



HAL
open science

A common factor underlying individual differences in confirmation bias

Vincent Berthet, Predrag Teovanović, Vincent de Gardelle

► **To cite this version:**

Vincent Berthet, Predrag Teovanović, Vincent de Gardelle. A common factor underlying individual differences in confirmation bias. *Scientific Reports*, 2024, 14 (1), pp.27795. 10.1038/s41598-024-78053-7. hal-04839711

HAL Id: hal-04839711

<https://cnrs.hal.science/hal-04839711v1>

Submitted on 16 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



OPEN

A common factor underlying individual differences in confirmation bias

Vincent Berthet^{1,5}✉, Predrag Teovanović^{2,3} & Vincent de Gardelle⁴

When they are asked to test a given hypothesis, individuals tend to be biased towards confirming evidence. This phenomenon has been documented on different cognitive components: information search, weighing of evidence, and memory recall. However, the interpretation of these observations has been debated, and it remains unclear whether they truly reflect a confirmation bias (as opposed to e.g., a bias towards positive information). In the present study we aimed at evaluating whether these biases might be subtended by a common factor. We adapted three classic experimental paradigms on hypothesis testing (Wason selection task, 2-4-6 task, and interviewee task) and examined the relation between these biases using an individual differences approach. Participants (N = 200) completed a total of nine behavioral tasks, in which each component of confirmation bias was measured in each of the three experimental paradigms. Correlations and factor analyses within a multitrait–multimethod framework indicated greater convergence of bias scores within each component across paradigms, than within experimental paradigms. This suggests that a common factor underlies the different measurements of confirmation bias across experimental paradigms, at least to some extent. In these paradigms, thus, biases towards confirming evidence may truly reflect a confirmation bias.

Keywords Confirmation bias, Hypothesis testing, Individual differences, Measurement

“I had also, during many years, followed a golden rule, namely, that whenever a published fact, a new observation or thought came across me, which was opposed to my general results, to make a memorandum of it without fail and at once; for I had found by experience that such facts and thoughts were far more apt to escape from the memory than favourable ones. Owing to this habit, very few objections were raised against my views which I had not at least noticed and attempted to answer.”

Charles Darwin, *The Autobiography of Charles Darwin*

Confirmation bias is commonly defined as the fact that “information is searched, interpreted, and remembered in such a way that it systematically impedes the possibility that the hypothesis will be rejected”¹. This bias is widespread and takes different forms^{1,2}, and recent research has highlighted in particular its potentially large societal impact. When applied to political beliefs for instance, confirmation bias may contribute to the persistence of misinformation and conspiracy beliefs^{4,5} and political polarization^{6,7}. Yet, and although confirmation bias has been described as “the best known and most widely accepted notion of inferential error to come out of the literature on human reasoning”⁸, it is still unclear how different manifestations of confirmation bias are related to each other.

Furthermore, one should distinguish between two aspects of confirmation bias commonly studied in the literature, which differ regarding the source of the hypotheses to test (internal vs. external), and their relevance to the participant’s values, beliefs and self-concept. On the one hand, confirmation bias can refer to the tendency to support one’s own prior views on a particular topic, for instance, death penalty⁶. In this case, it is sometimes referred to as a *myside bias*^{9–11}. On the other hand, confirmation bias was initially invoked to account for the participants’ tendency to confirm abstract/logical hypotheses that are not specifically related to their beliefs or values^{12,13}. This was the case, for instance, in the pioneering work of Wason^{14,15}, described below.

¹Psychology and Neuroscience Lab (2LPN), University of Lorraine, 23 boulevard Albert 1er, F-54000 Nancy, France. ²Faculty for Special Education and Rehabilitation, University of Belgrade, Belgrade, Serbia. ³Department of Psychology, Laboratory for Research of Individual Differences, Faculty of Philosophy, University of Belgrade, Belgrade, Serbia. ⁴French National Centre for Scientific Research, Paris School of Economics, Paris, France. ⁵Centre d’Economie de la Sorbonne, French National Centre for Scientific Research, Paris, France. ✉email: vincent.berthet@univ-lorraine.fr

The present study will only focus on this latter aspect of confirmation bias, with external hypotheses, rather than on myside bias. We have already noted that despite the apparent simplicity of the phenomenon, Oswald & Grosjean's definition implies that confirmation bias is an umbrella term that encompasses at least three related though distinct stages (seeking, interpreting, and remembering) of information processing¹⁻³. Our goal is to better understand whether these different manifestations are distinct phenomena or whether they can be considered as three facets of a single phenomenon.

The concept of confirmation bias in hypothesis testing was introduced by Wason¹⁴ who designed a rule discovery task known as the 2-4-6 task. Participants were provided with three numbers (e.g., “2-4-6”) as an instance of a numerical rule, which they had to discover. For that purpose, they were instructed to put forward triples and the experimenter would tell them whether or not each triple complied with the rule. It turned out that participants usually tested triples that were compatible with the hypothesis they had in mind (e.g., the triple “10-12-14” to test the hypothesis that the rule at hand is “add 2”), suggesting that they aimed to confirm their hypothesis rather than to falsify it. The four-card problem (or selection) introduced by Wason¹⁵ is also typically used to demonstrate the tendency of participants to search for confirmatory information when testing logical (if-then) rules¹⁶. A similar finding was found using a trait hypothesis testing task in which participants are put in the position of interviewers and select questions to ask an interviewee to test a hypothesis about her personality^{17,18}. For instance, to test the hypothesis that the interviewee is extraverted, participants selected significantly more questions assuming that it was true (e.g., “What kind of situations do you seek out if you want to meet new people?”) than questions assuming that the alternative, introversion hypothesis, was true.

These three experimental paradigms have been most influential in the confirmation bias literature. The card selection task introduced by Wason^{15,19} has given rise to hundreds of publications and received thousands of citations. Regarding the 2-4-6 task, Evans²⁰ highlighted that Wason's paper¹⁴ “fully deserves the many hundreds of citations it has received, which continue unabated to the present day.” (p. 1). Similarly, the interviewee task has been widely discussed in debates on confirmation bias^{1,10}.

There has been considerable debate in the literature about whether the pattern observed in these tasks demonstrates a true confirmation strategy in hypothesis testing^{1,8,10,20}. For instance, behavior in the Wason task might reflect a matching bias by which participants tend to choose whichever cards happen to be named in the rule²¹. A positive test strategy^{3,22} or a congruence bias²³ have been also suggested, by which participants would ask only questions for which they expect a positive answer. As noted by Evans²⁰, in the case of the classic 2-4-6 task, this positive test strategy fully overlaps with a confirmation strategy, as it would lead participants to only suggest sequences that would be confirmed by the experimenter. A diagnostic strategy, by which participants would ask questions that most clearly distinguish between the given hypothesis and its alternative(s) has also been discussed²⁴⁻²⁶.

While confirmation bias has been mainly studied with regard to information search using the different paradigms described above, it may also occur in the processes of weighing evidence and memory recall¹. Several studies have evidenced that participants evaluated information incongruent with their attitude much more critically than congruent information⁶, and retained their hypothesis when faced with ambiguous evidence²⁷ or even clearly falsifying evidence²⁸. With regard to memory recall, participants have a bias towards reporting information as previously encountered for confirming more than disconfirming information, whereas sensitivity measures of memory show no clear effect^{29,30}.

So far, most studies focused on only one component of information processing and only one task/paradigm. One notable exception is the study by Vedejová and Čavojevová³¹ who designed a unified procedure in which the three components of confirmation bias (i.e., information search, weighing of evidence, and memory recall) were evaluated within the same paradigm. However, this study has focused only on myside bias, and reported average effects for each component, without assessing score reliability and individual differences. Thus, how the different components of confirmation bias in hypothesis testing may relate to each other remains unclear. Within the variety of tasks associated with the phenomenon of confirmation bias, here we focus on tasks where hypotheses to be tested are given by the experimenter and not related to the identity or beliefs of participants.

Our main goal is to document how these tasks might relate to one another, and whether they could rely on a common underlying process. To do so, we use an individual differences approach and a multitrait-multimethod framework (MTMM)³². It is worth noting that few studies on confirmation bias followed an individual differences approach and most of them focused on myside bias³³. Rassin³⁴ introduced a self-report measure of confirmation proneness; however, composite scores from various single item behavioral tasks (e.g., 2-4-6, interviewee) used as a criterion showed poor reliability.

