Confirmation Bias and Misconceptions: Pupillometric Evidence for a Confirmation Bias in Misconceptions Feedback

Willem W.A. Sleegers
Travis Proulx
Ilja van Beest

Author Note
Willem W. A. Sleegers, Department of Social Psychology, Tilburg University
Travis Proulx, Department of Psychology, Cardiff University
Ilja van Beest, Department of Social Psychology, Tilburg University

Correspondence concerning this article should be addressed to Willem W. A. Sleegers, Department of Social Psychology, Tilburg University, PO Box 90153, 5000 LE Tilburg, +31 13 466 8364. E-mail: w.w.a.sleegers@tilburguniversity.edu

Word count: 5136
Abstract

It has long been supposed that the confirmation bias plays a role in the prevalence and maintenance of misconceptions. However, this has been supported more by argument than by empirical evidence. In the present paper, we show how different types of belief-feedback evoke physiological responses consistent with the presence of a confirmation bias. Participants were presented with misconceptions and indicated whether they believed each misconception to be true or false, as well as how committed they were to the misconception. Each response was followed by feedback that was either clear (i.e., “correct” or “incorrect”) or ambiguous (i.e., “partly correct” or “partly incorrect”). Pupillary response to each feedback condition was assessed. The results show an interaction between feedback accuracy and feedback clarity on pupil size. The largest pupil size was found in response to clear disconfirmatory feedback. The smallest pupil size was found in response to both clear and ambiguous confirmatory feedback. Crucially, the pupil responded to ambiguous confirmatory feedback as though it were wholly confirmatory. Moreover, pupil size in response to ambiguous disconfirmatory feedback was significantly smaller than response to clear disconfirmatory feedback, showing an overall trend towards confirmatory processing in the absence of clear disconfirmation. Additionally, we show a moderation by commitment towards the misconception. The greater the commitment, the larger the effect of belief-violating feedback on pupil size. These findings support recent theorizing in the field of misconceptions and, more generally, the field of inconsistency-compensation.

Keywords: confirmation bias, pupillometry, misconceptions, error-feedback, inconsistency compensation
Confirmation Bias and Misconceptions: Pupillometric Evidence for a Confirmation Bias in Misconceptions Feedback

The confirmation bias is the tendency to search for, interpret, favor and recall information in a way that confirms one's preexisting beliefs or hypotheses. Various psychological tendencies can be considered illustrations of a confirmation bias. For instance, people may start out overconfident in an initial belief, fail to give proper consideration to alternative hypotheses, or interpret ambiguous information in favor of a held belief (Klayman, 1995). Moreover, the confirmation bias has been subject to much empirical work (for an overview, see Nickerson, 1998) and has been argued to impact societally important areas like policy making (Tuchman, 1984), medical care (Pines, 2006), false beliefs and the spread of misinformation (Hughes, Lyddy, & Lambe, 2013; Vicario et al., 2016), and even science itself (Kuhn, 1970; Polya, 1954).

In the present paper, we investigate whether a confirmation bias process can be observed in people’s arousal response to belief-feedback. We investigate the possibility that confirmation biases may occur rapidly in the context of prevalent yet incorrect beliefs, i.e., misconceptions.

Misconceptions and the confirmation bias

People are often confronted with information that explicitly disconfirms their beliefs. Research has found that even in the presence of explicit negative feedback, people’s beliefs may persist (Nickerson, 1998). Rather than changing their minds, people frequently question or explain away information that conflicts with pre-existing beliefs (Henrion & Fischhoff, 1986; Lord, Ross, & Lepper, 1979). This is more likely to occur when the belief-violating information is ambiguous (Risinger, Saks, Thompson & Rosenthal, 2002; Ross & Anderson, 1982) and when commitment towards the prior belief is high (Jonas, Schulz-Hardt, Frey, & Thelen, 2001).

One area in which such belief-persistence is particularly relevant is the area of misconceptions. Misconceptions refer to knowledge and beliefs that are incongruent with the core concepts and empirical findings of a discipline (Hamza & Wickman, 2008; Taylor & Kowalski,
Common misconceptions include the belief that humans use 10% of their brain, that the expected life span during the Middle Ages was 30, or that Vikings wore horned helmets in battle. Although these beliefs are incorrect, they remain common; with prevalence rates ranging from 28% to 71% (Lillienfeld et al., 2009).

