

These results are again consistent with our three hypotheses. This suggests that our results extend to an English-speaking sample and to two other conspiracy mentality measures.

General Discussion

Research on conspiracy theories often focuses on people high in conspiracy mentality and adherence to unwarranted conspiracy claims. We considered the whole continuum of conspiracy mentality and used an SDT framework to study how participants reacted to both unwarranted and warranted conspiracy claims. Across two studies conducted in two cultural contexts—the French-speaking European context and the English-speaking American/British context—we found a consistent pattern of results.

First, our results revealed that response bias increased with conspiracy mentality (Hyp. 1a), ranging from an objective conservative bias (Hyp. 1b) at low levels of conspiracy mentality to an objective liberal bias at high levels. That latter (exploratory) result (i.e., the objective liberal bias) was not significant in Study 1, while it was in Study 2 and in the integrative analysis ($p < .001$).

Second, although results on response bias showed that participants low in conspiracy mentality appear biased, results on sensitivity revealed that they were still better than participants higher in conspiracy mentality at distinguishing warranted from unwarranted conspiracy claims (Hyp. 2). This difference remained when controlling for various personal interests relevant to performing well in such a task (in history, geopolitics, and politics; see supplementary material). These results call for further studies on the mechanisms responsible for this difference in accuracy.

Combined, our results better characterize both ends of the conspiracy mentality continuum. The conservative objective bias observed among participants low in conspiracy mentality aligns with Imhoff and Bruder's notion of "naïve trust." However, calling them "naïve" or "mindless sheeple" (Franks et al., 2017; Harambam & Aupers, 2016) would be too simplistic because, as we said, our results reveal they are also more accurate than their counterparts at distinguishing between warranted and unwarranted conspiracy claims. The latter result similarly enhances our understanding of the upper end of the spectrum. While the literature has shown that these participants are more likely to believe in unwarranted conspiracy claims, our studies are the first to demonstrate that they are also less accurate at distinguishing these claims from warranted ones.

Another contribution of this work concerns measurement, as conspiracy mentality scales conflate bias and accuracy in conspiracy detection. While the extensive length of our

paradigm might prevent it from being used as a measurement tool, developing an SDT-inspired measurement of conspiracy mentality constitutes a promising line of research.

We see at least four limitations. The first one pertains to our non-representative samples, both of which were recruited on Prolific in WEIRD countries. The second limitation pertains to the potential ambiguity of some items inherent to the reporting of real-world events, as responses may depend on participants' interpretation. The third limitation pertains to the proportion of 50% warranted and 50% unwarranted conspiracy claims we used. This proportion may underestimate the actual frequency of unwarranted conspiracy claims and may create a mismatch with participants' expectations. Future research could explore this issue by manipulating the proportion of warranted and unwarranted claims to better understand how people adapt their judgments. The fourth limitation is that our task rests on the assumption that the epistemic warrantedness of conspiracy claims is a reliable proxy for their truth or falsity. Although there are solid philosophical grounds for accepting this assumption despite the fallibility of epistemic authorities, our approach might fall short in cases where the truth of a conspiracy claim is not (yet) well established.

Despite these limitations the present work highlights the value of distinguishing response bias and sensitivity when considering individuals' tendency to endorse (or reject) conspiracy beliefs. Both gullibility and skepticism wear many disguises. The present studies shed light on what remains concealed behind these facades.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Béna, J., Rihet, M., Carreras, O., & Terrier, P. (2023). Repetition could increase the perceived truth of conspiracy theories. *Psychonomic Bulletin & Review*, *30*(6), 2397-2406. https://doi.org/10.3758/s13423-023-02276-4

Bierwiaczonek, K., Gundersen, A. B., & Kunst, J. R. (2022). The role of conspiracy beliefs for COVID-19 health responses: A meta-analysis. *Current Opinion in Psychology*, *46*, 101346. https://doi.org/10.1016/j.copsyc.2022.101346

Brotherton, R., French, C. C., & Pickering, A. D. (2013). Measuring belief in conspiracy theories: The generic conspiracist beliefs scale. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00279

Bruder, M., Haffke, P., Neave, N., Nouripanah, N., & Imhoff, R. (2013). Measuring individual differences in generic beliefs in conspiracy theories across cultures: Conspiracy mentality questionnaire. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00225

Caldwell, A. R. (2022). *Exploring equivalence testing with the updated TOSTER R Package*. https://doi.org/10.31234/osf.io/ty8de

Constantin, J., & Grundmann, T. (2020). Epistemic authority: Preemption through source sensitive defeat. *Synthese*, *197*(9), 4109-4130. https://doi.org/10.1007/s11229-018-01923-x

Curran, P. J., & Hussong, A. M. (2009). Integrative data analysis: The simultaneous analysis of multiple data sets. *Psychological Methods*, *14*(2), 81-100. https://doi.org/10.1037/a0015914

DeCarlo, L. T. (1998). Signal detection theory and generalized linear models. *Psychological Methods*, *3*(2), 186-205. https://doi.org/10.1037/1082-989X.3.2.186

DeCarlo, L. T. (2010). On the statistical and theoretical basis of signal detection theory and extensions: Unequal variance, random coefficient, and mixture models. *Journal of Mathematical Psychology*, *54*(3), 304-313. https://doi.org/10.1016/j.jmp.2010.01.001

de Leeuw, J.R., Gilbert, R.A., & Luchterhandt, B. (2023). jsPsych: Enabling an open-source collaborative ecosystem of behavioral experiments. *Journal of Open Source Software, 8*(85), 5351, https://joss.theoj.org/papers/10.21105/joss.05351

Franks, B., Bangerter, A., Bauer, M. W., Hall, M., & Noort, M. C. (2017). Beyond "monologicality"? Exploring conspiracist worldviews. *Frontiers in Psychology*, *8*, 861. https://doi.org/10.3389/fpsyg.2017.00861

Frenken, M., Reusch, A., & Imhoff, R. (2024). "Just because it's a conspiracy theory doesn't mean they're not out to get you": Differentiating the correlates of judgments of plausible versus implausible conspiracy theories. *Social Psychological and Personality Science*, 19485506241240506. https://doi.org/10.1177/19485506241240506

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* Wiley.