We measure in each individual the three different cognitive components of confirmation bias described in prior research (i.e., information search, weighing of evidence, and memory recall) using the three different experimental paradigms described above (Wason selection task, 2-4-6 task, and interviewee task). We expect that different measures of the same component would be correlated across paradigms, indicating that this cognitive component can be robustly estimated at the individual level using different methods. The correlation of confirmation bias scores across components, on the other hand, will indicate the extent to which these components should be treated as distinct aspects of confirmation bias or whether this phenomenon can be considered unitary. Finally, we investigate how these aspects of confirmation bias relate to relevant societal issues using individual measures of dogmatism, pseudo-scientific beliefs, and conspiracy beliefs.

Methods

Participants

A total of 200 participants (83 women) were recruited by the Fédération S2CH in France (CNRS and Paris 1 Panthéon-Sorbonne University) to complete a one-hour online session. Each participant received compensation of €10. The mean age was 27.08 years ($SD = 4.73$). Participants in our study were quite educated as only 4.5%

were not higher education graduates (59.5% of the participants reported having a master's degree). This research was reviewed and approved by the Institutional Review Board–Paris School of Economics (approval number: 2020–022). All experiments were performed in accordance with relevant guidelines and regulations. Participants gave their informed consent before taking part in the study.

In a prior study in which we evaluated correlations between several cognitive bias scores³⁵, we found a correlation of $r=0.28$ between confirmation bias and outcome bias. It can be reasonably expected that the correlations between various confirmation bias scores in the present study will be at least of this magnitude. A power analysis, conducted with G*Power 3.1.9.7 indicated that detecting such a correlation (in a correlation bivariate normal model) with 5% error probability and 80% power would require 77 participants. Our sample size ($N=200$) largely exceeds this number, it is also much higher than the sample size typically used in prior studies, for instance, with $N=18$ ²⁸, or with $N=44$ (study 1) and $N=80$ (study 2)³⁶, and it is consistent with a prior study by Vedejová and Čavojová³¹ examining different confirmation bias tasks (with $N=199$).

Confirmation bias tasks and measures

All tasks included confirming and disconfirming information, and some of them also included neutral information as filler items. Information type (confirming vs. disconfirming vs. neutral) was always manipulated as a within-subjects factor. Confirmation bias scores were always calculated as the mean response (typically over 3 or 4 items) to confirming information minus the mean response to disconfirming information, and could take values between -1 (disconfirmation bias) and 1 (full confirmation bias). The absence of neutral information in some tasks was either due to logical constraints (e.g., in information search, the numerical rule hypothesis testing task does not allow for neutral options) or practical reasons (e.g., to shorten the memory recall tasks). We report below for each task the instructions, one example of a rule/hypothesis that participants had to test, and the scoring rule. Full details of each task (including all stimuli) are included in the supplementary material.

Information search

Logical rule hypothesis testing task As in the original task¹⁵, participants were presented with four cards and with a conditional statement of “if P then Q” type. For example, one item featured the four cards “D”, “7”, “5”, and “K” together with the rule “If a card has a D on one side, then it has a 5 on the other side”. Participants were instructed to indicate which of the four cards they would turn in order to test this statement. Similar to Stanovich et al.³⁷, we used a variant in which participants reported whether or not they would turn over each card (Yes = 1 vs. No = 0). The task was repeated four times, with conditional statements from slightly different contexts (none of which included a deontic rule, i.e., involving actions forbidden or allowed, obligatory or not obligatory). For each item, confirmation bias was scored as the difference between the response to the confirming card (Q, i.e., card “5” in the example) and the response to the disconfirming card (not-Q, i.e., card “7”). Note that the P card (card “D”) was not included in the scoring as it is both confirming and disconfirming²⁸.

Numerical rule hypothesis testing task Similar to the original task¹⁴, participants were provided with the following instruction: “You start a game in which your opponent has set a rule generating sequences of three numbers. Your goal is to find out this rule. For that purpose, you put forward sequences of numbers and your opponent tells you whether or not each sequence complies with the rule”. Participants were presented with a valid sequence and with the hypothesis at hand (unlike in the original version of Wason). For instance, for one item this was: “Your opponent starts by telling you that the sequence 2-4-6 complies with the rule. You think that ‘add 2’ is the rule”. They were also presented with two candidate sequences, one confirming (e.g., “8-10-12”) and one disconfirming (e.g., “3-6-9”), and they were asked to report whether or not they would put forward each one (Yes = 1 vs. No = 0). The task was repeated three times with different rules and sequences. Confirmation bias was calculated as the mean difference between the response to the confirming sequence and the response to the disconfirming sequence.

Trait hypothesis testing task We used the adapted version of the interviewee's personality task (Snyder & Swann, 1978) introduced in our prior work^{35,38}. Participants were given a hypothesis regarding a personality attribute of a candidate such as “the candidate is extraverted” and they were instructed to test this hypothesis by selecting 4 questions among a set of 10 questions to ask during an interview. In this set, 4 questions were confirming (e.g., “What events make you feel popular with people?”), 4 were disconfirming (e.g., “What kind of events make you feel like being alone?”), and 2 were neutral (e.g., “What kind of charities do you like to contribute to?”). This task was repeated four times, each time with a different personality attribute (agreeableness, conscientiousness, emotional stability, extraversion). For each item, the confirmation bias score was calculated as the difference between the percentage of confirming questions selected and the percentage of disconfirming questions selected.

Weighing of evidence

In the WE tasks, participants were presented with the same contexts as used in the previous set of tasks, except that instead of selecting one option (e.g., one card to flip, or one sequence to submit to a test), here they were presented with a set of options already selected and the associated outcomes (e.g., the card has a 7 on the other side), and they were asked to indicate whether each outcome was informative or not to test a given hypothesis.

Logical rule hypothesis testing task In the Wason selection task, when testing a logical rule of the form “if P on one side then Q on the other side”, turning over each of the four cards (P, not-P, Q, not-Q) could produce three potential outcomes: confirming, disconfirming, or neutral. For instance, turning over a P card produces a confirming outcome if Q is observed and a disconfirming outcome otherwise. Turning over a Q card could produce

a confirming outcome if P is observed and a neutral outcome otherwise. Turning over a not-Q card produces a disconfirming outcome if P is observed and a neutral outcome otherwise. Finally, turning over a not-P card always produces a neutral outcome. On each of 4 trials, participants were presented with six cards (e.g., E, C, 5, A, 4, and 9) and they were given a rule to test (e.g., “If a card has a vowel on one side, then it has an even number on the other side”). They were then presented with the outcomes observed when turning over each card, and asked to indicate whether the outcome was informative or not (Informative = 1, Uninformative = 0) with regard to the rule to test. We excluded the two neutral outcomes (“not-P → Q”, and “Q → not-P”). We used four items (none of which included a deontic rule). In each item, confirmation bias was calculated as the difference between the mean response to the two confirming outcomes (e.g., “P → Q” and “Q → P”) and the mean response to the two disconfirming outcomes (e.g., “P → not-Q” and “not-Q → P”).

Numerical rule hypothesis testing task In the 2-4-6 task, participants have to put forward sequences of numbers in order to test a hypothesis given to them by the experimenter regarding the rule used by their opponent. For instance, one of the rules was that each card would “add 2” to the preceding. There are four possible outcomes, depending on whether the sequence proposed is consistent with the hypothesis, and whether it complies with the rule used by the opponent. For instance, if the current hypothesis is “add 1”, the sequence “1-2-3” would be consistent and the sequence “8-5-2” inconsistent with the hypothesis. Participants were presented with these four outcomes and were asked to indicate whether each outcome was informative or not (Informative = 1, Uninformative = 0) with regard to the hypothesis to test. The task was repeated three times, with different items. In each item, confirmation bias was calculated as the difference between the mean response to the confirming outcome (e.g., if the hypothesis is “add 1”, the outcome “1-2-3 → complies” is confirming) and the mean response to the two disconfirming outcomes (e.g., “3-4-5 → does not comply” and “3-6-9 → complies” if the hypothesis is “add 1”). Neutral outcomes (e.g., “3-6-9 → does not comply” is neutral with respect to the hypothesis “add 1”) are not used in the scoring.