Most of the research on misconceptions has been conducted in the field of educational psychology because of potential challenges that misconceptions pose for educators (Hughes, Lyddy, & Lambe, 2013). It has been speculated that misconceptions negatively impact the learning of new information (Chew, 2006; Hammer, 1996; Posner, Strike, Hewson, & Gertzog, 1982), resulting in worse academic performance (Kendeou & van den Broek, 2005). Besides negative consequences in education, misconceptions may also harm individuals in various domains of their lives, including romantic relationships (e.g., by believing that “opposites attract”), medical decisions (e.g., by believing that vitamin C protects against the common cold) or seeking help from authorities (e.g., by believing you have to wait more than 24 hours to file a missing persons report; van Maanen, 1994).

Researchers have also investigated how misconceptions can be reduced. Unfortunately, they have found that misconceptions are highly resistant to change. Exposure to educational courses were found to reduce some misconceptions but fail to eliminate many others (Gregg, Winer, Cottrell, Hedman, & Fournier, 2001; Lamal, 1995; Landau & Bavaria, 2003). For example, 30% of students believe that someone experiencing schizophrenia has a “split personality” even after completing psychology courses that teach this belief is false (Gardner & Dalsing, 1986). To explain this failure to reduce misconceptions, researchers have pointed at cognitive biases such as the confirmation bias (Lilienfeld, Lynn, Ruscio, & Beyerstein, 2010; Schick & Vaughn, 2014). Yet, as noted by Hughes, Lyddy, and Lambe (2013) in their review on the topic, these suggestions are limited due to lack of empirical investigation and “have generally been more supported by argument rather than empirical evidence” (p. 22).
In the present paper, we investigate whether a confirmation bias may indeed underlie, in part, the persistence of misconceptions. We propose that one method to demonstrate the relevance of the confirmation bias is by assessing physiological responses to belief-feedback. There are two main reasons why a physiological measure may be preferred over more traditional measures such as self-report: 1) physiological measures can potentially index psychological processes that are not reflected in self-reports due to limitations resulting from social desirability or the lack of introspective awareness and 2) physiological measures allow for more sensitive measures such as greater temporal resolution compared to self-report (Tomarken, 1995). In the context of responding to belief-feedback, we believe that social desirability may be an important factor because there are often specific norms as to how people should respond to belief feedback. Additionally, a more sensitive measure also enables us to detect quick, preconscious responses to belief-feedback, which may reveal differences between various types of belief-feedback that can be obscured by self-report measures.

**Physiological responses to belief-violating feedback**

Although humans appear to frequently demonstrate a confirmation bias, they are not immune to belief feedback. After all, it is functional to have, to some degree, an accurate representation of one’s environment. Consequently, one of the main functions of the central nervous system is to constantly compare sensory input from the environment to expectations derived from one’s beliefs. Inconsistencies between sensory input and one’s expectations indicate a need for belief-updating. Various brain areas have been implicated in this monitoring function, such as the anterior cingulate cortex and the posterior parietal cortex (O’Reilly et al., 2013). Based on this research, it appears that people do not update their beliefs in a fully rational, Bayesian, manner, and often do so insufficiently, giving rise to the confirmation bias (Charness & Dave, 2017).

The observation that people do not always rationally update their beliefs following negative belief-feedback lies at the core of recent models in the inconsistency-compensation literature (Jonas
et al., 2014; Proulx, Inzlicht & Harmon-Jones, 2012). According to these models, people construct mental models that enable them to predict and interpret subsequent experiences, providing them with a sense of understanding, even meaning (Proulx & Inzlicht, 2012). For instance, people frequently maintain the mental model that the world is just (Lerner, 1980). They believe good things happen to good people and bad things to bad people. Mental models such as the belief in a just world are not held passively. Once adopted, people are committed to these models, as seen in the defensive behaviors following just world belief violating events such as natural disasters, accidents, and criminal acts that harm innocent people. Rather than changing their belief in a just world, people often blame the victim in order to maintain their just world belief (van den Bos & Maas, 2009; van der Bruggen & Grubb, 2014).

