

To Believe or Not to Believe in Conspiracy Claims? That Is a Question for Signal Detection Theory

Psychological Science
1–11

© The Author(s) 2026

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/09567976261417245

www.psychologicalscience.org/PS



Maude Tagand¹ , Dominique Muller^{1,2} , Cécile Nurra³ ,
Olivier Klein⁴ , Benjamin Aubert-Teillaud⁵ , and
Kenzo Nera^{4,6} 

¹Université Grenoble Alpes, Université Savoie Mont Blanc, LIP/PC2S, France; ²Institut Universitaire de France;³Université Grenoble Alpes, LaRAC, France; ⁴Center for Social and Cultural Psychology,Université libre de Bruxelles, Belgium; ⁵Université Paris Cité, Laboratoire de Psychologie Sociale, France;and ⁶Fonds de la Recherche Scientifique (FRS-FNRS), Belgium

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 adult participants from France, Switzerland, and Belgium (Study 1) and 576 English-speaking adult participants from the United States and the United Kingdom (Study 2), both groups recruited via Prolific. 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 the perceived truthfulness of conspiracies.

Keywords

conspiracy claims, conspiracy theory, conspiracy mentality, signal detection theory, response bias, sensitivity

Received 1/10/25; revision accepted 12/26/25

Because conspiracy-theory researchers explore ways to address the challenges posed by conspiracy theories, they are often criticized for disregarding the existence of conspiracies (e.g., Pigden, 1995). But, of course, conspiracies do exist and failing to detect them when they happen can have deleterious consequences. Interestingly, the literature has primarily 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 mobilized signal detection theory (SDT; Green & Swets, 1966) because it provides a framework to disentangle

two (nonexclusive) tendencies: seeing conspiracies everywhere versus nowhere, reflecting a response bias, and distinguishing between unwarranted conspiracy claims (referred to as “conspiracy theories”) and warranted ones, reflecting sensitivity.

Corresponding Authors:

Maude Tagand, Université Grenoble Alpes

Email: maude.tagand@gmail.com

Dominique Muller, Université Grenoble Alpes

Email: dominique.muller@univ-grenoble-alpes.fr

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 et al., 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 claims (conspiracy theories) that are rejected by relevant epistemic authorities (experts, scientific institutions, and governments) and are based on poor evidence (Brotherton et al., 2013). We define warranted conspiracy claims (e.g., the Watergate scandal) as conspiracy claims attested by relevant epistemic authorities. Although this distinction does not strictly equate to *true* versus *false* conspiracy claims—because epistemic authorities are fallible and can provide only 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 a “*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 with 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, raise important questions.

First, on the basis of 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? 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 bias), 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 concern the response bias, which ranges from an extreme conservative bias for lower values (corresponding to a probability of a “true” answer of 0) to extreme liberal bias for higher values (corresponding to a probability of a “true” answer of 1; a probability of .5 representing 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 (Hypothesis 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 (Hypothesis 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 predicted 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 (Hypothesis 2).

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

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 the Agence Nationale de la Recherche (ANR) and a grant allocated by the Fonds de la Recherche Scientifique (FRS-FNRS, Belgium). **Artificial intelligence:** ChatGPT (version 3.5, OpenAI, 2022) was employed to generate stimulus ideas and 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 at understanding cognitive processes. **Open Science Framework (OSF):** To ensure long-term preservation, we registered all OSF files at <https://osf.io/kjgca/overview>.

Study 1 disclosures

Preregistration: The hypotheses, methods, and analysis plan were preregistered (<https://osf.io/k73x5>) prior to data collection. Minor deviations are detailed in the Supplemental Material (Table S1). Additionally, nonpreregistered 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>). **Computational reproducibility:** The computational reproducibility of the results in the main article (but not the Supplemental Material) has been independently confirmed by the journal’s STAR team.

Study 2 disclosures

Preregistration: The hypotheses, methods, and analysis plan were preregistered (<https://osf.io/emrt3>) prior to data collection. There were no deviations from the preregistration. Nonpreregistered 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>). **Computational reproducibility:** The computational reproducibility of the results in the main article (but not the Supplemental Material) has been independently confirmed by the journal’s STAR team.

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). On the basis of 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 and 16 who responded unrealistically quickly (in less than 7 min); no one failed the attention check. The final sample included 310 participants (147 men, 153 women, 10 others), aged between 20 and 68 years ($M = 31$ years, $SD = 10$).

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 completing the 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 questions 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 warranted conspiracy claims (i.e., conspiracies such as the Watergate scandal or the MK-Ultra CIA project, whose existence is acknowledged by authorities) and 26 referred to 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 and 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 affected 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.