Trait hypothesis testing task Participants were provided with the following instructions: “Imagine that you work as a human resources recruiter for a large company. At the end of the interviews, the candidate took a personality test (a set of statements to which she answered TRUE or FALSE).” They were presented with six items of the test and the answers of the candidate, and they were asked to indicate whether each outcome was informative or not (Informative = 1, Uninformative = 0) with regard to the hypothesis that the candidate has a given personality attribute. For instance, one hypothesis was that “the candidate is an agreeable person” (other hypotheses mentioned a different personality trait). Among these six outcomes, two were confirming (e.g., “I always try to stay in a good mood → TRUE”), two were disconfirming (e.g., “I am not irritated by anyone → FALSE”), and two were neutral (e.g., “I am a subscriber to a magazine → FALSE”). Each type of outcome included a “TRUE” answer and a “FALSE” answer. Four items were used. In each item, confirmation bias was calculated as the difference between the mean response to the two confirming outcomes and the mean response to the two disconfirming outcomes.

Memory recall

Memory recall tasks were based on the same material as the previous tasks. Here, participants were presented with several outcomes, some of which were actually presented in the corresponding weighing of evidence task that they had done before, while the others were new. Participants had to indicate whether each outcome had been presented before or not. Confirmation bias was measured regardless of accuracy, as the tendency to rate confirming information as encountered previously more often than disconfirming information.

Logical rule hypothesis testing task Participants were provided with the following instructions: “Previously, we showed you the outcomes observed after turning the cards over and we asked you to indicate whether each outcome was informative or not to test a given rule”. Participants were then presented with eight outcomes and asked to indicate whether each outcome was actually shown before or not (Yes = 1 vs. No = 0). Four of them were actually presented, among which two confirming (P → Q and Q → P) and two disconfirming (P → not-Q and not-Q → P), while four were completely new (two confirming and two disconfirming). This task was repeated 4 times with items corresponding to the ones used in the WE version of the task. In each item, confirmation bias score was calculated as the difference between the mean response to the four confirming outcomes (two actual, two new) and the mean response to the four disconfirming outcomes (two actual, two new).

Numerical rule hypothesis testing task Participants were provided with the following instructions: “Previously, we showed you the opponent’s answers to various sequences of numbers (complies with the rule or not) and we asked you to indicate whether each outcome was informative or not with regard to the hypothesis to test”. Participants were then presented with eight outcomes and asked to indicate whether each outcome was actually shown before or not (Yes = 1 vs. No = 0). Four outcomes were actually presented before (one confirming, two disconfirming, and one neutral) and 4 were completely new (with again one confirming, two disconfirming, one neutral). The task was repeated four times, with different items. In each item, confirmation bias was calculated as the difference between the mean response to the two confirming outcomes (one actual, one new) and the mean response to the four disconfirming outcomes (two actual, two new).

Trait hypothesis testing task Participants were provided with the following instructions: “Previously, we showed you the answers of a candidate to several questions of a personality test, and we asked you to indicate whether each outcome was informative or not to test a hypothesis regarding a given personality attribute”. After that, participants were presented with eight outcomes and asked to indicate whether each outcome was actually

presented before or not (Yes = 1 vs. No = 0). Four of them were actually presented, among which two confirming and two disconfirming, while four were new (two confirming and two disconfirming). Again, the task was repeated four times, with different items. In each item, confirmation bias was calculated as the difference between the mean response to the four confirming outcomes (two actual, two new) and the mean response to the four disconfirming outcomes (two actual, two new).

Procedure

After providing consent, participants completed the tasks in the following order: information search, weighing of evidence, memory recall. Within each block, participants were first presented with the logical rule hypothesis testing task, then the numerical rule hypothesis testing task, and finally trait hypothesis testing task. Participants eventually completed the 7-item³⁹ and multiple-choice version of the Cognitive Reflexion Test (CRT MCQ-4⁴⁰), and the Decision Styles Scale (DSS)⁴¹ but this latter measure was not included in the analysis.

Follow-up session

To control for cognitive abilities and evaluate the relations between our measures of confirmation bias and socially relevant issues, we contacted the same participants again, to complete an additional online session. In this session, we used the matrix (11 items) and verbal (14 items) reasoning tasks from the International Cognitive Ability Resource⁴², as well as 3 scales measuring dogmatism (DOG scale)⁴³, pseudoscientific beliefs (short version of the PSEUDO scale)^{44,45}, and conspiracy beliefs (single-item scale from Lantian et al⁴⁶). We also collected participants' socioeconomic status: 37.38% of the participants in the subsample reported managerial and professional occupations; 20.56% were employed; and 27.10% were students.

The DOG scale⁴³ included 20 items (e.g., "Anyone who is honestly and truly seeking the truth will end up believing what I believe") to which participants responded using a 7-point Likert scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). The PSEUDO scale⁴⁴ included 8 items (e.g., "While it is true that evolution is a fact, there are issues that require an intelligent intervention to be explained") to which participants responded using a 5-point Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). The conspiracy beliefs scale⁴⁶ included a single item ("I think that the official version of the events given by the authorities very often hides the truth") to which participants responded using a 9-point Likert scale ranging from 1 (*completely false*) to 9 (*completely true*).

One hundred and seven participants (mean age: 29.27 years, 39% women) from the original sample completed this follow-up session.

Analysis

We investigated the convergent and discriminant properties of these measures, within a multitrait-multimethod framework³² in which a set of t traits are each measured by m methods. In the Campbell-Fiske approach to MTMM analysis, one would expect: (a) the measures of the same trait to converge across methods (i.e., convergent validity is high when the mean monotrait-heteromethod correlation is high); and (b) the measures of different traits using the same method to diverge (i.e., discriminant validity is high when the mean heterotrait-monomethod is low). In our case, we used the three aforementioned paradigms as methods, and the three components of information processing (information search, weighing of evidence, memory recall) as traits (note that within this MTMM approach, the reverse framework was possible but conceptually less relevant. In fact, while information search, weighing of evidence, memory recall could be defined as different methods/paradigms, rule and trait hypothesis testing could barely be described as psychological traits). Accordingly, we expected the measures of the same component to be correlated across paradigms (e.g., the measures of confirmation bias in information search should be correlated whether participants are asked to test a logical or a numerical hypothesis) while we aimed to find out the extent to which the three components within the same paradigm were correlated.

Regarding the analysis of convergent and discriminant validity, we applied confirmatory factor analysis (CFA)^{47,48} on the correlation matrix, following the correlated trait-correlated uniqueness (CTCU) model. The CTCU model avoids the common estimation and convergence problems arising with the correlated trait-correlated method (CTCM) model^{49–51}. The main difference between the CTCU and the CTCM models is the estimation of method effects⁵². In the CTCM model, these effects are assessed by specifying method factors while in the CTCU model, they are estimated by adding correlated uniquenesses (errors) among the indicators of a same method. As each uniqueness term reflects a combination of random measurement error variance and systematic method variance components, the correlations between uniquenesses within each method represent the variance common to that method⁴⁹. The CTCU model was implemented using the *lavaan* package in R.

Results

Descriptive statistics and reliability

Figure 1 shows the mean of participants' responses (aggregated across items) within each type of information (confirming, disconfirming, neutral), and Table 1 also indicates the reliability of scores in each task.