The motivation for people to maintain their beliefs is further demonstrated by studies showing that inconsistencies result in a state of physiological arousal. For example, the experience of cognitive dissonance elicits increased skin conductance responses (e.g., Croyle & Cooper, 1983). Similarly, interacting with stereotype-inconsistent individuals has been found to produce aversive arousal as assessed with impedance cardiography (Mendes, Blascovich, Hunter, Lickel, & Jost, 2007; Townsend, Major, Sawyer, & Mendes, 2010). Even trivial perceptual inconsistencies such as reverse-colored playing cards produce sympathetic nervous system arousal, as assessed with pupillometry (Sleegers, Proulx, & van Beest, 2015). Consequently, it has been postulated that all experiences of inconsistency evoke a state of aversive arousal. The state of arousal in turn motivates compensatory behaviors aimed at reducing the aversive arousal, whether it is through a change in the belief or an alternative strategy (see Jonas et al., 2014; Proulx et al., 2012; Sleegers & Proulx, 2015).

Arousal and pupillometry

One increasing popular measure of arousal is pupillometry. Pupillometry (i.e., measuring changes in pupil size) is a popular tool to assess neuroaffective arousal (for reviews, see Beatty &
PUPILLARY RESPONSE TO FEEDBACK ABOUT MISCONCEPTIONS

Lucero-Wagoner, 2000; Sirois & Brisson, 2014). The relationship between pupil size and arousal stems from its association with the locus coeruleus-norepinephrine (LC-NE) system. This system plays a role in the regulation of engagement or withdrawal from a task by releasing NE through projections from the LC in the forebrain (for a review, see Aston-Jones & Cohen, 2005). Research has shown that pupil size correlates with LC activity in monkeys (Rajkowski et al., 1993) as well as in humans (Gilzenrat, Nieuwenhuis, Jepma, & Cohen, 2010; Murphy, O’Connell, O’Sullivan, Robertson & Balsters, 2014), and work by Beatty and colleagues on the pupillary system is consistent with LC responses to task-events (Beatty, 1982a, b; Richer & Beatty, 1978; see Aston-Jones & Cohen, 2005, for a discussion).

The link between pupil size and the LC-NE system allows researchers to infer a broad range of both cognitive and emotional processes from the extent of pupil dilation. Notably, the pupil dilates in response to task error (Brown et al., 1999; Critchley, Tang, Glaser, Butterworth, & Dolan, 2005) and violations of expectations (Preuschoff, ’t Hart & Einhäuser, 2011; Proulx, Sleegers & Tritt, 2017; Raisig, Welke, Hagendorf & van der Meer, 2010; Sleegers, Proulx & Van Beest, 2015). More specifically, Nassar et al. (2012) have shown that prediction error (i.e., unexpectedness) was positively correlated with pupil size and linked this response to each participant’s learning rate (i.e., belief-updating). They conclude that a relationship between arousal state and learning rate is likely a result of a coordinated learning-arousal network including the locus coeruleus and the ACC.

We propose that pupillometry affords a physiological demonstration of the confirmation bias. As prior research has shown, when events violate people’s beliefs, this should result in pupil dilation and a potential display of the confirmation bias, particularly when the feedback is ambiguous. When feedback is ambiguous, it can be implicitly interpreted as either confirming or violating the expectations that follow from prior beliefs. Given that people are motivated to maintain their prior beliefs, we propose that people are particularly likely to assimilate ambiguous feedback in such a way that it wholly confirms their existing beliefs. Consequently, this assimilation
should affect physiological arousal levels following feedback. If feedback is assimilated as confirmatory, the feedback should not evoke elevated levels of arousal, while if feedback is interpreted as disconfirmatory, elevated levels of arousal should follow. Additionally, these effects should be more pronounced the more one is committed to the belief. If one is highly committed to a belief, one has a stronger expectation to be correct and should thus be more likely to interpret feedback as confirmatory feedback.

**Hypotheses**

Using a 2 (feedback accuracy: correct/incorrect) x 2 (feedback clarity: clear/ambiguous) x 1 (commitment) design, we investigate the impact of clear and ambiguous feedback about misconceptions on the pupillary response, with an aim to gain empirical support for the role of a confirmation bias process. We hypothesize that when participants receive feedback about the veracity of their beliefs, they will show an increase in pupil dilation when the feedback is belief-violating (i.e., “incorrect”) relative to when the feedback is confirmatory (i.e., “correct”). More importantly, and indicative of a confirmation bias, we hypothesize that ambiguous feedback (i.e., “partly correct” and “partly incorrect”) will be assimilated as a confirmation of existing beliefs, thus evoking a pupillary response that is similar to the pupillary response following wholly confirmatory feedback (i.e., “correct”), and distinct from wholly disconfirmatory (i.e., incorrect). Finally, we hypothesize that relative commitment to the misconception will moderate the pupillary response, heightening both the increased (disconfirmation) and decreased (confirmation) arousal that corresponds to belief veracity feedback.