Imhoff, R., & Bertlich, T. (2024). Are conspiracy beliefs a sign of flawed cognition? Reexamining the association of cognitive style and skills with conspiracy beliefs. *Harvard Kennedy School Misinformation Review*. https://doi.org/10.37016/mr-2020-168

Imhoff, R., & Bruder, M. (2014). Speaking (un–)truth to power: Conspiracy mentality as a generalised political attitude. *European Journal of Personality*, *28*(1), 25-43. https://doi.org/10.1002/per.1930

Jolley, D., Mari, S., & Douglas, K. M. (2020). Consequences of conspiracy theories. In M. Butter & P. Knight (Éds.), *Routledge Handbook of Conspiracy Theories* (1re éd., p. 231-241). Routledge. https://doi.org/10.4324/9780429452734-2_7

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*(1), 54-69. https://doi.org/10.1037/a0028347

Harambam, J., & Aupers, S. (2016). 'I am not a conspiracy theorist': Relational identifications in the Dutch conspiracy milieu. *Cultural Sociology, 11*(1), 113-129. https://doi.org/10.1177/1749975516661959

Lakens, D. (2017). Equivalence tests: A practical primer for *t* tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, *8*(4), 355-362. https://doi.org/10.1177/1948550617697177

Lantian, A., Bagneux, V., Delouvée, S., & Gauvrit, N. (2021). Maybe a free thinker but not a critical one: High conspiracy belief is associated with low critical thinking ability. *Applied Cognitive Psychology*, *35*(3), 674-684. https://doi.org/10.1002/acp.3790

Lantian, A., Muller, D., Nurra, C., & Douglas, K. M. (2016). Measuring belief in conspiracy theories: Validation of a french and english single-item scale. *International Review of Social Psychology*, *29*(1), 1. https://doi.org/10.5334/irsp.8

Pigden, C. (2006). Popper revisited, or what is wrong with conspiracy theories?. In *Conspiracy theories* (pp. 17-43). Routledge.

R Core Team. (2022). *R: A language and environment for statistical computing* (Version 4.3.0) [Computer software]. Retrieved from http://www.R-project.org

Spiller, S. A., Fitzsimons, G. J., Lynch, J. G., & McClelland, G. H. (2013). Spotlights, floodlights, and the magic number zero: Simple effects tests in moderated regression. *Journal of Marketing Research*, *50*(2), 277-288. https://doi.org/10.1509/jmr.12.0420

Sutton, R. M., & Douglas, K. M. (2014). Examining the monological nature of conspiracy theories. In J.-W. van Prooijen & P. A. M. van Lange (Eds.), *Power, politics, and paranoia: Why people are suspicious of their leaders* (pp. 254–273). Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9781139565417

Swami, V., Chamorro-Premuzic, T., & Furnham, A. (2010). Unanswered questions: A preliminary investigation of personality and individual difference predictors of 9/11 conspiracist beliefs. *Applied Cognitive Psychology*, *24*(6), 749-761. https://doi.org/10.1002/acp.1583

Swami, V., Coles, R., Stieger, S., Pietschnig, J., Furnham, A., Rehim, S., & Voracek, M. (2011). Conspiracist ideation in Britain and Austria: Evidence of a monological belief system and

associations between individual psychological differences and real-world and fictitious conspiracy theories: Conspiracist ideation. *British Journal of Psychology*, *102*(3), 443-463. https://doi.org/10.1111/j.2044-8295.2010.02004.x

Prooijen, J.-W. van, & Vugt, M. van. (2018). Conspiracy theories: Evolved functions and psychological mechanisms. *Perspectives on Psychological Science*, *13*(6), 770–788. https://doi.org/10.1177/1745691618774270

Wagner-Egger, P., De Oliveira Fernandes, D., Carrel, J., & Delouvée, S. (2023). Do real and invented scandals fuel beliefs in conspiracy theories? *European Journal of Psychology Open*, *82*(1), 1-7. https://doi.org/10.1024/2673-8627/a000034

Wood, M. J., Douglas, K. M., & Sutton, R. M. (2012). Dead and alive: Beliefs in contradictory conspiracy theories. *Social Psychological and Personality Science*, *3*(6), 767-773. https://doi.org/10.1177/1948550611434786

Yelbuz, B. E., Madan, E., & Alper, S. (2022). Reflective thinking predicts lower conspiracy beliefs: A meta-analysis. *Judgment and Decision Making*, *17*(4), 720-744. https://doi.org/10.1017/S1930297500008913

Yzerbyt, V. Y., Muller, D., & Judd, C. M. (2004). Adjusting researchers' approach to adjustment: On the use of covariates when testing interactions. *Journal of Experimental Social Psychology, 40*(3), 424-431. https://doi.org/10.1016/j.jesp.2003.10.001