Ethics statement. This research was conducted under French jurisdiction. In accordance with applicable regulations, formal ethical approval was not required for experiments aimed at understanding cognitive processes. The studies complied with the ethical guidelines and legal requirements of the authors’ institutions and the countries in which the research was conducted, and all participants provided informed consent.

Results

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 Supplemental Material available online (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 (DeCarlo, 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 (DeCarlo, 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). For fixed effects, we specified the type of conspiracy claim (coded -0.5 for unwarranted and 0.5 for warranted); conspiracy mentality, centered at the 16th percentile (i.e., participants low in conspiracy mentality); and the interaction between the two variables. The 16th percentile corresponds to 1 standard deviation below the mean, which is a standard reference point for simple effects. Using this percentile helps to avoid issues related to nonnormal distribution, in which the standard deviation might be biased. For 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). Because there were only a few simplifications, we report only the cases in which they were applied.

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

Confirmatory analysis. Consistent with Hypothesis 1a, the effect of conspiracy mentality was significant, $b = 0.28$, $z = 9.24$, $p < .001$, odds ratio (OR) = 1.32, 95% confidence interval (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 Hypothesis 1b, an objective conservative bias for participants with a low conspiracy mentality. As can be seen in Figure 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 the

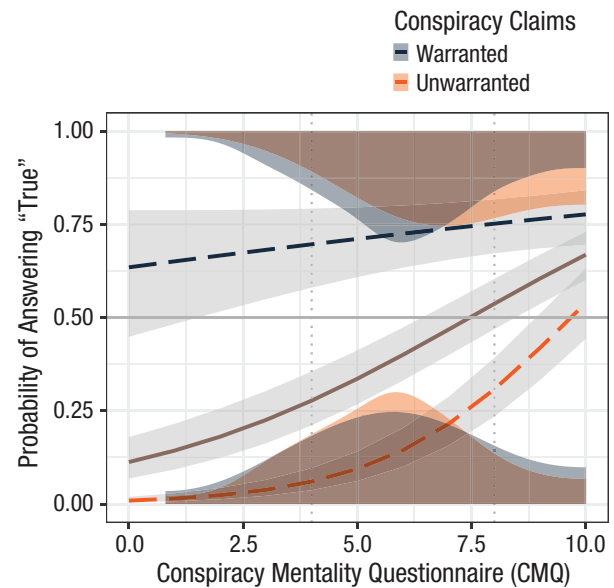


Fig. 1. Probability of answering “true” as a function of conspiracy mentality (as assessed by the CMQ) and type of conspiracy claim (warranted vs. unwarranted). The dashed lines represent predictions for warranted or unwarranted conspiracy claims, whereas the solid brown line shows the overall trend. In signal-detection-theory 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. Grey areas around slopes represent 95% confidence intervals for the predictions. The solid gray horizontal line at a probability of .50 indicates the actual prevalence of warranted conspiracy claims, whereas the two dotted vertical lines represent the CMQ scores at the 16th and 84th percentiles.

section on “Exploratory Analysis” in the Supplemental Material). 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 the section called “Floodlight Analysis” on p. 1 of the Supplemental 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 (Hypothesis 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 Supplemental Material (see the “Controlled Variables” section, 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).

Exploratory analysis. Our confirmatory analyses showed that the response bias increases with conspiracy mentality (Hypothesis 1a) and that participants low in conspiracy mentality underestimated the actual prevalence of conspiracies (Hypothesis 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 1 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 was 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 believed² 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. Because of 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, 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 only conducted afterward.

Five hundred and seventy-six English-speaking participants from the United States and the United Kingdom 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 being vaccinated against COVID-19 (Bierwaczek 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 min)—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-speaking 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 the “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 CMS (Imhoff & Bruder, 2014; $\alpha = .93$) consists of 12 items (e.g., “Those at the top do whatever they want”) to which participants responded using a 7-point Likert scale ranging from 1, *strongly disagree*, to 7, *strongly agree*. The 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 CMQ (Bruder et al., 2013; $\alpha = .87$), ensuring that the CMQ was the first conspiracy-beliefs scale completed to maximize similarity between Studies 1 and 2.

Ethics statement. This research was conducted under French jurisdiction. In accordance with applicable regulations, formal ethical approval was not required for

experiments aimed at understanding cognitive processes. The studies complied with the ethical guidelines and legal requirements of the authors' institutions and the countries in which the research was conducted, and all participants provided informed consent.

Results

Confirmatory analysis. We ran the same model as in Study 1 for each of the three conspiracy-mentality scales (i.e., the CMQ, the CMS, and the SICBS; Fig. 2). In line with Hypothesis 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, consistent with Hypothesis 1b across all three scales, an objective conservative bias for participants low in conspiracy mentality. As can be seen in Figure 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 the Exploratory Analysis section in the Supplemental Material). 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 the "Floodlight Analysis" section on p. 1 of the Supplemental Material).