Note that we opted for manipulating the ambiguity of both the confirming and disconfirming feedback, rather than employing a design with only one type of ambiguity. Semantically, the two ambiguous options should be processed identically, but framing effects have been shown to be prevalent (e.g., Schwarz, 1999), hence we decided to include both options and test the exploratory
hypothesis that positive ambiguity is more likely to be processed as whole confirmatory feedback than negative ambiguity.

Method

Participants

We recruited a total of 51 undergraduate psychology students from Tilburg University who took part for course credit. They had an average age of 19.65 ($SD = 1.84$, min = 17, max = 24; 2 missing) and the majority of the participants were women (82.35%). Sample size was based on prior research1 (e.g., Bradley, Micelli, Escrig, & Lang, 2008; Laeng, Ørbo, Holmlund, & Miozzo, 2011, Partala & Surakka, 2003) and data was collected over the course of one week of reserved lab time. No additional data was collected after data analysis commenced.

Procedure and Design

Participants were welcomed into the lab and seated in standardized cubicles. The eye tracker was calibrated using the calibration extension from E-Prime 2 (Psychology Software Tools, Pittsburgh, PA). Hereafter, the misconceptions task began.

We employed a 2 x 2 x 1 within-subjects design to test the effect of feedback accuracy (correct/incorrect), feedback ambiguity (clear/ambiguous) and misconception commitment on pupillary reactivity. Participants were presented with misconceptions and asked to indicate whether they believed the misconception to be true or false. Following each misconception, but before they received the feedback, participants were asked several questions to assess their commitment towards the misconception. At the end, we assessed demographics, debriefed, and thanked participants for their participation.

-------------------

1 In the present paper we analyzed our results using linear mixed models. Calculating power for these types of analyses is challenging and unsupported in popular power calculation tools (e.g., G*Power), hence we based our sample size on previous studies in the literature.
Figure 1: Overview of a single trial. Durations are presented below each screen. Note that participants received three commitment questions, one by one, with no response window.

Materials

Task. The task contained 100 trials. Trials began with a fixation cross (Courier New, 30pt, black) that was displayed for 1000 ms on a gray background, followed by a misconception (Arial, 18pt, black). Participants could press the ‘1’ key to indicate that the misconception is correct or ‘2’ for incorrect, with a response window of 10 s. Next, participants answered three commitment questions about the misconception. Hereafter the words "Your answer is..." appeared. When the participant fixated on the words for 1000 ms, the words disappeared and the feedback appeared. The feedback was either clearly confirming (“correct”), clearly disconfirming (“incorrect”), ambiguously confirming (“partly correct”), or ambiguously disconfirming (“partly incorrect”). Participants received each feedback type equally often, so the feedback was not tied to their actual response. The four feedback options were presented in random order. The feedback was presented in Arial, 18pt, black font, in the center of the screen for a duration of 1500 ms. For an overview of a trial, see Figure 1.

Misconceptions. Misconceptions were collected from various sources, such as the Internet and books on the topic (e.g., Lilienfeld et al., 2010; van Maanen, 1994). A list of the misconceptions is available in Appendix A.

Commitment. To assess commitment, we asked three questions about each misconception: "How certain are you of your answer?", "How important is the statement to you?", and “How much
fun do you think the statement is?”. We included the question on fun because many of the misconceptions are often referred to as fun facts, indicating that while some misconceptions may be trivial, people may derive value from the misconceptions in terms of enjoyment. The misconceptions were referred to as statements (“stelling” in Dutch), in order to prevent the participants from realizing all statements were misconceptions. Each question consisted of a Likert scale that ranged from 1 (“Not at all”) to 7 (“Completely”). The commitment measure was created by averaging the responses on the certainty, importance, and fun measure together. Cronbach's $\alpha$ was calculated for each of the 100 misconceptions and was found to be acceptable ($M = 0.68$, $SD = 0.11$).

