LONDON’S GLOBAL UNIVERSITY

UCL

How Messenger Characteristics Influence Expertise Learning and Information-Seeking Choices

Joseph Alexander Marks
Department of Experimental Psychology
University College London (UCL)

Thesis submitted for the degree of
Doctor of Philosophy (Ph.D.)
March 2022

Declaration

I, Joseph Alexander Marks confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed: Joseph Alexander Marks, 27th March 2022.

Acknowledgements

This thesis would not have been possible without the faith, guidance, and support of my friends, colleagues, collaborators, funders, mentors, and loved ones.

First and foremost, I want to thank Adam Harris for enthusiastically adopting me as his PhD student. Adam is everything one could ask for in a supervisor: incredibly insightful, appropriately sceptical, and always responsive; warm, patient, and compassionate; and just generally an absolute joy to work with. I could not have hoped for a better academic mentor. Likewise, I will be forever grateful to Tali Sharot for taking me under her wing, involving me in her inspiring research programme, investing vast amounts of time and effort into my research endeavours, and teaching me countless invaluable lessons.

A huge thanks also to Cass Sunstein. Working with one of the founding fathers of behavioural economics – who also happens to be one of the kindest and most affirming people in academia – was an extremely humbling experience. And to David Lagnado who, like Adam, welcomed me with open arms into his lab and was always there to provide thoughtful, astute, and measured guidance when needed.

Special thanks go to Oliver Harrison, Aleksandar Matic, and Telefonica Alpha; Nick Barron, Alex Bigg, Kate Pogson, and the rest of the team at MHP; as well as John Draper and the PALS Division at UCL for providing the funding for my PhD. I would not have been able to complete this thesis without your support.

One of the perks of splitting my time between two labs during my PhD was that I got to spend it with two groups of amazing people. A big thank you to all the members, past and present, of the Affective Brain Lab and the (extended) Causal Cognition Lab for filling my time at UCL with fun, educational, and inspiring conversations. And to Rebecca Saxe and the SaxeLab for taking me in as one of your own during my time at MIT.

A special mention to Filip Gesiarz, Chris Kelly, Bastien Blain, Marius Vollberg, and Otto Simonsson. Filip for sharing his analysis scripts with me, teaching me how to fit computational models, and eating countless Tesco's wraps with me in Russell

Square; Chris for being my partner in crime at MIT, letting me win at table-tennis, and always offering to buy the first round; Bastien for filling so many lunchtimes with raucous laughter and always offering advice and assistance on all things Matlab; Marius for being a wonderful tour guide and source of support and friendship in Massachusetts; and Otto for constantly pushing me to get involved in his research activities, introducing me to his collaborators, and sprucing my decision-making research portfolio with more exotic meditation-based projects.

A special mention also to the MSc students I have had the pleasure of working with, Eloise Copland, Helen St Quintin, and Saumya Mody. You were all brilliant and helped me greatly with my research.

Thank you to all the staff in the UCL PALS Division who have helped me along the way, in particular Jeremy Skipper, Eva Krumhuber, Lee De-Wit, Dimitrios Tsivrikos, Lasana Harris, Antonietta Esposito, Pia Horbacki, Hannah Spikesley, and Steve Williams.

I also owe a huge debt of gratitude to Steve Martin, who kindly invited me to write a book with him on the same topic as this thesis back in 2017. Steve has been a constant source of encouragement and support over the last five years. Co-authoring a book while doing a PhD was no easy feat, but Steve made the experience as enjoyable as it was rewarding.

Lastly, a thank you to my family – Hilary, Richard, Olly, Jules, Shirley, John, Valerie, Bernard, my wonderful cousins, aunts and uncles – and friends (you know who you are) for all your support, wise words, and love.

Abstract

When trying to form accurate beliefs and make good choices, people often turn to one another for information and advice. But deciding whom to listen to can be a challenging task. While people may be motivated to receive information from accurate sources, in many circumstances it can be difficult to estimate others' task-relevant expertise. Moreover, evidence suggests that perceptions of others' attributes are influenced by irrelevant factors, such as facial appearances and one's own beliefs about the world. In this thesis, I present six studies that investigate whether messenger characteristics that are unrelated to the domain in question interfere with the ability to learn about others' expertise and, consequently, lead people to make suboptimal social learning decisions.

Studies one and two explored whether (dis)similarity in political views affects perceptions of others' expertise in a non-political shape categorisation task. The findings suggest that people are biased to believe that messengers who share their political opinions are better at tasks that have nothing to do with politics than those who do not, even when they have all the information needed to accurately assess expertise. Consequently, they are more likely to seek information from, and are more influenced by, politically similar than dissimilar sources.

Studies three and four aimed to formalise this learning bias using computational models and explore whether it generalises to a messenger characteristic other than political similarity. Surprisingly, in contrast to the results of studies one and two, in these studies there was no effect of observed generosity or political similarity on expertise learning, information-seeking choices, or belief updating.

Studies five and six were then conducted to reconcile these conflicting results and investigate the boundary conditions of the learning bias observed in studies one and two. Here, we found that, under the right conditions, non-politics-based similarities can influence expertise learning and whom people choose to hear from; that asking people to predict how others will answer questions enhances learning from observed outcomes; and that it is unlikely that inattentiveness explains why we observed null effects in studies three and four.

Impact Statement

Much of what people know about the world comes from information they receive from others. If an individual chooses to seek information and learn from sources that possess useful and reliable information, they can improve the accuracy of their beliefs, increase the rewards they receive from their decisions, and avoid costly mistakes. At a cultural level, societies advance and thrive when useful and accurate information is transmitted through social networks with high fidelity. Accordingly, it is surprising that people often selectively seek out and believe information from sources that they find congenial rather than prioritising accuracy. Moreover, in an era with mass-communication, the Internet, and easy national and international travel, individuals now have more freedom than ever before to choose where they get information from and can therefore, to a large extent, filter what they see and hear. There is a growing concern that the combination of an ever-increasing ability to choose where to go to for information and a human tendency to receive information from like-minded sources is leading individuals to believe 'fake news' and form maladaptive beliefs, groups to form 'echo chambers', and societies to become more polarised.