In line with Hypothesis 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 we controlled for participants' interest in history, politics, or geopolitics. See the Supplemental Material (the "Controlled Variables" section; 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.

Exploratory analysis. As in Study 1, we aimed to test whether participants with a high conspiracy mentality exhibited an objective liberal bias—although this question was not central to our main research question. Following Bates et al. (2015), we removed the nonsignificant correlation between random intercepts and conspiracy-mentality slopes at the event level in all three models. Unlike 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, we also tested the objective bias for participants high in conspiracy mentality, although this test was only exploratory. Following Bates et al. (2015), in this model we removed the nonsignificant 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 bias 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

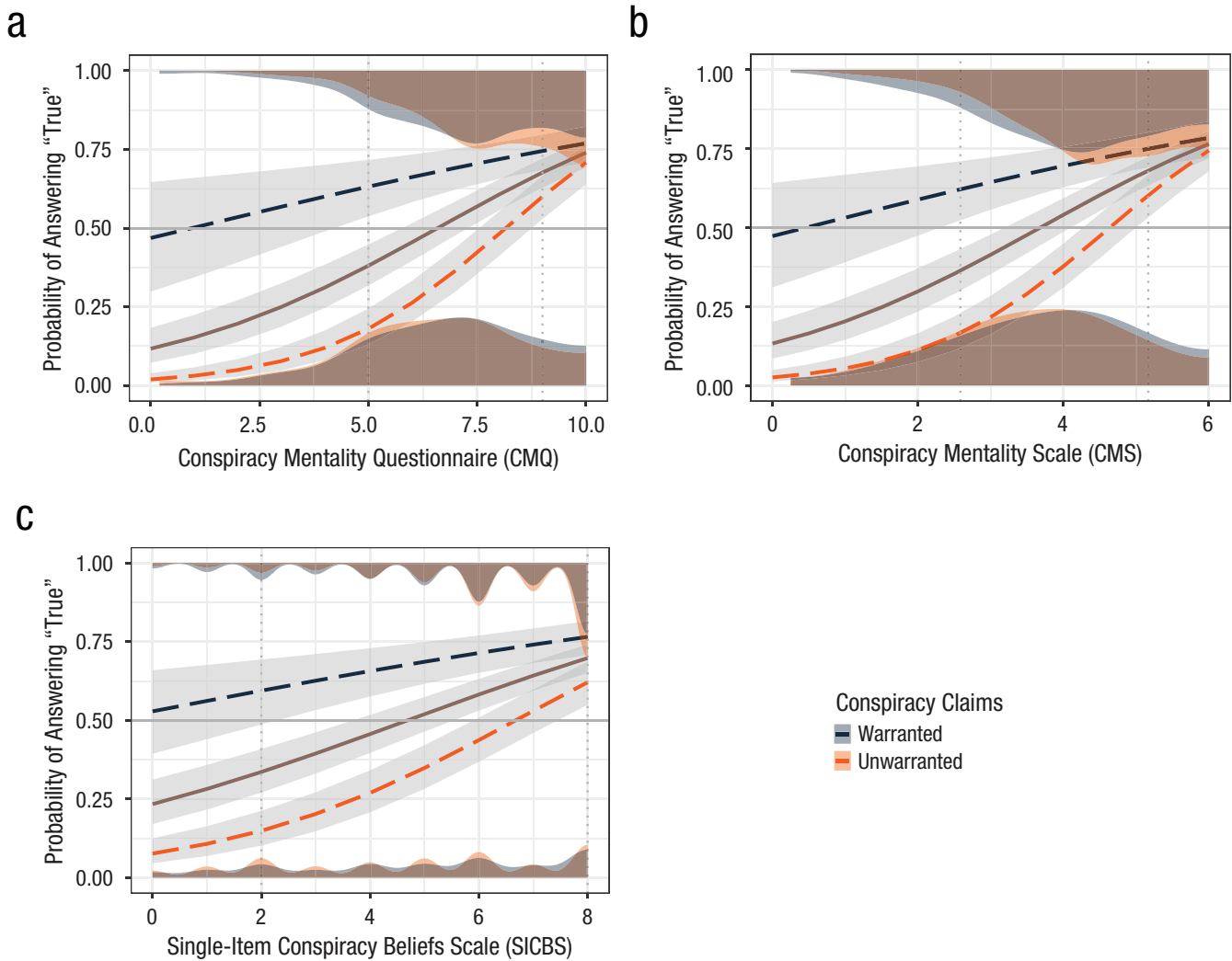


Fig. 2. Probability of answering "true" as a function of conspiracy mentality and conspiracy-claim type (warranted vs. unwarranted) for (a) the Conspiracy Mentality Questionnaire, (b) the Conspiracy Mentality Scale, and (c) the Single-Item Conspiracy Belief Scale. The dashed lines represent predictions for warranted or unwarranted conspiracy claims, whereas 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. Grey areas around slopes represent 95% confidence intervals for the predictions. The solid gray horizontal line at a probability of .50 indicates the actual prevalence of warranted conspiracy claims, whereas the two dotted vertical lines represent the conspiracy-mentality scores at the 16th and 84th percentiles.