We first investigated confirmation bias at the group level by analyzing the data in a 2 (Component: information search, weighing of evidence, memory recall) × 2 (Paradigm: logical rule hypothesis testing, numerical rule hypothesis testing, trait hypothesis testing) × 2 (Type of information: confirming, disconfirming) ANOVA with repeated measures. Crucially, this analysis yielded a significant main effect of Type of information, $F(1, 198) = 158.12, p < 0.001$, participants being more prone to select the confirming options ($M = 69.1\%$, $SD = 13.6$) than the disconfirming ones ($M = 48.2\%$, $SD = 15.1$) (Cohen's $d = 0.89$) across all tasks. Type of information and Paradigm interacted significantly, $F(2, 396) = 63.49, p < 0.001$. The effect of Type of information was actually larger in the logical rule hypothesis testing ($d = 0.72$) and the numerical rule hypothesis testing ($d = 0.91$) paradigms

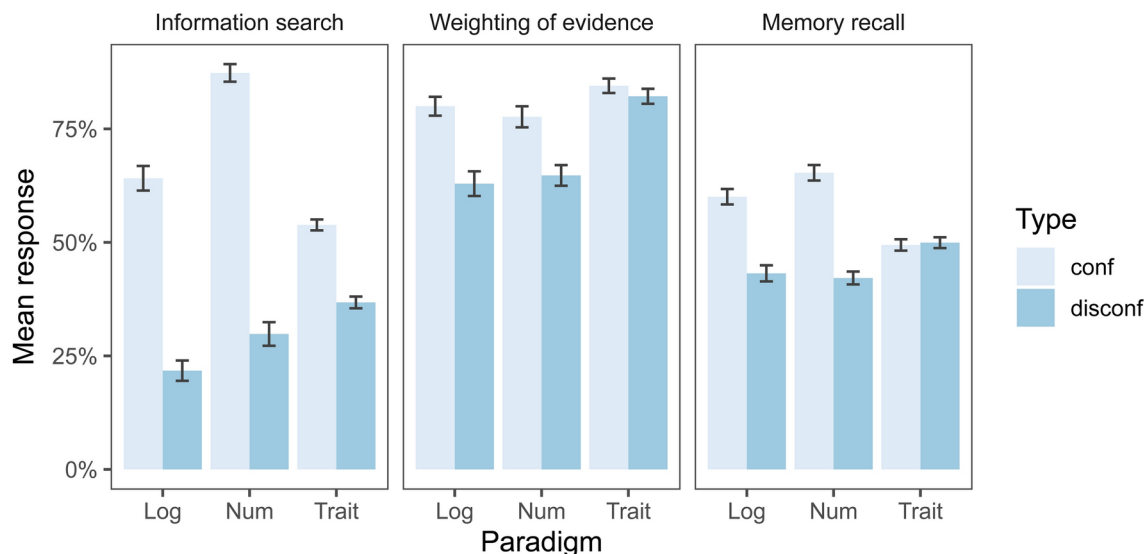


Fig. 1. Mean participants' responses (Yes = 1, No = 0) as a function of Component, Paradigm, and Type of information. Error bars indicate standard errors.

Task	Number of items	Confirming information	Disconfirming information	Neutral information	Confirming - Disconfirming			
		Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	<i>t</i> (199)	Cohen's <i>d</i>	Cronbach's α
<i>Logical rule hypothesis testing</i>								
IS	4	64.1 (38.3)	21.8 (31.5)	15.8 (23.7)	42.4 (55.0)	10.90, $p < .001$	0.77	.84
WE	4	80.0 (29.4)	62.9 (38.5)	23.8 (32.7)	17.1 (51.7)	4.66, $p < .001$	0.33	.94
MR	4	60.0 (24.0)	43.2 (25.1)	-	16.9 (40.6)	5.88, $p < .001$	0.42	.86
<i>Numerical rule hypothesis testing</i>								
IS	3	87.3 (27.3)	29.8 (36.7)	-	57.5 (55.5)	14.65, $p < .001$	1.04	.75
WE	3	77.7 (32.8)	64.8 (32.1)	22.8 (35.0)	12.9 (48.4)	3.78, $p < .001$	0.27	.81
MR	3	65.3 (24.1)	42.2 (20.1)	-	23.2 (31.6)	10.38, $p < .001$	0.72	.55
<i>Trait hypothesis testing</i>								
IS	4	53.8 (17.1)	36.8 (18.3)	19.6 (19.3)	17.1 (33.3)	7.26, $p < .001$	0.51	.64
WE	4	84.5 (22.5)	82.2 (23.4)	21.6 (23.9)	2.30 (26.5)	1.23, $p = .22$	0.09	.69
MR	4	49.4 (17.7)	49.9 (16.8)	-	-0.5 (18.6)	-0.40, $p = .70$	-0.02	.48

Table 1. Mean and standard deviation of participants' responses (in percent) to confirming, disconfirming and neutral options in the nine confirmation bias tasks (N = 200). For all tasks, the observed range of confirmation bias scores was always between -100 and 100 across participants.

than in the trait hypothesis testing paradigm ($d = 0.36$). In fact, the latter did not produce any confirmation bias regarding weighing of evidence and memory recall. Moreover, Type of information and Component interacted significantly, $F(2, 396) = 85.38, p < 0.001$. Indeed, the effect of Type of information was larger in information search ($d = 1.18$) than in weighing of evidence ($d = 0.31$) and memory recall ($d = 0.59$). This may reflect the fact that these three paradigms were originally designed to investigate confirmation bias in information search.

Overall, confirmation bias appears as a general phenomenon at the group level as it can be observed in three cognitive components through various tasks. Note that adding gender as covariate did not affect these results. In additional analyses, we also verified that confirmation bias scores in our data cannot be explained by the presence of negations in disconfirmatory statements (see Supplementary Material).

Second, we examined the internal consistency of confirmation bias scores, using Cronbach's α coefficient. Six out of the nine measures reached an adequate level of reliability (see Table 1). The three measures showing the lowest internal consistency were the trait hypothesis testing paradigm for the measurement of confirmation bias in information search ($\alpha = 0.64$) and memory recall ($\alpha = 0.48$), and the numerical rule hypothesis testing paradigm for the measurement of confirmation bias in memory recall ($\alpha = 0.55$). This finding suggests that the adapted versions of the tasks introduced produce reliable scores despite a low number of items (three or four). This preliminary check of score reliability confirmed that subsequent analyses based on individual differences in confirmation bias scores could be reasonably conducted.

CFA–MTMM analyses

We then analyzed the structure of correlations between the different confirmation bias scores, using a multitrait-multimethod approach, where the three components were used as traits and the three paradigms were used as methods. The complete 9×9 matrix of pairwise correlations between all tests is shown in Table 2. Following the analytic approach originally introduced by Campbell and Fiske³², we calculated the mean monotrait-heteromethod correlation r_1 , mean heterotrait-monomethod correlation r_2 , and mean heterotrait-heteromethod correlation r_3 . According to this approach, r_1 is a measure of convergent validity while r_2 is a measure of discriminant validity, therefore the expected pattern is $r_1 > r_2 > r_3$. In our data, the mean correlations exhibited the expected order (mean $r_1 = 0.32 > \text{mean } r_2 = 0.27 > \text{mean } r_3 = 0.18$). The average r_1 indicated moderate convergent validity overall, with lowest values for information search (pairwise correlations of 0.13, 0.16 and 0.30 across tests) and highest values for weighing of evidence (with pairwise correlations of 0.34, 0.42, and 0.65). The average heterotrait-monomethod r_2 indicated also that the different traits, as measured by the same method, were substantially correlated. These two correlation coefficients exceeded the average heterotrait-heteromethod correlation (mean $r_3 = 0.18$).

Unsurprisingly, the 2-4-6 and interviewee tasks were the most correlated ($r = 0.30, p < 0.001$), confirming that both reflect a positive test strategy, by which participants tend to select options for which a yes answer would confirm the hypothesis tested⁴². However, scores on the 2-4-6 and the Wason selection tasks were also significantly correlated ($r = 0.16, p < 0.001$), even though performance in the latter could barely reflect a positive test strategy as no Yes/No answer was expected and, importantly, the scoring method used was designed to capture confirmation bias only. The evidence of convergent validity suggests that the three measures of information search share common, though limited variance reflecting confirmation bias.