This thesis presents a novel account of why people choose to seek information from like-minded sources. In particular, the empirical results reported here demonstrate that in some circumstances people judge those who share their values or political beliefs as more competent at unrelated tasks than those with differing views, even when they have all the information needed to accurately assess expertise in the relevant domain, consequently leading to suboptimal information-seeking, belief updating and decision-making. These findings highlight that people may believe that they are learning from accurate sources, even in cases where the evidence in front of them indicates that they are not, and thus provide important contributions to the wider literature on social learning and person perception.

The results of this research have been disseminated widely and have already had a broad impact in the mainstream media and on many subfields within psychology and the social sciences. In particular, the studies reported in Chapter 2 were published in a top-ranking psychology journal (Cognition) and discussed in the New York Times, the Guardian and UCL News. Since publication, many researchers from different labs and institutions, in fields such as communication research, political science, management, and psychology, have cited this work. Furthermore, I have had the honour of being invited to write about this research in mainstream publications and discuss it in podcasts, lectures and lab meetings in both the UK and the US. The broad interest in this work indicates that the findings presented in this thesis provide value both inside and outside of academia.

Table of Contents

List of Figures

List of Tables

Contributions

The work reported in this thesis is entirely my own except for the contributions acknowledged below. All chapters have benefited from guidance and advice from my supervisors, Dr Adam Harris, Prof. Tali Sharot, Prof. Cass Sunstein, and Prof. David Lagnado.

The Blap [Blup] Task used in Chapters 2 and 4 was adapted from an unpublished study originally conducted by Eleanor Loh and Prof. Tali Sharot. Chapter 2 was conducted in collaboration with Eloise Copland, an MSc Behaviour Change student at the UCL Division of Psychology and Language Sciences. This was co-supervised by myself, Prof. Cass Sunstein, and Prof. Tali Sharot. Eloise assisted with the study design, implementation in Qualtrics, data collection, and analysis, under my supervision.

Chapter 4, Study 5 was conducted in collaboration with Helen St Quintin, an MSc Psychological Sciences student at the UCL Division of Psychology and Language Sciences. The research was co-supervised by myself and Dr Adam Harris. Helen assisted with the study design, implementation of the study in Qualtrics, and data collection, under my supervision.

Notes to Examiners

The findings from Chapter 2 have been published in a peer-reviewed article:

Marks, J., Copland, E., Loh, E., Sunstein, C. R., & Sharot, T. (2019). Epistemic spillovers: Learning others' political views reduces the ability to assess and use their expertise in nonpolitical domains. *Cognition*, *188*, 74-84.

Other works conducted during the course of the PhD but not included in this thesis include:

Martin, S. & Marks, J. (2019). *Messengers: Who We Listen To, Who We Don't, and Why*. Penguin Random House.

Marks, J., & Sharot, T. (2019). Forecasting the US Primary Elections From Convenience Samples Using Behavioural Science Tools. Poster presented by Joseph Marks at the Society for Neuroeconomics, Dublin, Ireland.

Simonsson, O., Narayanan, J., & Marks, J. (2021). Love thy (partisan) neighbor: Brief befriending meditation reduces affective polarization. *Group Processes and Intergroup Relations.* https://doi.org/10.1177/13684302211020108

Marks, J., Czech, P., & Sharot, T. (Under Review). Observing others give & take: A computational account of bystanders' feelings and actions. *Available at SSRN:* https://dx.doi.org/10.2139/ssrn.3674051.

Chapter 1. Theoretical Background

To form accurate beliefs and make good decisions, humans learn from others. From infancy, we build our understanding of the world not only through direct experiences but also by observing and listening to those around us (Bandura, 1977; Csibra & Gergely, 2011; Harris & Corriveau, 2011). As we grow older, we turn to one another for information and advice on both benign matters, such as which movies to watch, as well as the more consequential, such as medical, legal, and financial questions.

Indeed, much of what we know comes from information we receive from others. For example, our knowledge of history is almost entirely comprised of information provided to us by other people. The same goes for our understanding of scientific entities and ontological features of the world that we cannot observe for ourselves, such as the existence of germs, the shape of the earth, and the relationship between mind and matter. We are unable to learn about such facts by making first-hand observations, so we rely on second-hand knowledge – which may have been passed through multiple sources – to build our understanding of the environment (Coady, 1992).

The success of humanity is often attributed to this ability to share technological, social, conventional, and institutional knowledge and skills (Boyd et al., 2011; Csibra & Gergely, 2011; Henrich & McElreath, 2003; Tomasello et al., 2005). Complex practices and ideas are developed and honed over generations, being passed from one to the next. Therefore, human culture, unlike other animal populations, is characterised by a rachet effect, whereby there is an increase in the complexity or efficiency of technology over time (Tennie et al., 2009; Tomasello, 1999). As a result, most human societies today are so reliant on sophisticated technologies that individuals have no choice but to trust and use information from others (Boyd & Richerson, 1985).

Social learning has been defined as "learning from, or in interaction with, other individuals" (Olsson et al., 2020, p. 202). Given the centrality of social learning to belief formation and behaviour, it is important to understand how people decide which sources to seek information from and trust. Individuals do not indiscriminately accept claims made by others for risk of being unintentionally led astray or purposefully exploited. Nor do they seek information from every source available, as there are often costs involved in acquiring new information, such as time and energy loss (Boyd & Richerson, 1985; Morrison & Vancouver, 2000). Rather, people display considerable selectivity in when they listen to others, whom they listen to, and what they believe (for a review, see Kendal et al., 2018).