**Pupillometry.** Pupil data was collected using a Tobii T60 eye tracker (Tobii, Stockholm, Sweden). The Tobii T60 is integrated in a 17” TFT monitor and records at a rate of 60Hz. Each measurement has a validity indication that ranges from 0 (the system is certain that all data belongs to the particular eye) to 4 (gaze data is missing or incorrect). Only recordings with a validity score of 0 were used, resulting in 16.6% missing data during the relevant pupil response period (500-1500 ms). Pupil size from each eye were averaged together to create a single pupil size score. To address missing data, pupil measures were linearly interpolated using the “zoo” package (Zeileis & Grothendieck) and filtered with a modified repeated median filter (outer width: 25, inner width 15) using the “robfilter” package (Fried, Schettlinger, & Borowski, 2014). Hereafter, the pupil size was controlled for baseline differences by subtracting the average pupil size during a 500 ms fixation period before feedback presentation from the subsequent pupil measurements during feedback presentation (Beatty & Lucero-Wagoner, 2000). A 500 ms period after feedback presentation served as the light reflex period. Finally, an average pupil size was calculated over the 500-1500 ms time feedback period, resulting in an average pupil size change score for each trial.
Data analysis

Data was analyzed using R (R Core Team, 2018). All analyses consisted of linear multilevel models, using both a frequentist and Bayesian framework. For the frequentist analyses, the 'lme4' package (Bates, Maechler, Bolker, & Walker, 2015) was used in combination with ‘lmerTest’ (Kuznetsova, Brockhoff, & Christensen, 2017), which adds p-values and degrees of freedom for the t-test on the model parameters. For the Bayesian analyses, we used the ‘brms’ package (Bürkner, 2017). We defined random intercepts for each participant and each misconception in our models. Feedback accuracy (0 = correct; 1 = incorrect) and feedback ambiguity (0 = clear; 1 = ambiguous) were dummy coded and added as categorical predictors. Commitment was added as a (mean centered) covariate. Effects were inspected using additional contrast coding (e.g., changing the reference category). In the case of an interaction with commitment, both simple slopes were analyzed and the effect of feedback at both the minimum and maximum level of commitment. For the Bayesian analyses, we set normal distribution priors on the regression coefficients ($\mu = 0$, $\sigma = 1$). With a mean of 0 and a standard deviation of 1 we regularize the prior and allow enough variation for the data to inform the posterior.

Results

Response descriptives

Participants believed, on average, that the majority of the misconceptions were true ($M = 67.86\%, SD = 8.86$), despite all misconceptions being factually untrue. This indicates that 1) people’s beliefs were indeed misconceptions and 2) participants did not realize all statements were misconceptions and consequently did not adopt the response style of simply stating each statement was false.
Main analyses

**Frequentist analyses.** We conducted a linear mixed model with feedback accuracy (correct/incorrect), feedback ambiguity (clear/ambiguous), commitment (mean centered), and their interactions as predictors. We found a significant main effect of feedback accuracy, $t(4981.72) = 7.21, p < .001$ $b = 0.034, SE = 0.0047, 95\% CI [0.025, 0.043]$. Given our dummy coding, this main effect indicates that participants showed a larger pupil dilation following clear belief violating feedback than after clear belief confirming feedback (also see Table 1). There was no main effect of feedback clarity, $t(4979.96) = 0.38, p = .70, b = 0.0018, SE = 0.0047, 95\% CI [-0.0082, 0.011]$. Importantly, there was a significant interaction effect between feedback clarity and feedback accuracy, $t(4980.59) = -2.46, p = .014, b = -0.016, SE = 0.0067, 95\% CI [-0.030, -0.0035]$. As can be seen in Figure 2, participants did not differ in the amount of pupil dilation following either clear confirming feedback and ambiguous correct feedback, while ambiguous violating feedback resulted in greater pupil dilation compared to ambiguous confirming feedback, $t(4982.24) = 3.73, p < .001$, $b = 0.018, SE = 0.0047, 95\% CI [0.0077, 0.027]$, and clear confirming feedback, $t(4984.98) = 4.12, p < .001, b = 0.020, SE = 0.0047, 95\% CI [0.0096, 0.029]$. Additionally, we find that when the feedback violated people’s beliefs, there was less pupil dilation following ambiguous violating feedback than after clear violating feedback, $t(4984.77) = -3.09, p = 0.002, b = -0.015, SE = 0.0047, 95\% CI [-0.024, -0.0042].

Table 1: Descriptives of average pupil size change in response to each feedback condition.