To further investigate the convergent and discriminant validity, we applied CFA on the correlation matrix, following the CTCU model. Note that our $N/pratio$ ($200/9 = 22.2$) was above the recommended values of 20⁵³, which suggested that we had enough power to conduct CFA on our data. With regard to the cut-off criteria recommended by Hu and Bentler⁵⁴ (namely, CFI > 0.95 , RMSEA < 0.06 , SRMR < 0.08), the CTCU model yielded an excellent fit to the data, $\chi^2(15) = 21.76, p = 0.11$, CFI = 0.981, RMSEA = 0.047, SRMR = 0.041. The standardized parameter estimates of the CTCU model are presented in Table 3 (see also Fig. 2A). Consistent with the analysis of the MTMM correlation matrix, the three measures displayed moderate to good convergent validity as the trait factor loadings were all substantial and statistically significant (loadings ranging from 0.26 to 0.87, mean loading was 0.57), indicating that they tend to converge in their assessment of each confirmation bias component. Once again, convergent validity was the lowest for the measurement of confirmation bias in information search and it was the highest for the measurement of confirmation bias in weighing of evidence. Moreover, the CTCU model revealed that the intercorrelations among the confirmation bias components were moderate to large in size (mean $r = 0.60$) and highly significant, showing low evidence of discriminant validity. Finally, virtually all correlations between residuals among indicators of the same method were non-significant and small, practically indicating the absence of method effects for logical rule hypothesis testing (mean $r = 0.15$), numerical rule hypothesis testing (mean $r = 0.15$), and trait hypothesis testing (mean $r = 0.11$). In other words, the variance observed in each task was primarily due to the trait factor rather than to the method factor.

Our previous analysis indicates that the three components of confirmation bias are substantially correlated. Consequently, another way to represent this structure is by introducing a general confirmation bias as a second-order factor over these three components. We illustrate this model in Fig. 2B. Note that fit indices for this model are identical to those of the model in the previous section, as both models share an equivalent structure: the

	Logical rule hypothesis testing			Numerical rule hypothesis testing			Trait hypothesis testing		
	IS	WE	MR	IS	WE	MR	IS	WE	MR
<i>Logical rule hypothesis testing</i>									
IS	.84								
WE	<i>.23**</i>	.94							
MR	<i>.12</i>	<i>.48***</i>	.86						
<i>Numerical rule hypothesis testing</i>									
IS	.16*	<i>.41***</i>	<i>.28***</i>	.75					
WE	<i>.10</i>	.65***	<i>.31***</i>	<i>.42***</i>	.81				
MR	<i>-.07</i>	<i>.31***</i>	.41***	<i>.26***</i>	<i>.35***</i>	.55			
<i>Trait hypothesis testing</i>									
IS	.13	<i>.26***</i>	<i>.13</i>	.30***	<i>.31***</i>	<i>.10</i>	.64		
WE	<i>.18**</i>	.42***	<i>.35***</i>	<i>.20**</i>	.34***	<i>.18**</i>	<i>.17***</i>	.69	
MR	<i>-.05</i>	<i>.10</i>	.28***	<i>.05</i>	<i>.05</i>	<i>.17**</i>	<i>.05</i>	<i>.33***</i>	.48

Table 2. Multitrait–multimethod correlation matrix ($N = 200$). IS: Information search, WE: Weighing of evidence, MR: Memory recall. Monotrait–heteromethod correlations are in boldface. Heterotrait–monomethod correlations are in italics. Internal consistencies are along the diagonal. * $p < .05$, ** $p < .01$, *** $p < .001$, two-tailed.

Parameter	Information search	Weighing of Evidence	Memory Recall
<i>Factor loadings</i>			
Logical rule hypothesis testing	.26**	.87***	.80***
Numerical rule hypothesis testing	.65***	.74***	.54***
Trait hypothesis testing	.45***	.50***	.30***
<i>Latent factor correlations</i>			
Information search	1.00		
Weighing of evidence	.73***	1.00	
Memory recall	.45***	.61***	1.00
<i>Correlations among uniqueness</i>			
Logical rule hypothesis testing			
Information search	1.00		
Weighing of evidence	.16	1.00	
Memory recall	.12	.17	1.00
Numerical rule hypothesis testing			
Information search	1.00		
Weighing of evidence	.12	1.00	
Memory recall	.13	.21*	1.00
Trait hypothesis testing			
Information search	1.00		
Weighing of evidence	.01	1.00	
Memory recall	.02	.30***	1.00

Table 3. Correlated traits–correlated uniqueness (CTCU) standardized parameter estimates. * $p < .05$, ** $p < .01$, *** $p < .001$.

three pairwise correlations between the components are now described as the loadings of each component on the general confirmation bias factor.

To further shed light on the factorial structure underlying all tasks, we tested two additional models (see Fig. 2C and 2D). First, we evaluated a bifactor model with one general factor (“confirmation bias”) and three domain-specific factors (IS, WE, and MR). Because this model did not converge, we also allowed for a correlation between the domain-specific IS factor and the general confirmation bias factor. The resulting model converged, but it yielded worse fit indices than our original model ($\chi^2(17) = 54.08$, $p < 0.001$, CFI = 0.895, RMSEA = 0.104, SRMR = 0.054). Second, we evaluated a single-factor model where a general confirmation bias factor accounts for all 9 tasks at once, without the intermediate layer of 3 components. Again, this model did not result in better fit indices than our original model ($\chi^2(18) = 58.42$, $p < 0.001$, CFI = 0.888, RMSEA = 0.106, SRMR = 0.064). These findings suggest that our data are best explained by three domain-specific factors, each representing a confirmation bias across different experimental paradigms.

Controlling for cognitive ability and cognitive impulsivity (CRT)

One could argue that confirmation bias is confounded with cognitive ability. In fact, participants with low cognitive ability might have more difficulty to understand the tasks, and therefore be more prone to confirmation bias. Moreover, it has been reported that cognitive ability does correlate with the tendency to avoid some thinking biases including the matching bias on the Wason selection task⁵⁵. To evaluate this possibility, we used the scores on the matrix and verbal reasoning tasks as measures of cognitive ability, in a subsample of participants who completed a follow-up session. Scores on the matrix reasoning task ($M = 0.64$, $SD = 0.24$, Cronbach's $\alpha = 0.72$) and on the verbal reasoning task ($M = 0.71$, $SD = 0.20$, $\alpha = 0.76$) were generally high in comparison to the norms reported by Condon and Revelle⁴², and reached satisfactory levels of internal consistency. As expected, these two scores were highly correlated across individuals ($r = 0.61$, $p < 0.001$), as already described by Condon and Revelle⁴², so we averaged them to produce a measure of participants' cognitive ability. We conducted for each confirmation bias measure a regression with cognitive ability across participants, and then we took the residuals of this regression (i.e., the variation in confirmation bias which is not accounted for by cognitive ability scores). From then, we replicated all of our analyses (correlation matrix, CFA–MTMM analysis) on these residuals (see Supplementary Materials). All of our findings were unchanged, suggesting that the results are not driven by a confound between confirmation bias and cognitive abilities.

Besides, it was reported that the CRT was a strong independent predictor of performance on rational thinking tasks, even after controlling for cognitive ability³⁹. As we have also measured CRT in our participants (with Cronbach's $\alpha = 0.75$), we could also replicate all of our analyses when partialling out both cognitive ability and CRT scores from all confirmation bias measures. Again, the results were unchanged (see Supplementary Materials).