This thesis is concerned with how people decide whom to learn from and, in particular, the role that beliefs about irrelevant messenger characteristics (also known as "source characteristics") play in shaping how people learn about and utilise others' task-relevant expertise. Here, a messenger is defined as an agent that transmits information to others, where the transmission of information may be intentional, as when a parent teaches their child about some fact of life, or unintentional, as when a parent speaks without knowing that their child is listening. A messenger characteristic is defined as a feature or quality, such as a trait, behaviour, or piece of information, that serves to define the messenger's identity. Starting from the premise that there is a 'right' way to decide whom to learn from, this thesis will explore whether humans are able to approximate this optimal decision-making process by learning about relevant messenger characteristics (here, their task-relevant expertise) or whether the ability to learn about relevant messenger characteristics, and thus optimally make decisions of whom to learn from, is biased by beliefs about irrelevant messenger characteristics. The rationale underlying this research question will be explained in this chapter.

Chapter 1 will thus be dedicated towards introducing a normative theory of social learning, as well as key descriptive results from empirical research, to place the hypotheses and experiments that follow in a theoretical context. As the focus of the thesis is primarily on one aspect of social learning – namely, information-seeking decisions – particular attention will be paid to how people should decide whom to seek information from and what they should do upon receiving this information. Nonetheless, as social learning decisions are expected to be governed by the same rules as non-social decisions (e.g., Behrens et al., 2008), this chapter begins by introducing normative models that have been devised to prescribe how people should make judgements and decisions more generally.

Normative Models of Judgement and Decision-Making

To assess the optimality of social learning decisions, we need a normative theory that prescribes how people should make these decisions, given the context and their goals, against which to compare descriptive results. Social learning occurs under conditions of uncertainty. There is no point in acquiring and evaluating information from others if already certain about the veracity of a proposition or the best decision to make. In conditions of uncertainty, Subjective Expected Utility Theory and Bayesian Probability Theory provide the foundations for normative models, prescribing how people should learn about and from others. Each of these theories will be outlined in the following subsections, with a particular focus on how they can be applied to social learning decisions.

Subjective Expected Utility Theory

The standard normative model of decision-making under uncertainty is Subjective Expected Utility Theory (SEU) (Savage, 1954; Von Neumann & Morgenstern, 1944). At its simplest level, this model dictates that when confronted with a decision people should choose the option that is expected to provide them with the most subjective value, or 'utility'. How an individual assigns subjective utility will depend on their goals, with outcomes that have a large impact on important goals receiving more weight than those that do little to further or hinder one's goals (Baron, 1996). More formally, SEU prescribes that, in order to arrive at a decision, individuals should multiply the subjective utility of each possible outcome of a decision by the respective subjective probability that the outcome will occur.

To give a concrete example, imagine a person, Thomas, is trying to decide whether to take an umbrella with him when leaving the house. To act in accordance with SEU, he will first need to consider the possible outcomes associated with each choice. If he takes the umbrella, he may prevent himself from getting rained on. But it might not rain, in which case he will have to carry the umbrella around for no good reason. On the other hand, if he does not take the umbrella and it does rain, he will get wet. Whereas, if it doesn't rain, he will not get wet or have the burden of carrying the umbrella. Thomas can assign a utility value to each of these potential outcomes. Given that Thomas does not want to get wet or carry an umbrella around, the utility of each outcome might look like that presented in Table 1.

Table 1

The Possible Outcomes and Utilities Under Consideration When an Individual Is Deciding Whether to Take an Umbrella Out with Them

Options in the choice set	Weather Event	
	Rain	No rain
Umbrella	-10	-10
No umbrella	-100	0

Expected utilities (EUs) are calculated by multiplying the utility of each outcome by the probability of it occurring. Therefore, if there is a 30 percent chance it will rain and a 70 percent chance it will not rain, the expected utilities will be as in Table 2.

Table 2

The Expected Utility of Each Outcome When an Individual Is Deciding Whether to Take an Umbrella Out with Them and There Is A 30 Percent Chance of Rain

Options in the choice set	Weather Event (probability)	
	Rain (30%)	No rain (70%)
Umbrella	-3	-7
No umbrella	-30	0

Finally, Thomas can calculate the EU of each choice by summing across the different outcomes that may occur if that choice is made. So, the EU of taking an umbrella is computed by adding the EU of taking an umbrella when it rains (-3) and the EU of taking an umbrella when it doesn't rain (-7; $EU_{Umbrella}$ = -3 - 7 = -10). Likewise, the EU of not taking an umbrella is computed by adding the EU when it does rain (-30) and the EU when it doesn't (0) ($EU_{No\ Umbrella}$ = -30 - 0 = -30). Thomas then simply chooses the action with highest EU, which in this case is to take the umbrella.

Of course, when making real-world decisions, people do not usually have full knowledge of all the possible options available to them or the potential outcomes and associated probabilities and utilities (Simon, 1979). SEU suggests that the probabilities and utilities assigned to each outcome are subjective, meaning that they reflect the individual's personal beliefs and preferences rather than objective truths. Thus, in order to arrive at a decision, individuals must estimate the probabilities and utilities of the outcomes that come to mind from the choices under consideration. These constraints need to be taken into account in order to assess the optimality of human decision-making. That is, one must consider what a decision-maker knows, or should know, in a given environment (Anderson, 1990).

Bayesian Probability Theory

Bayesian Probability Theory provides a normative framework for handling uncertainty in probability estimates. Bayesians conceptualise probabilities as subjective degrees of belief, rather than objective frequencies. On the Bayesian account, the probability of rain in the above example reflects Thomas's subjective belief, which may change if he receives new information. In the Bayesian framework, beliefs are therefore represented by probability distributions over different hypotheses (Strevens, 2006).

Although the Bayesian approach introduces subjectivity into probability judgements, there are still rules that constrain Bayesian probability and thus allow it to serve as a normative model (Lindley, 1994; Rosenkrantz, 1992). Most notably, the coherence principle dictates that a person's probability judgments must adhere to the fundamental axioms of probability (Cox, 1946; Howson & Urbach, 1996). These are that: 1) all probabilities are real numbers between zero and one; 2) the probabilities of all the possible outcomes in a sample space add up to one, and 3) the probability of one of two mutually exclusive outcomes occurring is the sum of their individual probabilities. Other mathematical rules of probability can then be derived from these axioms.