<table>
<thead>
<tr>
<th>Feedback</th>
<th>Clarity</th>
<th>M</th>
<th>SD</th>
<th>SE</th>
<th>Min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>correct</td>
<td>clear</td>
<td>0.053</td>
<td>0.057</td>
<td>0.0080</td>
<td>-0.076</td>
<td>0.16</td>
</tr>
<tr>
<td>correct</td>
<td>ambiguous</td>
<td>0.054</td>
<td>0.053</td>
<td>0.0075</td>
<td>-0.056</td>
<td>0.18</td>
</tr>
<tr>
<td>incorrect</td>
<td>clear</td>
<td>0.087</td>
<td>0.071</td>
<td>0.010</td>
<td>-0.074</td>
<td>0.31</td>
</tr>
<tr>
<td>incorrect</td>
<td>ambiguous</td>
<td>0.072</td>
<td>0.058</td>
<td>0.0082</td>
<td>-0.050</td>
<td>0.24</td>
</tr>
</tbody>
</table>
There was a significant main effect of commitment, $t(4725.27) = -3.02, p = .0026, b = -0.0092, SE = 0.0030, 95\% CI [-0.015, -0.0025]$ and a significant interaction effect of commitment and feedback accuracy, $t(4988.67) = 3.68, p < .001, b = 0.016, SE = 0.0043, 95\% CI [0.0066, 0.0025]$, see Figure 3. The main effect indicates that when participants received clear confirming feedback, higher levels of commitment were associated with less pupil dilation. Additionally, we find that when participants received clearly violating feedback, higher levels of commitment was associated with more pupil dilation, $t(4704.92) = 2.20, p = .028, b = 0.0067, SE = 0.0030, 95\% CI [0.00061, 0.013]$. 

Figure 2: Average pupil dilation in response to feedback accuracy (correct/incorrect) and feedback clarity (clear/ambiguous), with commitment as a covariate. Error bars reflect 95\% confidence intervals.
We inspected the effect of clear feedback accuracy at varying levels of commitment. At the lowest level of commitment, there was no effect of feedback accuracy, $t(4987.21) = 0.74, p = 0.46$, $b = 0.0079, SE = 0.011, 95\% CI [-0.015, 0.028]$ while at the highest level there was an effect of feedback accuracy, $t(4987.11) = 5.36, p < .001, b = 0.066, SE = 0.012, 95\% CI [0.040, 0.092]$. Finally, we did not find an interaction effect of commitment and feedback ambiguity, $t(4988.08) = 0.79, p = .43, b = 0.0035, SE = 0.0044, 95\% CI [-0.0053, 0.0012]$, nor a three-way interaction effect, $t(4988.96) = -0.88, p = .38, b = -0.0054, SE = 0.0062, 95\% CI [-0.018, 0.0066].$\(^2\)

\(^2\) We also repeated our main analysis without commitment as a covariate and found the same significant main effect of feedback accuracy, $t(4985.59) = 7.17, p < .001, b = 0.034, SE = 0.0047, 95\% CI [0.024, 0.044]$, nonsignificant effect of feedback ambiguity, $t(4983.95) = 0.30, p = .76, b = 0.0014, SE = 0.0047, 95\% CI [-0.0073, 0.0098]$, and significant interaction effect, $t(4984.92) = -2.42, p = .015, b = -0.016, SE = 0.0067, 95\% CI [-0.029, -0.0024]$. 

Figure 3: Average pupil dilation in response to positive and negative feedback as a function of commitment. Error bars reflect 95\% confidence intervals.
Bayesian analyses. The previous analyses were repeated using a Bayesian framework. Consistent with the frequentist analyses, we find that participants showed a larger pupil dilation following belief violating feedback than after belief confirming feedback, $b = 0.034$, $SE = 0.0048$, 95% HDI [0.025, 0.043]. Similarly, results indicate no main effect of feedback clarity, as shown by a small effect and the 95% HDI overlapping with 0, $b = 0.0017$, $SE = 0.0048$, 95% HDI [-0.0077, 0.011].

The previously found interaction effect is also supported by the Bayesian analysis, $b = -0.016$, $SE = 0.0069$, 95% HDI [-0.030, -0.0032]. For an overview of these results, see Figure 4. Notably, these finding reiterate the results of the frequentist analyses, but importantly increase...
confidence that the pupillary response to partly correct feedback matches the pupillary response to wholly correct feedback—a conclusion that is more difficult to draw using the frequentist framework.