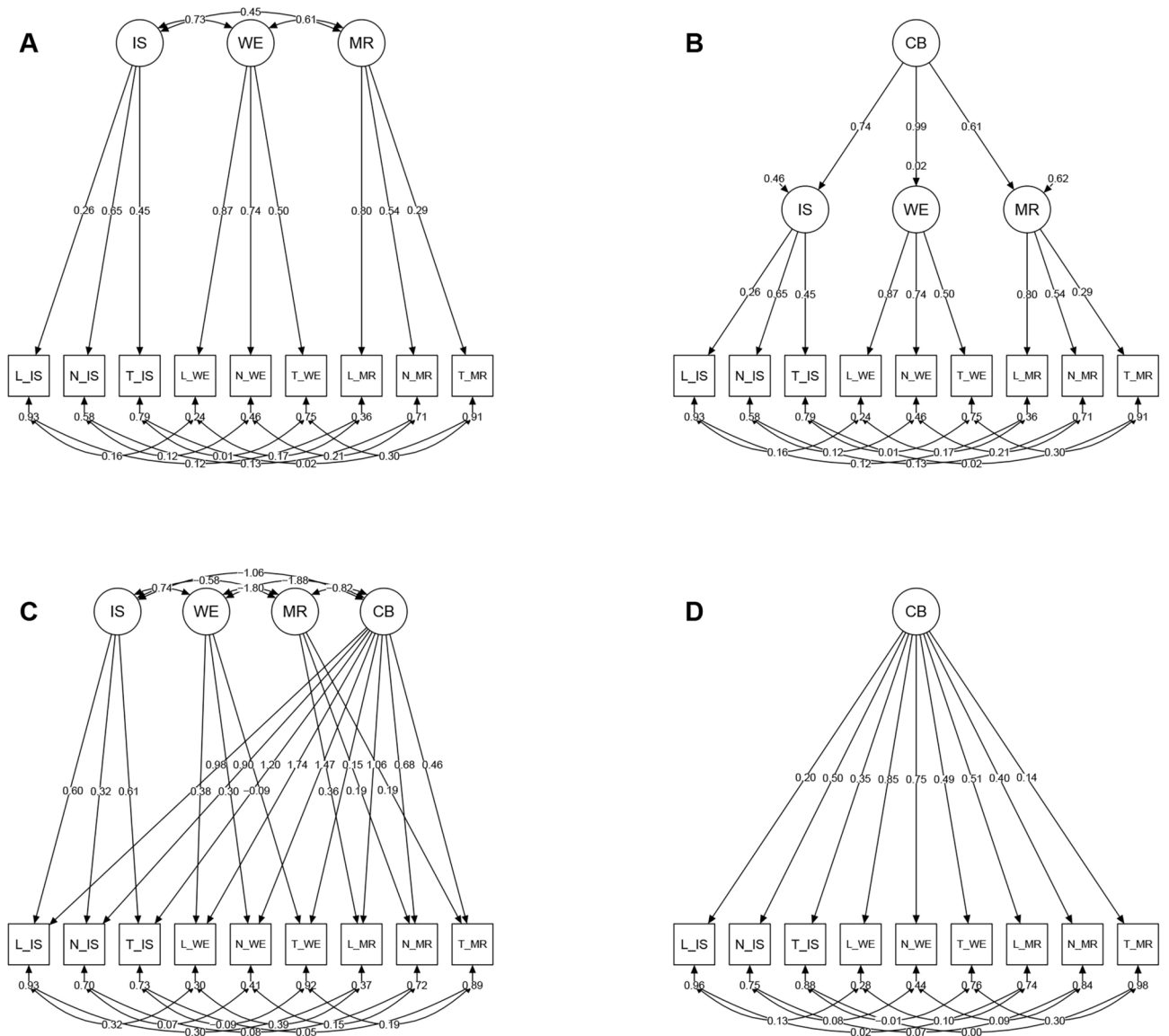


Fig. 2. Confirmatory factor analysis (CFA) of the MTMM matrix. Panel A: Correlated trait–correlated uniqueness (CTCU) model with three correlated first-order factors, representing the three components of confirmation bias. Panel B: CTCU model with a general confirmation bias as a second-order factor, accounting for the correlations between the three first-order factors. Panel C: Bifactor model with a general confirmation bias factor and three domain-specific factors (components). Panel D: Single-factor model where a general confirmation bias factor accounts for all 9 tasks simultaneously. The models shown in Panels A and B produce the best fit indices, and these indices are identical for both models, $\chi^2(15) = 21.76, p = .11, CFI = .981, RMSEA = .047, SRMR = .041$. L: Logical rule hypothesis testing, N: Numerical rule hypothesis testing, T: Trait hypothesis testing, IS: Information search, WE: Weighing of evidence, MR: Memory recall.

Correlations with external measures

To examine how confirmation bias may relate to current questions in social sciences⁷, in our follow-up session we also collected individuals’ scores on three scales measuring dogmatism (unchangeable and unjustified certainty confidence in one’s own beliefs), pseudoscientific beliefs, and the general tendency to believe in conspiracy theories. The dogmatism scale (Cronbach’s $\alpha = 0.89$) and the pseudoscientific beliefs scale ($\alpha = 0.81$) reached excellent levels of internal consistency, suggesting that subsequent correlational analyses could be reasonably conducted (note that the conspiracy beliefs scale included a single item, which prevented us from estimating its reliability).

We then investigated the correlations between these three external measures and the factor scores for the three components of confirmation bias. These correlations are presented in Table 4. First, we found that scores on the three scales were positively and significantly correlated with each other. In addition, the dogmatism and pseudoscientific beliefs scales were negatively and significantly correlated with both cognitive ability and CRT

	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8
1. Information search	0.06	0.71	–							
2. Weighing of evidence	0.03	0.83	.91***	–						
3. Memory recall	0.00	0.81	.66***	.77***	–					
4. Dogmatism	3.09	0.96	.06	–.01	.08	–				
5. Pseudoscientific beliefs	3.18	0.79	.20*	.21*	.23*	.19*	–			
6. Conspiracy beliefs	5.67	2.50	.05	.00	.10	.33**	.39***	–		
7. Cognitive ability	0.68	0.20	–.35***	–.33**	–.31**	–.35***	–.27**	–.14	–	
8. CRT	3.06	2.01	–.34***	–.32**	–.34***	–.20*	–.27**	–.05	.54***	–

Table 4. Correlations between the factor scores on the three components of confirmation bias and external measures ($N = 107$). * $p < .05$, ** $p < .01$, *** $p < .001$, two-tailed.

scores, providing evidence of their validity. Most importantly, scores on the three components of confirmation bias were positively and significantly correlated with pseudoscientific beliefs.

Discussion

The present study was designed to investigate the construct validity of three paradigms measuring confirmation bias in hypothesis testing. In line with previous literature¹, we found that when asked to test a given hypothesis, participants were prone to (1) search for confirming information more than for disconfirming information, (2) rate confirming information as more informative in comparison to disconfirming information, (3) rate confirming information as more previously encountered than disconfirming information. In other words, confirmation bias was present in three components of information processing.

Beyond the analysis of confirmation bias in each task separately, we then followed a CFA–MTMM approach to explore the correlations between the paradigms (convergent validity) as well as the correlations between the three components of confirmation bias (discriminant validity). To the best of our knowledge, our study is the first to address the construct validity of confirmation bias measures following a MTMM approach. Such an analysis requires having reliable measures of individual differences in each task. This requirement was almost achieved as six out of the nine measures reached an adequate level of reliability ($\alpha \geq 0.69$). This finding is noteworthy knowing that the original tasks were designed to produce robust experimental effects, which do not necessarily translate to reliable measures of individual differences, a phenomenon described by Hedge et al.⁵⁶ as the “reliability paradox”. From a methodological perspective, the tasks used in our study might be useful for individual differences research addressing the issue of a general factor of rationality.^{35,57–60}

Our CFA–MTMM results shed light on the construct validity of confirmation bias, especially in information search which has been the most studied¹. First, our data showed that, regardless of the hypothesis tested (logical rule, numerical rule, personality trait), the three paradigms used in this study converged in the measurement of each latent component of confirmation bias. In the CTCU model, all trait factor loadings were significant and only three of them were below 0.50. Interestingly, the information search component converged the less in our data set, which was manifest both in the MTMM correlation matrix (monotrait–heteromethod correlations) and the CTCU model, probably because the Wason selection, 2–4–6, and interviewee tasks are not pure measures of confirmation bias. Indeed, prior studies suggested that the 2–4–6 and interviewee tasks reflect a positive test strategy²² while the Wason selection task involves a matching bias²¹. While our finding that these three tasks are significantly correlated means that they share common variance, we acknowledge that the nature of this common variance remains subject to debate: all three tasks might involve a “true” confirmation bias, and/or a matching bias, for instance.

Second and most importantly, we found that the three components of confirmation bias were significantly and substantially correlated. This finding is also meaningful regarding the construct validity of confirmation bias in each component. For instance, in the case of weighing of evidence, confirming and disconfirming outcomes were equally informative in our design, so it is unlikely that participants’ choices are due to a diagnostic strategy (choosing the options that distinguish the most between the hypothesis at hand and its alternatives). Rather, because scores in this component are correlated with scores in information search and memory recall, the simplest explanation of this common variance is a general latent factor that captures individual differences in the tendency to confirm the hypothesis at hand.