One such rule is Bayes' theorem, which prescribes how beliefs should be updated in light of new evidence. Let us imagine that upon waking up Thomas has no idea whether it will rain or not. In the language of the theory, his prior belief that it will rain is P(0.5). After seeing the weather forecast, Bayes' theorem can be applied to calculate how much he should update this belief:

$$P(h|e) = \frac{P(h)P(e|h)}{P(h)P(e|h) + P(\neg h)P(e|\neg h)} \tag{1}$$

The left-hand side of this equation represents his posterior degree of belief in the hypothesis, *h* (that it will rain), given the observed evidence, *e* (the information he receives from the weather forecast). The right-hand side shows that this value can be normatively derived, using the rules of probability theory, by multiplying his prior belief, *P(h),* by the likelihood of observing that evidence if the hypothesis is true, *P(e|h)*, and normalising by the probability of the evidence (regardless of the truth or falsity of *h*).

Bayesian learners therefore update their beliefs according to the strength of the evidence with which they are presented. If *P(e|h) > P(e|¬h)*, then the learner will increase their belief in *h*; if *P(e|h) < P(e|¬h)*, then their belief in *h* will decrease; and if *P(e|h) = P(e|¬h)*, then their belief will remain unchanged. The degree to which *P(e|h)* and *P(e|¬h)* differ will influence how much the learner will update their belief, with larger differences representing stronger evidence and stronger evidence having a greater effect on belief updating than weaker evidence.

Information-Seeking Decisions

Information-seeking decisions can be included as choice options in normative models of decision-making (Edwards, 1965; Kobayashi & Hsu, 2019; Moutoussis et al., 2011; Stigler, 1961). This is because information may carry instrumental benefits that improve the decision-maker's knowledge of the world and thus help them to make better decisions. The utility of information can be mathematically formalised as the change in the expected utility of a decision from accruing said information.

Let us imagine again that Thomas's prior belief that it will rain is 0.5 and the utility of each outcome takes the same value as in Table 1. The EUs of taking and not taking an umbrella are calculated as:

$$EU_{Umbrella} = -10 \cdot 0.5 - 10 \cdot 0.5 = -10 \tag{2}$$

$$EU_{No\ umbrella} = -100 \cdot 0.5 + 0 \cdot 0.5 = -50 \tag{3}$$

Assuming that Thomas is fully deterministic and always chooses the option with the highest utility, this equation can be re-framed to indicate the utility of the decision in his current knowledge state, S$_0$:

$$EU(S_0) = \max[-10 \cdot 0.5 - 10 \cdot 0.5, -100 \cdot 0.5 + 0 \cdot 0.5] = -10 \tag{4}$$

Now we can factor in the effect that acquiring information has on Thomas's probability estimates. In general, he knows that checking the weather forecast usually improves his ability to predict whether it will rain or not but will likely not provide him with a definitive answer one way or the other. The degree to which information can discriminate between a particular hypothesis and its alternatives represents the 'diagnosticity' of that information. Here, the estimated diagnosticity of the weather forecast will be defined as π (where 0.5 ≤ π ≤ 1). The diagnosticity of information will influence the level of certainty Thomas believes he will possess after receiving that information and updating his belief, as it equates to the strength of the evidence in Bayes' theorem (Good, 1950). He can then calculate the EU of the future states he may enter. If the weather forecast suggests it is unlikely to rain, he will enter state S_+, whereas if rain is forecast, he will enter state S_-:

$$EU(S_+) = \max[-10 \cdot (1-\pi) - 10 \cdot \pi, -100 \cdot (1-\pi) + 0 \cdot \pi] \tag{5}$$

$$EU(S_-) = \max[-10 \cdot \pi - 10 \cdot (1-\pi), -100 \cdot \pi + 0 \cdot (1-\pi)] \tag{6}$$

In this example, if the estimated diagnosticity of the weather forecast, π, is less than or equal to 0.9 then Thomas will not bother to check it; he will take an umbrella with him regardless of what the forecast predicts, because his aversion to getting wet is so strong. For example, if Thomas estimates that π = 0.9 then:

$$\begin{aligned} EU(S_+) &= \max[-10 \cdot 0.1 - 10 \cdot 0.9, -100 \cdot 0.1 + 0 \cdot 0.9] \\ &= \max[-10, -10] \end{aligned} \tag{7}$$

$$\begin{aligned} EU(S_-) &= \max[-10 \cdot 0.9 - 10 \cdot 0.1, -100 \cdot 0.9 + 0 \cdot 0.1] \\ &= \max[-10, -90] \end{aligned} \tag{8}$$

Equation 7 indicates that if Thomas checks the weather forecast and sees that it is not expected to rain that day, he will be indifferent as to whether to take an umbrella out with him or not. That is, the expected utility of taking an umbrella and not taking an umbrella are equal and $EU(S_+) = -10$. If, on the other hand, Thomas checks the forecast and sees that it is expected to rain, then he will choose to take an umbrella out with him, as shown in Equation 8, and $EU(S_-) = -10$. Therefore, it does not benefit Thomas to check the weather forecast, as the EU of taking an umbrella will be equal to or greater than the EU of not taking an umbrella in either case. However, if π > 0.9 then Thomas will check the weather forecast because it may influence his decision; if after checking the forecast Thomas estimates the probability of rain is less than 0.1, he will not take the umbrella as the expected utility of leaving without it will be greater than the expected utility of taking it. For example, if Thomas estimates that π = 0.95 then:

$$\begin{aligned} EU(S_+) &= \max[-10 \cdot 0.05 - 10 \cdot 0.95, -100 \cdot 0.05 + 0 \cdot 0.95] \\ &= \max[-10, -5] \end{aligned} \tag{9}$$

$$\begin{aligned} EU(S_-) &= \max[-10 \cdot 0.95 - 10 \cdot 0.05, -100 \cdot 0.95 + 0 \cdot 0.05] \\ &= \max[-10, -95] \end{aligned} \tag{10}$$

Equation 9 shows that if Thomas checks the more diagnostic weather forecast and sees that it is not expected to rain that day, he will not take an umbrella and $EU(S_+) = -5$. If, however, he sees that it is forecast to rain, then Equation 10 indicates that Thomas will choose to take an umbrella out with him and $EU(S_-) = -10$.