We similarly find that commitment affects the pupillary response following confirmatory feedback, \( b = -0.0092, \ SE = 0.0030, \ 95\% \ HDI \ [-0.015, \ -0.0033] \) and that commitment interacts with feedback accuracy, \( b = 0.016, \ SE = 0.0042, \ 95\% \ HDI \ [0.0076, \ 0.024] \) in a manner consistent with the frequentist analyses. Also consistent with the frequentist analysis, we do not find support for a three-way interaction between commitment, feedback accuracy, and feedback ambiguity, \( b = -0.0054, \ SE = 0.0060, \ 95\% \ HDI \ [-0.017, \ 0.0063] \).

Discussion

We conducted an experiment to investigate the physiological response to feedback about held misconceptions. By observing and manipulating both the accuracy and clarity of the feedback we demonstrated that feedback disconfirming beliefs (i.e., being mistaken) led to an increase in pupil size compared to feedback confirming beliefs (i.e., being correct). Crucially, when feedback was confirmatory but ambiguous (“partly correct”), the amount of pupil dilation was identical to that in response to wholly confirmatory feedback (“correct”). Moreover, when feedback was belief-violating but ambiguous (“partly incorrect”), this led to less pupil dilation compared to wholly violating feedback (“incorrect”), showing an overall trend towards confirmatory processing in the absence of clear disconfirmation.

Also, we found that the amount of arousal was moderated by commitment towards the misconception belief. At low levels of commitment, we found no effect of feedback on arousal, but at higher levels of commitment there was an increasingly larger difference in arousal between confirming and disconfirming feedback. At high commitment, disconfirmatory feedback violates valued expectations, resulting in more arousal. At low commitment, the absence of expectations about the veracity of the misconception excludes a violation of expectations, resulting in a reduced,
PUPILLARY RESPONSE TO FEEDBACK ABOUT MISCONCEPTIONS

if not absent, arousal response. Additionally, confirmatory feedback was associated with less
arousal at increasing levels of commitment, to the extent that committed expectations were being
confirmed. When the feedback was disconfirmatory, greater commitment was associated with
higher levels of arousal. Notably, we did not find support for a three-way interaction. This means
we did not find empirical evidence for the possibility that a confirmation bias is more likely when
commitment towards the misconception is high.

Theoretical implications

Confirmation bias is a ubiquitous phenomenon, with many studies demonstrating this
behavior, but there have not been any empirical investigations of a confirmation bias in response to
misconceptions, despite their presupposed relevance (Hughes, Lyddy, & Lambe, 2013). Our study
contributes to this literature by demonstrating a confirmation bias in response to feedback about
misconceptions, as assessed with a direct psychophysiological measure. It appears that people
interpret ambiguous feedback about their misconceptions in a manner consistent with clear belief-
confirming feedback when the feedback is positively framed, and less so when the ambiguous
feedback is negatively framed.

Our results provide additional support for inconsistency-compensation models (Jonas et al.,
2014; Proulx, Inzlicht, & Harmon-Jones, 2012). According to these models, people experience a
state of aversive arousal following the detection of inconsistencies, such as those following negative
belief-feedback. Numerous studies have shown that inconsistencies cause a state of aversive arousal
and that they are motivated to reduce this arousal (Proulx & Inzlicht, 2012; Sleegers & Proulx,
2015). The results of the present study show that participants show increased arousal in response to
negative belief-feedback, and predictably show a reduced response when provided with ambiguous
belief-feedback. These findings provide new avenues for future research. For example, the absence
of an arousal response following positively framed ambiguous feedback should make subsequent
compensation behaviors, such as belief-updating, unnecessary—a prediction that can be tested in future studies.