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 (Hypothesis 1a), ranging from an objective conservative bias (Hypothesis 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, but 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 at distinguishing warranted from unwarranted conspiracy claims than participants higher in conspiracy mentality (Hypothesis 2). This difference remained when we controlled for various personal interests relevant to performing well in such a task (interests in history, geopolitics, and politics; see the Supplemental Material). These results call for further studies on the mechanisms responsible for this difference in sensitivity.

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. Although 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, because conspiracy-mentality scales conflate bias with accuracy in conspiracy detection. Although the extensive length of our paradigm might prevent it from being used as a measurement tool, it is clear that 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 nonrepresentative samples, both of which were recruited on Prolific in Western, educated, industrialized, rich, and democratic (WEIRD) countries. The

second limitation pertains to the potential ambiguity of some items inherent to the reporting of real-world events, because 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 in which 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.

Transparency

Action Editor: Yoel Inbar

Editor: Simine Vazire

Author Contributions

Maude Tagand: Conceptualization; Formal analysis; Funding acquisition; Investigation; Methodology; Software; Visualization; Writing – original draft.

Dominique Muller: Conceptualization; Formal analysis; Funding acquisition; Methodology; Supervision; Writing – original draft.

Cécile Nurra: Conceptualization; Funding acquisition; Supervision; Writing – review & editing.

Olivier Klein: Conceptualization; Funding acquisition; Supervision; Writing – review & editing.

Benjamin Aubert-Teillaud: Conceptualization; Methodology; Writing – review & editing.

Kenzo Nera: Conceptualization; Funding acquisition; Methodology; Supervision; Validation; Writing – original draft.

Declaration of Conflicting Interests

The authors declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding

The research was supported by a grant allocated by the Agence Nationale de la Recherche (ANR) and a grant allocated by the Fonds de la Recherche Scientifique (FRS-FNRS, Belgium).

Artificial Intelligence

ChatGPT (version 3.5, OpenAI, 2022) was employed to generate stimulus ideas and 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 at understanding cognitive processes. Open Science Framework (OSF): To ensure long-term preservation, we registered all OSF files at <https://osf.io/kjgca/overview>.

Open Practices

Open practices for this article are described in the Research Transparency Statement section, which appears at the end of the Introduction section in the main text.

ORCID iDs

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 Aubert-Teillaud  <https://orcid.org/0000-0003-4470-1856>

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

Supplemental Material

Additional supporting information can be found at <http://journals.sagepub.com/doi/suppl/10.1177/09567976261417245>

Notes

1. Other authors have investigated differences between plausible and implausible conspiracy claims (Frenken et al., 2024; Hattersley et al., 2022). Although related to our research question, these works do not utilize SDT, nor do they seek to characterize participants low in conspiracy mentality.
2. A recording error in the CMQ 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 of learning something from a potentially nonsignificant interaction. Nevertheless, the hypothesis of an interaction effect was still also preregistered.

References

- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <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>
- Bierwiazzonek, 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, Article 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, Article 279. <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, Article 225. <https://doi.org/10.3389/fpsyg.2013.00225>
- Caldwell, A. R. (2022). *Exploring equivalence testing with the updated TOSTER R Package*. PsyArXiv. <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), Article 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, Article 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*, 16(6), 684–696. <https://doi.org/10.1177/19485506241240506>
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. Wiley.
- 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>
- Hattersley, M., Brown, G. D., Michael, J., & Ludvig, E. A. (2022). Of tinfoil hats and thinking caps: Reasoning is more strongly related to implausible than plausible conspiracy beliefs. *Cognition*, 218, Article 104956. <https://doi.org/10.1016/j.cognition.2021.104956>
- 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 (Eds.), *Routledge Handbook of Conspiracy Theories* (1st ed., pp. 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>
- 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), 1–14. <https://doi.org/10.5334/irsp.8>
- Pigden, C. (1995). Popper revisited, or what is wrong with conspiracy theories? *Philosophy of the Social Sciences*, 25(1), 3–34.
- R Core Team. (2022). *R: A language and environment for statistical computing* (Version 4.3.0) [Computer software]. <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>
- Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, 31(1), 137–149. <https://doi.org/10.3758/BF03207704>
- 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: 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. *British Journal of Psychology*, 102(3), 443–463. <https://doi.org/10.1111/j.2044-8295.2010.02004.x>
- van Prooijen, J.-W., & van Vugt, M. (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>