The utility of the weather forecast can be calculated by multiplying the difference between $EU(S_+)$ and $EU(S_0)$ by the probability of entering state S+. As Thomas thought it was equally likely to rain as not rain in the above example, the probability of entering each of these futures states was P(0.5). If the diagnosticity of the weather forecast, π, is 0.95, then $EU(S_+) = -5$, $EU(S_-) = -10$, and $EU_{Information} = 0.5 \cdot 5 = 2.5$.

The above calculation prescribes that a decision-maker will always seek information if that information may alter their decision. This is, of course, unrealistic, as the cost of obtaining new information may outweigh the benefits that could be accrued from changing the decision. If there is a cost to acquiring information, which in the real world there invariably will be (even if merely a time or energy cost), then this should be factored into the expected utility equation. As the information cost, *c*, is sunk – the decision-maker cannot retrieve it after deciding to seek information – the cost can be subtracted from the utility of each potential outcome, as in equations 11 and 12, which are adapted from Kobayashi and Hsu (2019).

$$\begin{aligned} EU(S_+, c) = \max[&u(x_1 - c) \cdot (1 - \pi) + u(x_2 - c) \cdot \pi, u(x_3 - c) \cdot (1 - \pi) \\ &+ u(x_4 - c) \cdot \pi] \end{aligned} \tag{11}$$

$$\begin{aligned} EU(S_-, c) = \max[&u(x_1 - c) \cdot \pi + u(x_2 - c) \cdot (1 - \pi), u(x_3 - c) \cdot \pi \\ &+ u(x_4 - c) \cdot (1 - \pi)] \end{aligned} \tag{12}$$

Here the different outcomes under consideration are represented by x_{1-4}, where x_1 reflects the case when Thomas takes the umbrella and it does rain; x_2 when he takes the umbrella and it does not rain; x_3 when he does not take an umbrella and it does rain; and x_4 when he does not take an umbrella and it does not rain.

This is clearly still too simplistic a model to dictate how people should make real-world decisions, as the probabilities and utilities are represented by single numbers (or parameters) rather than distributions that allow for uncertainty in those values, there are only a small number of possible choices and outcomes, and so on. Nonetheless, it demonstrates that if people are aware that they might update their beliefs in response to new information, the utility of information can be calculated according to the degree to which it is expected to improve their decision-making in conjunction with the costs associated with its acquisition.

Source Credibility

Key to the information-seeking calculus above is the principle that decision-makers are forward-looking agents who can estimate the impact that information will have on the utility of their decision. Decision-makers should therefore seek to learn from others whom they believe can provide utility-enhancing information. As illustrated in the previous section, information can serve to increase the utility of a decision by improving the decision-maker's ability to predict the probability of different outcomes occurring. Accordingly, decision-makers should be more motivated to seek information and follow the advice of messengers with highly diagnostic information – in other words, messengers who are deemed as credible.

Theoretical evolutionary analyses of social learning have suggested that fitness benefits are achieved by following social learning strategies, which dictate the conditions under which social information should be relied upon (Boyd & Richerson, 1985; Giraldeau et al., 2002; Rogers, 1988). Most evolutionary scholars have based theories on the assumption that social learning will carry more fitness benefits if agents selectively learn from 'successful' members of the population (Boyd & Richerson, 1985; Henrich & Broesch, 2011; Henrich & Gil-White, 2001; Henrich & McElreath 2003; Laland, 2004). That is, individuals should copy the choices or strategies of those who are receiving the highest payoffs. For example, an animal could choose to copy the foraging patch choice of the individual who seems to be reaping the greatest returns (Schlag, 1998).

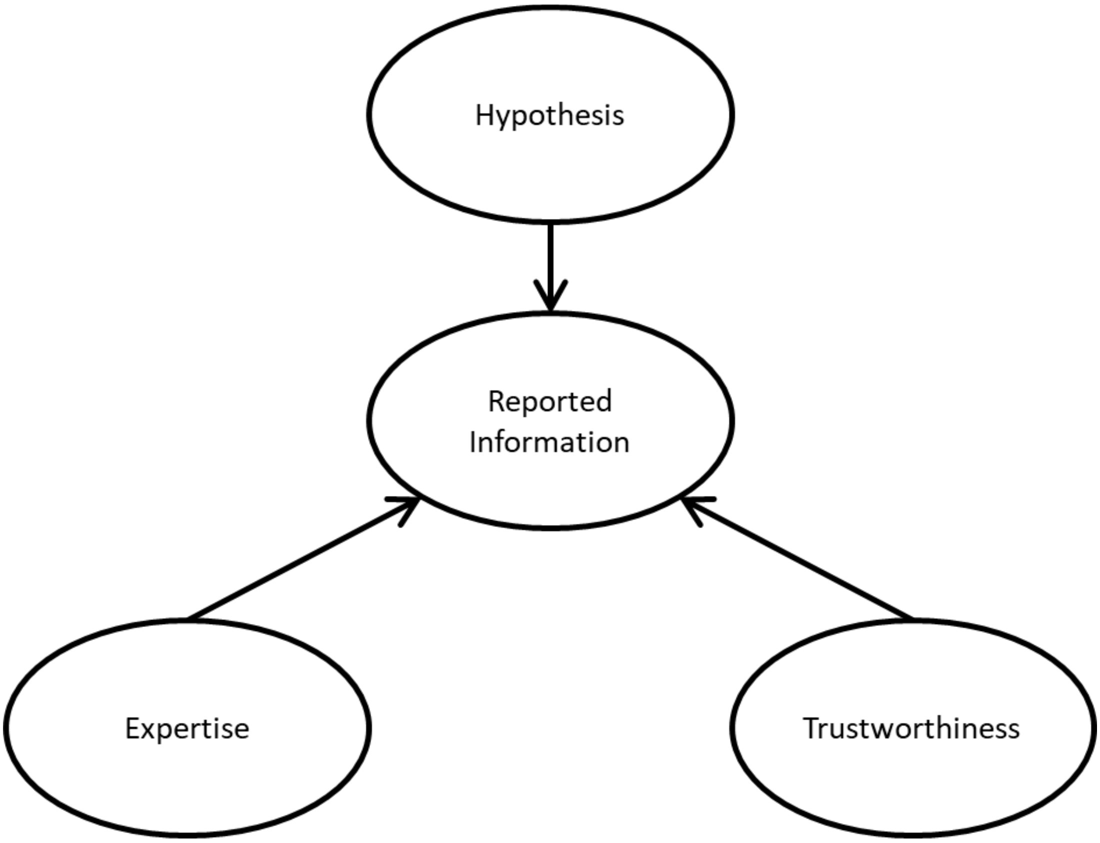
Unlike other species, humans do not have to directly observe the relationships between others' choices and their outcomes in order to estimate the instrumental benefits associated with learning from specific individuals. We can verbally transmit information and use explicit, reportable metacognitive abilities – that is, conscious knowledge about our own and others' cognitive processes – to selectively seek information from those with relevant, valuable information (Heyes, 2016).