We found no effect of feedback ambiguity when the feedback was confirmatory, but when the ambiguous feedback was belief disconfirmatory, the feedback did result in greater pupil dilation relative to clear or ambiguous confirmatory feedback. This finding may be consistent with a negativity bias, the notion that negative stimuli have a greater effect on one's psychological state and processes than do neutral or positive stimuli. Myriad studies have demonstrated this bias (for reviews, see Baumeister et al., 2001; Cacioppo, Gardner, & Berntson, 1999; Peeters & Czapinski, 1990; Taylor, 1991). However, negativity bias fails to account for the general pattern of effects, insofar as ambiguous confirmatory feedback—which is partially disconfirmatory—did not produce the same magnitude of pupillary response as wholly disconfirmatory feedback. Moreover, ambiguous disconfirmatory feedback produced significantly less pupillary response than wholly disconfirmatory feedback, which could not have occurred if negativity were the dominant assimilation bias underlying feedback processing. Instead, a dominant negativity bias would have produced a pattern of effects whereby both ambiguous feedback responses would have significantly differed from the wholly confirmatory feedback, the converse of what we report. The fact that any feedback not wholly disconfirmatory produces relatively diminished pupillary response suggests a relatively dominant role for confirmation bias in belief-feedback.

Limitation

In the current study we used a 2 x 2 design to test the effect of feedback accuracy and feedback clarity. The consequence is that this introduced two versions of ambiguous feedback: partly correct, partly incorrect. The advantage of this design is that it allowed us to test whether both forms of ambiguity are equally likely to be assimilated. A potential limitation, however, is that we inadvertently may have made the ambiguous feedback less ambiguous by including the words ‘correct’ and ‘incorrect’ in the ambiguous feedback. This suggests that it may be possible to vary the
severity of ambiguity, and thus vary the impact of ambiguity on arousal. In fact, Holroyd, Hajcak, and Larsen (2006) demonstrated that wholly ambiguous feedback (non-informative performance feedback vs. positive or negative performance feedback) elicited heightened feedback error-related negativity to the same extent as negative feedback. Similarly, Hirsh and Inzlicht (2008) revealed individual differences in response to ambiguous performance feedback (a question mark vs. positive (+) or negative (-) feedback), with some individuals reacting more strongly to uncertain feedback than negative feedback (see also Gu, Ge, Huang, & Luo, 2010). Although beyond the scope of the present research, we suggest that these differences in findings could be due to variations in the level of ambiguity. Future research could investigate whether the level of ambiguity also affects whether people are more or less likely to display a confirmation bias. Our findings show that people may be more likely to display the confirmation bias when the ambiguous feedback contains positive framing.

Although pupillometry is a widely used technique to assess physiological arousal (e.g., see Beatty & Lucero-Wagner, 2000; Laeng, Sirois, & Gredebäck; 2011; Sirois & Brisson; 2014), it is limited in important ways. For instance, pupillometry cannot be used to assess whether the experienced arousal is aversive. Research has shown that pupil dilation occurs following both positive and negative stimuli (Bradley, Miccoli, Escrig & Lang, 2008). Hence, it cannot be concluded that participants felt bad after certain types of belief-feedback. In fact, the valence of inconsistencies remains a debated topic (see e.g., Kruglanski et al. 2018).

An additional limitation of pupillometry is that it may not be clear what exact psychological process is being assessed in reaction to the stimuli. Pupil responses have been linked to several psychological processes implicated in the LC-NE system, such as cognitive load, interest value, and emotional content (Sirois & Brisson, 2014). Based on pupil size alone, it cannot be ascertained what exact psychological process was responsible for the changes in pupil size. Nevertheless, the pattern of pupillary reactivity suggest that this process reacted to ambiguous stimuli as though it were
wholly or partially assimilated as confirmatory. In the present study, we assume that this pattern of reactivity required implicit interpretation of the feedback. Future research may specify which of these processes we are assessing with our pupillary measure, using different psychophysiological assessments.

Finally, we present only one study in support of a confirmation bias process underlying misconception feedback. Future work should aim to replicate the our findings and extend our work, for example by testing whether the reduced pupillary responses are also associated with less belief-updating and whether the arousal induced by negative belief-feedback is indeed experienced as aversive.

Conclusion

We presented participants with feedback about their misconceptions and observed that feedback violating beliefs elicits more physiological arousal, as measured through pupil dilation, than feedback confirming beliefs. We found evidence for a confirmation bias. Ambiguous confirmatory feedback was assimilated as wholly confirmatory feedback, while ambiguously belief-violating feedback appeared to be partially assimilated as confirmatory. This response was moderated by a commitment towards the misconception—no effect of feedback was found at the lowest levels of commitment and the largest effect of feedback was found at the highest levels of commitment. With this data we have empirical support for the role of a confirmation bias in response to feedback about misconceptions.


PUPILLARY RESPONSE TO FEEDBACK ABOUT MISCONCEPTIONS