It could be argued that there is some overlap between tasks within the same paradigm, which may have inflated the correlations observed (e.g., in the search paradigm, participants were presented with two choices, one of which was confirming while the other was not; this not only requires search but some form of weighing of evidence). This possibility actually strengthens our findings. Indeed, we report that monotrait–heteromethod correlations are higher than heterotrait–monomethod correlations, which is evidence of convergent validity between the three cognitive components of confirmation bias (information search, weighing of evidence, memory recall). In other words, despite the possibility that our procedure may have induced some correlations between the components within the same paradigm (i.e., heterotrait–monomethod correlations), our results indicate that empirically such correlations are weaker than the correlations between measures of a same component (i.e., monotrait–heteromethod correlations).

Several scholars outlined the “many guises” of confirmation bias^{2,3} or even questioned its existence at all. For instance, Mercier¹⁰ suggested that “there is no such thing as a general tendency to confirm whatever one

thinks about, only a tendency to find arguments that support one's own views – a myside bias” (p. 99–100). Our empirical approach was to investigate tasks in which individuals face hypotheses that have no relation to their own prior beliefs or views, i.e., tasks which are difficult to interpret in terms of a myside bias. Our finding of a common structure across cognitive components in these tasks suggests that individuals exhibit a general tendency to confirm the hypothesis at hand, even in this case. More research is needed however, to explore the structure of confirmation bias more generally, by also including tasks associated with myside bias in order to evaluate the potential unity of the construct across both internal and external hypotheses. For that purpose, one might build on the unified paradigm designed by Vedejová and Čavojská³¹, and evaluate the reliability of these tasks.

Before we conclude, one should acknowledge that the sample of our study and the social context in which it was conducted may limit the generalizability of our findings. Indeed, as is the case in many laboratory studies in psychology, our experiment involved a WEIRD sample of participants⁶¹. For instance, our participants were more educated than the general French population (59.5% of them reported having a master's degree), and their cognitive ability scores were above average as well. Moreover, our study was conducted in the aftermath of the COVID-19 pandemic, which constituted a specific social context, during which a significant portion of individuals in Western countries may have endorsed a conspiracy and confirmatory mentality^{62,63}. These considerations are of particular importance given that behavioral biases may be prone to cultural variation (e.g., see Mezulis et al.⁶⁴, for a meta-analysis of the self-serving attributional bias; and see Knobloch-Westerwick et al.⁶⁵, for an example of cultural influence on myside bias). These specific conditions regarding the sample and the context of our study invite caution regarding the generalizability of our results. However, although one can anticipate that the magnitude of confirmation bias may vary with context or culture, it is less clear why the structure of correlations between the different components of confirmation bias should change as well. Yet, this remains an empirical question.

Finally, one important avenue for future studies on confirmation may be to evaluate how individuals' confirmation bias scores, which can be reliably measured in laboratory tasks, relate to decision making in the real world. In particular, several scholars have argued that confirmation bias contributes to ideological extremism⁶⁶, polarization of opinions^{67,67}, and conspiracy beliefs⁴. We note that such studies are often based on a measure of confirmation that is in fact a myside bias, although this is not always the case. For instance, a recent study by Hattersley et al.⁴ reported correlations between confirmation bias scores in the Wason selection task and beliefs in implausible conspiracy theories (e.g., “The earth is flat, and scientists are concealing it from society”, “Feminism and campaigns for gender equality are just part of a secret agenda to hurt male rights”). In our data, we found that confirmation bias in information search, weighing of evidence and memory recall was positively and significantly correlated with scores on the pseudoscientific beliefs scale. Such results outline a relation between confirmation bias and irrational beliefs that foster conflicts between groups and individuals, giving credit to Lilienfeld et al.'s statement that “research on combating extreme confirmation bias should be among psychological science's most pressing priorities”⁶⁶.

Data availability

All data files and all materials used in the tasks are available at: <https://osf.io/saj4c/>.

Received: 24 April 2024; Accepted: 28 October 2024

Published online: 13 November 2024

References

- Oswald, M. E. & Grosjean, S. Confirmation bias. In *Cognitive illusions: A handbook on fallacies and biases in thinking, judgement and memory* (ed. Pohl, R. F.) 79–96 (Psychology Press, 2004).
- Nickerson, R. S. Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Rev. Gen. Psychol.* **2**(2), 175–220 (1998).
- Klayman, J. Varieties of confirmation bias. In *Decision making from a cognitive perspective* (eds Busemeyer, J. et al.) 365–418 (Academic Press, 1995).
- Hattersley, M., Brown, G. D. A., Michael, J. & Ludvig, E. A. Of tinfoil hats and thinking caps: Reasoning is more strongly related to implausible than plausible conspiracy beliefs. *Cognition* **218**, 104956 (2022).
- Zhou, Y. & Shen, L. Confirmation Bias and the Persistence of Misinformation on Climate Change. *Communication Research* **49**(4), 500–523 (2022).
- Lord, C. G., Ross, L. & Lepper, M. R. Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology* **37**(11), 2098–2109 (1979).
- Jost, J. T., Baldassarri, D. S. & Druckman, J. N. Cognitive-motivational mechanisms of political polarization in social-communicative contexts. *Nature reviews psychology* **1**(10), 560–576 (2022).
- Evans, J. S. B. T. *Bias in human reasoning: Causes and consequences* (Erlbaum, 1989).
- Baron, J. Myside bias in thinking about abortion. *Thinking and Reasoning* **1**, 221–235 (1995).
- Mercier, H. (2017). Confirmation bias—Myside bias. In R. F. Pohl (Ed.), *Cognitive illusions: Intriguing phenomena in thinking, judgment and memory* (pp. 99–114). New York, NY: Routledge/Taylor & Francis Group.
- Stanovich, K. E. *The bias that divides us: The science and politics of myside thinking* (The MIT Press, 2021).
- Baron, J. *Thinking and deciding* 4th edn. (Cambridge University Press, 2008).
- Poletiek, F. H. *Hypothesis-testing behaviour* (Psychology Press, 2001).
- Wason, P. C. On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology* **12**, 129–140 (1960).
- Wason, P. C. Reasoning. In *New horizons in psychology I* (ed. Foss, B. M.) 106–137 (Penguin, 1966).
- Wason, P. C. & Johnson-Laird, P. N. *Psychology of Reasoning. Structure and Content* (Harvard University Press, 1972).
- Snyder, M. & Swann, W. B. Hypothesis-testing processes in social interaction. *Journal of Personality and Social Psychology* **36**(11), 1202–1212 (1978).
- Snyder, M. & White, P. Testing Hypotheses about Other People: Strategies of Verification and Falsification. *Personality and Social Psychology Bulletin* **7**(1), 39–43 (1981).
- Wason, P. C. Reasoning about a Rule. *Quarterly Journal of Experimental Psychology* **20**(3), 273–281 (1968).