When receiving information from others, normative models suggest that Bayesian learners should utilise their beliefs about the messenger's expertise and trustworthiness to evaluate its diagnosticity (Bovens & Hartmann, 2003; Hahn et al., 2009; Hahn et al., 2013; Harris et al., 2016; Madsen, 2016, 2019a, 2019b; see also, Schum, 1981; Walton, 1997). Here, expertise refers to the amount of relevant knowledge the messenger possesses, while trustworthiness refers to the degree that the messenger attempts to convey information honestly. A messenger's perceived credibility can thus be quantified through an amalgamation of their perceived expertise and trustworthiness. The relationships between a messenger's expertise and trustworthiness, as well as the truth of a particular hypothesis, and the information reported by that messenger can be represented within a simple Bayesian belief network, as in Hahn et al. (2013) and Harris et al. (2016) (Figure 1).

In the Bayesian framework, the messenger's perceived credibility, and by extension expertise and trustworthiness, are captured by the ratio of $P(e|h)$ and $P(e|\neg h)$ (i.e., the likelihood ratio). In particular, the likelihood of receiving accurate or inaccurate information regarding the veracity of a proposition will be conditional on the messenger's expertise and trustworthiness. The overall likelihood function represents a generative model of how the messenger's credibility relates to the observed evidence, and can be represented by a linear function, where greater credibility leads to a higher probability of producing accurate information (as in Behrens et al., 2008; Harris et al., 2016; Leong & Zaki, 2018).

Figure 1

A Bayesian Network Representing the Veracity of a Hypothesis, The Information Reported by A Messenger and The Expertise and Trustworthiness of That Messenger



Note. This figure demonstrates that a learner can use a Bayesian model to estimate the probability that a hypothesis is true, given the values of the other elements. If the truth of the hypothesis is already known, the learner can update their beliefs about the messenger's expertise and/or trustworthiness upon receiving the reported information.

This can be illustrated by revisiting the earlier example. A perfectly expert weather forecaster would be able to tell Thomas whether it will rain with 100 percent accuracy. If the expert weather forecaster wanted to intentionally deceive Thomas, they could provide him with information that would certainly prove false. However, without any relevant expertise, the weather forecaster would not be able to provide diagnostic information in either direction. Thus, by estimating a messenger’s expertise and trustworthiness, a learner can make predictions about the utility of the information they can receive from others.[1]

Notably, even a Bayesian learner may not know how credible others are. In the real world, we often lack information about the quality of our evidence – that is, the likelihood ratio might not be known – and if a learner misestimates the credibility of their potential information sources they may make information-seeking choices that lead to inaccurate beliefs and poor decision-making (Hahn et al., 2018). However, like other beliefs, our estimates of others’ credibility are not static; individuals typically receive more and more evidence about what others are like over time and should therefore seek to update their beliefs about them in an appropriate manner. Learners can thus use Bayesian principles to infer others’ expertise and trustworthiness. That is, they can compute the posterior probability that a messenger is credible (e.g., predicts rain with 90% accuracy) by combining their prior belief with the likelihood of the observed evidence. Thus, if Thomas checks the weather forecast and subsequently determines that its prediction was correct, he should update his belief about that forecast’s credibility and be more inclined to check it in the future. His posterior belief will become his prior the next time he observes new evidence, reflecting how our beliefs about others’ credibility evolve over time (Behrens et al., 2008; Leong & Zaki, 2018).

Is Optimality a Reasonable Prescriptive Standard?

Much of the work in decision-making psychology has examined whether people’s behaviour is consistent with the predictions made by normative models. Early evidence suggested that humans are “intuitive statisticians” who unconsciously use Bayesian principles to form intuitive judgments (Peterson & Beach, 1967) and make choices that maximize their expected utility under uncertainty (Edwards, 1954; Newell et al., 1958). This view of human judgment and decision-making came under

[1] Note, the Bayesian model of source credibility advanced by Hahn and colleagues provides a normative framework specifically for assessing the degree to which a messenger’s testimony should be believed. This approach could be generalised to situations where a learner is interested in determining the best course of action to take in a given environment. For example, the likelihood ratio could reflect the relationship between the likelihood of a messenger endorsing a choice given that it is optimal and a choice that is not.

pressure in the 1970s after a series of experiments demonstrated that people systematically violate the laws of logic (Wason, 1968), probability theory (Tversky & Kahneman, 1974), and expected utility theory (Kahneman & Tversky, 1979). The existence of cognitive biases in people's judgements and decisions was taken to suggest that the mind substitutes optimal procedures for fast but fallible cognitive strategies – 'heuristics' that characteristically only use a limited amount of the information available and perform less computation than would be required to compute the statistically optimal function (Tversky & Kahneman, 1981). Examples of such biases include framing effects (Tversky & Kahneman, 1981), anchoring bias (Tversky & Kahneman 1974), base-rate neglect and the conjunction fallacy (Kahneman & Tversky 1972), among many others (Gilovich et al., 2002; Kahneman, 2011). Consequently, people often fail to choose the best available option, given the information they have available, in decision-making tasks (Kahneman & Tversky 1979).

Yet, this view of the mind lies in contrast with the remarkable abilities people display in more basic forms of cognition. For example, to produce vision, the brain infers the intrinsic properties of objects, such as colour, from ambiguous data supplied to the retina. Even though the same object reflects a different spectrum to the eye when it is viewed under different illuminations, the visual system is generally able to provide consistent representations of an object's colour (Brainard & Freeman, 1997). This phenomenon, known as colour constancy, cannot be achieved through deductive or certain inference. Rather, the ambiguity can only be resolved by combining the image data with an accurate probabilistic model of the environment (Brainard et al., 2006). Similarly, to understand what someone is saying from noisy speech data the mind must use its knowledge of language to provide probabilistic constraints on which words are likely to have been uttered in a given context (Chater & Manning, 2006).

Given that the mind can produce Bayesian-like low-level cognition, around the turn of the century researchers began to question whether conscious judgements may also be rooted in Bayesian inference (e.g., Chater & Oaksford, 1999; Cheng, 1997; Tenenbaum, 1999; Tenenbaum & Griffiths, 2001; Oaksford & Chater, 2001). One particularly striking study found that when making hypothetical predictions about quantities and durations, such as how much money movies will make, how long others will live, and how long politicians will remain in office, the median participants' judgments tended to be close to, if not indistinguishable from, ideal Bayesian predictions generated by applying Bayes theorem to empirical prior distributions (Griffiths & Tenenbaum, 2006). Participants in this study appeared to use their prior beliefs (e.g., how much money movies tend make) and the evidence they were given (e.g., how much money the movie had already grossed) to appropriately generate a posterior belief (e.g., how much money the movie would make in total). Notably, their predictions were also sensitive to the type of distribution underlying the values they were judging. For example, predictions about lifespans were seemingly generated from a Gaussian distribution, while those about movie grosses from a power-law distribution, in accordance with the true underlying distribution of each category. These findings are inconsistent with the view that human reasoning is non-probabilistic and incapable of implementing Bayesian-like functions.

That humans can achieve near optimal performance in perception (Knill & Pouget, 2004; Knill & Richards, 1996; Körding & Wolpert, 2004), learning (Fiser et al., 2010; Goodman et al., 2008; Griffiths & Tenenbaum, 2009; Xu & Tenenbaum, 2007), and reasoning and prediction tasks (Hahn & Oaksford, 2007; Frank & Goodman, 2012; Griffiths & Tenenbaum, 2006) has led scholars to reconsider the possibility that the human mind is a probabilistic reasoning machine. Many now argue in favour of a 'Bayesian brain' hypothesis (e.g., Doya et al., 2007; Tenenbaum et al., 2011), which argues that evolutionary dynamics have produced neural and cognitive mechanisms that allow people to generate near optimal solutions to the computational problems they face in their environment. This idea is not so different to earlier conceptualisations of humans as intuitive statisticians. It is important to note this hypothesis does not assume that people consciously conduct Bayesian calculations; it is inconceivable that people keep a multitude of hypotheses in mind and update the probability of each one in accordance with Bayes' theorem. Rather, the theory postulates that psychological processes that have been subject to strong evolutionary pressures over a long period of time are likely to be well-optimised, but conscious reasoning about probability is unlikely to have been shaped by strong selection pressures (Chater et al., 2006; Suchow et al., 2017).

Advocates of the Bayesian brain hypothesis contend that agents cannot be expected to find perfectly optimal solutions to the computational problems that they face – indeed, the calculations required to find them are typically computationally intractable (Simon, 1955, 1956) – but should rather be expected to use strategies to approximate them (e.g., Sanborn & Chater, 2016). An additional caveat is that living beings have limited cognitive capacities – our brains are only so big – due to biophysical and metabolic constraints on information processing (Lieder & Griffiths, 2020). It has thus been suggested that the heuristic mechanisms that humans and non-human animals use to make judgements and decisions may reflect the optimal use of their limited computational resources and time (Gershman et al., 2015; Griffiths et al., 2015).

According to this view, normative models need to account for the fact that people might not have perfect knowledge of the situations that confront them; exhaustive lists of options and future consequences; or the time and cognitive resources to solve computationally complex problems. Under these conditions, a cognitive function (or computer program; Gershman et al., 2015) can be considered bounded-optimal if it maximises performance compared to other computations that it could implement using its available information-processing capacities in a given environment (Gigerenzer, 2008; Lewis et al., 2014; Lieder & Griffiths, 2020; Russell & Subramanian, 1994).

When realistic assumptions about the environmental and cognitive constraints faced by humans are made, many ostensible irrationalities can be reinterpreted as optimal trade-offs between the benefits of increased accuracy and the costs of resource allocation (Bossaerts & Murawski, 2017). For example, computational models of cognitive control that take into account the opportunity costs of performing resource intensive cognitive operations indicate that the mind does a surprisingly good job at performing this cost-benefit analysis (Shenhav et al., 2013; Shenhav et al., 2017). Similarly, decisions over how much information to acquire before making a choice, which have traditionally been characterised as rash, often approximate optimality when time and effort costs are included into normative models (Tajima et al., 2016). As the gains of additional information are often small when performing everyday tasks, utility can be maximized globally by sampling very little on each decision (Vul et al., 2014).