20. Evans, J. S. B. T. Reasoning, Biases and Dual Processes: The Lasting Impact of Wason (1960). *Quarterly Journal of Experimental Psychology* **69**(10), 2076–2092 (2014).
21. Evans, J. S. Interpretation and matching bias in a reasoning task. *The Quarterly Journal of Experimental Psychology* **24**(2), 193–199 (1972).
22. Klayman, J. & Ha, Y.-W. Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review* **94**(2), 211–228 (1987).
23. Baron, J., Beattie, J., & Hershey, J. C. (1988). Heuristics and biases in diagnostic reasoning: II. Congruence, information, and certainty. *Organizational Behavior and Human Decision Processes*, **42**(1), 88–110.
24. Trope, Y. & Bassok, M. Confirmatory and diagnosing strategies in social information gathering. *Journal of Personality and Social Psychology* **43**(1), 22–34 (1982).
25. Trope, Y. & Bassok, M. Information-gathering strategies in hypothesis-testing. *Journal of Experimental Social Psychology* **19**(6), 560–576 (1983).
26. Devine, P. G., Hirt, E. R. & Gehrke, E. M. Diagnostic and confirmation strategies in trait hypothesis testing. *Journal of Personality and Social Psychology* **58**(6), 952–963 (1990).
27. Hoch, S. J. & Ha, Y. W. Consumer Learning Advertising and the Ambiguity of Product Experience. *Journal of Consumer Research* **13**(2), 221–233 (1986).
28. Beattie, J. & Baron, J. Confirmation and matching biases in hypothesis testing. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology* **40**(2), 269–297 (1988).
29. Eagly, A. H., Chaiken, S., Chen, S. & Shaw-Barnes, K. The impact of attitudes on memory: An affair to remember. *Psychological Bulletin* **125**(1), 64–89 (1999).
30. Stangor, C. & Mcmillan, D. Memory for expectancy-congruent and expectancy-incongruent information: A review of the social and social developmental literatures. *Psychological Bulletin* **111**(1), 42–61 (1992).
31. Vedejová, D. & Čavojová, V. Confirmation bias in information search, interpretation, and memory recall: Evidence from reasoning about four controversial topics. *Thinking & Reasoning* **28**(1), 1–28 (2022).
32. Campbell, D. T. & Fiske, D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin* **56**(2), 81–105 (1959).
33. Stanovich, K. E., West, R. F. & Toplak, M. E. Myside bias, rational thinking, and intelligence. *Current Directions in Psychological Science* **22**(4), 259–264 (2013).
34. Rassin, E. Individual differences in the susceptibility to confirmation bias. *Netherlands Journal of Psychology* **64**(2), 87–93 (2008).
35. Berthet, V., Autissier, D. & de Gardelle, V. Individual differences in decision-making: A test of a one-factor model of rationality. *Personality and Individual Differences* **189**, 111485 (2022).
36. Dawson, E., Gilovich, T. & Regan, D. T. Motivated reasoning and performance on the Wason Selection Task. *Personality and Social Psychology Bulletin* **28**(10), 1379–1387 (2002).
37. Stanovich, K. E., West, R. F. & Toplak, M. E. *The rationality quotient: Toward a test of rational thinking* (MIT Press, 2016).
38. Berthet, V. The Measurement of Individual Differences in Cognitive Biases: A Review and Improvement. *Frontiers in psychology* **12**, 630177 (2021).
39. Toplak, M. E., West, R. F. & Stanovich, K. E. Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning* **20**(2), 147–168 (2014).
40. Sirota, M. & Juanchich, M. Effect of response format on cognitive reflection: Validating a two- and four-option multiple choice question version of the Cognitive Reflection Test. *Behavior research methods* **50**(6), 2511–2522 (2018).
41. Hamilton, K., Shih, S. I. & Mohammed, S. The Development and Validation of the Rational and Intuitive Decision Styles Scale. *Journal of personality assessment* **98**(5), 523–535 (2016).
42. Condon, D. M. & Revelle, W. The international cognitive ability resource: Development and initial validation of a public-domain measure. *Intelligence* **43**, 52–64 (2014).
43. Altemeyer, B. Dogmatic behavior among students: testing a new measure of dogmatism. *The Journal of social psychology* **142**(6), 713–721 (2002).
44. Fasce, A., Avendaño, D. & Adrián-Ventura, J. Revised and short versions of the pseudoscientific belief scale. *Applied Cognitive Psychology* **35**(3), 828–832 (2021).
45. Fasce, A. & Picó, A. Conceptual foundations and validation of the Pseudoscientific Belief Scale. *Applied Cognitive Psychology* **33**(4), 617–628 (2019).
46. Lantian, A., Muller, D., Nurra, C. & Douglas, K. M. Measuring belief in conspiracy theories: Validation of a French and English single-item scale. *International Review of Social Psychology* **29**(1), 1–14 (2016).
47. Marsh, H. W. & Grayson, D. Latent-variable models of multitrait-multimethod data. In *Structural equation modeling: Issues and applications* (ed. Hoyle, R. H.) 177–198 (Sage, 1995).
48. Millsap, R. E. (1995). The statistical analysis of method effects in multitrait-multimethod data: A review. In P. E. Shrout & S. T. Fiske (Eds.), *Personality research, methods, and theory: A festschrift honoring Donald W. Fiske* (pp. 93–109). Hillsdale, NJ: Erlbaum.
49. Kenny, D. A. & Kashy, D. A. Analysis of the multitrait-multimethod matrix by confirmatory factor analysis. *Psychological Bulletin* **112**(1), 165–172 (1992).
50. Marsh, H. W. Confirmatory Factor Analyses of Multitrait-Multimethod Data: Many Problems and a Few Solutions. *Applied Psychological Measurement* **13**(4), 335–361 (1989).
51. Marsh, H. W. & Bailey, M. Confirmatory factor analyses of multitrait-multimethod data: A comparison of alternative models. *Applied Psychological Measurement* **15**(1), 47–70 (1991).
52. Brown, T. A. *Confirmatory Factor Analysis for Applied Research* 2nd edn. (The Guilford Press, 2015).
53. Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E. & Tatham, R. L. *Multivariate data analysis* 6th edn. (Pearson Prentice Hall, 2006).
54. Hu, L. T. & Bentler, P. M. Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives. *Structural Equation Modeling* **6**, 1–55 (1999).
55. Stanovich, K. E. & West, R. F. On the relative independence of thinking biases and cognitive ability. *Journal of personality and social psychology* **94**(4), 672–695 (2008).
56. Hedge, C., Powell, G. & Sumner, P. The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior research methods* **50**(3), 1166–1186 (2018).
57. Stanovich, K. E. & West, R. F. Individual differences in reasoning: Implications for the rationality debate?. *Behavioral and Brain Sciences* **23**(5), 645–665 (2000).
58. Burgoyne, A. P., Mashburn, C. A., Tsukahara, J. S., Hambrick, D. Z. & Engle, R. W. Understanding the relationship between rationality and intelligence: A latent-variable approach. *Thinking & Reasoning* **29**(1), 1–42 (2023).
59. Erceg, N., Galić, Z. & Bubić, A. Normative responding on cognitive bias tasks: Some evidence for a weak rationality factor that is mostly explained by numeracy and actively open-minded thinking. *Intelligence* **90**, 101619 (2022).
60. Teovanović, P., Knežević, G. & Stankov, L. Individual differences in cognitive biases: Evidence against one-factor theory of rationality. *Intelligence* **50**, 75–86 (2015).
61. Henrich, J., Heine, S. J. & Norenzayan, A. The weirdest people in the world?. *The Behavioral and brain sciences* **33**(2–3), 61–135 (2010).

62. Pummerer, L. et al. Conspiracy Theories and Their Societal Effects During the COVID-19 Pandemic. *Social Psychological and Personality Science* **13**(1), 49–59 (2022).
63. Bertin, P., Nera, K. & Delouvée, S. Conspiracy Beliefs, Rejection of Vaccination, and Support for hydroxychloroquine: A Conceptual Replication-Extension in the COVID-19 Pandemic Context. *Frontiers in psychology* **11**, 565128 (2020).
64. Mezulis, A. H., Abramson, L. Y., Hyde, J. S. & Hankin, B. L. Is there a universal positivity bias in attributions? A meta-analytic review of individual, developmental, and cultural differences in the self-serving attributional bias. *Psychological bulletin* **130**(5), 711–747 (2004).
65. Knobloch-Westerwick, S., Liu, L., Hino, A., Westerwick, A. & Johnson, B. K. Context impacts on the confirmation bias: Evidence from the 2017 Japanese snap election compared with American and German findings. *Human Communication Research* **45**(4), 427–449 (2019).
66. Lilienfeld, S. O., Ammirati, R. & Landfield, K. Giving debiasing away: Can psychological research on correcting cognitive errors promote human welfare?. *Perspectives on Psychological Science* **4**(4), 390–398 (2009).
67. Del Vicario, M., Scala, A., Caldarelli, G., Stanley, H. E. & Quattrociocchi, W. Modeling confirmation bias and polarization. *Scientific reports* **7**, 40391 (2017).

Acknowledgements

This work has benefitted from the support of Agence Nationale de la Recherche [grant ANR-19-CE28-0019-01 “Ambisense” to VdG].

Author contributions

V.B.: Conceptualization, Methodology, Writing—Original Draft, Writing—Review & Editing P.T.: Conceptualization, Methodology V.G.: Methodology, Writing—Original Draft, Writing—Review & Editing.

Funding

Agence Nationale de la Recherche, ANR-19-CE28-0019-01.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-78053-7>.

Correspondence and requests for materials should be addressed to V.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024