Moreover, Bayesian cognitive models that operate under the assumptions of limited information and cognitive capacity systematically generate examples of classic probabilistic reasoning errors that have been documented in humans. For example, Sanborn and Chater (2016) show that an efficient and scalable implementation of Bayesian inference, which uses sampling to represent relative posterior probabilities, will reproduce the unpacking effect, base-rate neglect, and the conjunction fallacy. A sampling-based Bayesian approach to inference can also explain why individual judgements in Griffiths and Tenenbaum's (2006) study of hypothetical predictions were often far from perfect, even though the median participant's judgement so closely matched the optimal Bayesian solution (Vul et al., 2014). Taken together, these finding highlight the importance of considering resource constraints when assessing the optimality of judgements and decisions.

Bayesian Social learning

A large body of evidence now indicates that cognitive processes relevant to social learning approximate statistically optimal solutions. In particular, studies have demonstrated that people continually track statistics of the environment, including those that relate to a messenger's expertise and trustworthiness, to form accurate representations of those around them and make utility-maximizing social learning decisions (Apesteguia et al., 2007; Behrens et al., 2007; Behrens et al., 2008; Biele et al., 2011; Bonaccio & Dalal, 2006; Boorman et al., 2013; Diaconescu et al., 2014, 2017; Harvey & Fischer, 1997; Leong & Zaki, 2018; Soll & Larrick, 2009; Shafto et al., 2012; Van Swol & Sniezek, 2005; Yaniv & Kleinberger, 2000; Yoshida et al., 2008). For example, humans display a remarkable ability to track both the probability that a choice will produce reward and the probability that a messenger will give accurate advice and combine these two sources of information into an overall probability estimate to determine their choices (Behrens et al., 2008).

The neural data from such studies suggests that learning others' 'informational value' relies on a large network consisting of both domain-general neural mechanisms that track the values associated with different stimuli (i.e. non-social as well as social) and domain-specific mechanisms that appear to be crucial for inferring what is going on in others' minds (Behrens et al., 2008; Behrens et al., 2009; Frith & Frith, 2012; Hackel et al., 2015; Saxe, 2006; Schilbach et al., 2013; Zaki et al., 2016). From an evolutionary perspective, it is unsurprising that our brains should have adapted to the demands of living in social groups; advocates of the 'social brain hypothesis' note that the relative brain size of different species can be explained by the number and complexity of social interactions that they are likely to experience (Dunbar & Shultz, 2007).

A parallel line of research from the field of argumentation has found that people weight testimony from others not only based on the content of the arguments put forward but also the credibility of the messengers delivering those arguments (Bovens & Hartmann, 2003; Hahn et al., 2009; Harris et al., 2016; Madsen, 2016, 2019a). For example, Hahn et al. (2009) presented participants with arguments and systematically manipulated both the source credibility and strength of the argument. That is, the source was either portrayed as credible or non-credible and the argument they put forward was either strong or weak. As expected, both source credibility and argument strength had positive main effects on the convincingness of the message. Crucially, these two factors interacted so that participants found strong arguments put forward by a credible source more convincing than would be expected by an additive model, consistent with the prescriptions of a Bayesian model that accounts for both factors in the likelihood function.

The ability to make interferences about others' task-relevant knowledge emerges early in humans. Research shows that three-year-olds can spontaneously monitor the accuracy of adults' behaviours, by comparing it against their own knowledge, and then use their assessments of expertise when deciding whom to learn from (for a review, see Harris & Corriveau, 2011). For example, in one study 3- and 4-year-olds observed different adults labelling objects (e.g., a cup) either correctly or incorrectly but were not told directly when the adults were correct and when they were not (nor were the adults rewarded for correct answers). When subsequently asked to make judgements about which adult was more competent at the task, the children were able to accurately identify the more accurate adult. They also used their knowledge of the adults' competence to guide their information-seeking decisions. When asked to name unfamiliar objects themselves, they chose to ask the previously accurate adult more often than the previously inaccurate adult. When the two adults both provided answers, and the answers conflicted, the children tended to side with the adult who had been accurate in the past (Koenig et al., 2004; see also, Birch et al., 2008; Corriveau & Harris, 2009; Pasquini et al., 2007; Corriveau et al., 2009; Koenig & Harris, 2005).

In addition to monitoring accuracy, young children also recognise that cues of confidence are indicative of expertise and use verbal confidence cues when deciding whom to learn from. If an adult expresses verbal uncertainty about her testimony (e.g., "I think this is a spoon", when evaluating a spoon-like object), three-years-olds are less likely to believe her than if she simply labels the object declaratively (e.g., "This is a spoon"; Jaswal & Malone, 2007; Sabbagh & Baldwin, 2001). Young children also track and utilise non-verbal confidence cues. In a study examining young children's sensitivity to others' non-verbal cues of confidence, two- and three-year-olds saw a confident looking adult (e.g., displaying facial expressions of recognition and satisfaction) and an unconfident looking adult (e.g., displaying puzzled facial expressions, shoulder shrugging, etc.) perform a task. They subsequently indicated that they thought the adult who exhibited confident non-verbal behaviours knew more than the uncertain looking adult and selectively copied their task behaviours (Birch et al., 2010). These findings suggest that young children possess metacognitive abilities that provide them with an understanding of others' knowledge and use this information to learn from credible sources.

Animal research indicates that some non-human animals also possess adaptive psychological mechanisms that dictate their social learning strategies (for reviews, see Kendal et al., 2018; Laland, 2004). Consistent with evolutionary and normative theories, some species, including rats (Galef et al., 2008), nine-spined stickleback fish (Kendal et al., 2009; Pike et al., 2010), monkeys (Coussi-Korbel & Fragaszy, 1995), and chimpanzees (Kendal et al., 2015), preferentially copy the behaviour of 'successful' individuals. Interestingly, sticklebacks copy the foraging patch choices of others in proportion to the rewards they observe those others receiving (Pike et al., 2010). This is notable because theoretical evolutionary analysis indicates that using a 'proportional observation strategy' is more efficient than always copying the behaviour of the most successful individual in conditions where the information learners receive is unreliable and noisy (Schlag, 1998). Taken together, the results from a broad range of research fields implicate the emergence of cognitive mechanisms that drive individuals to selectively learn from others in a manner that increases the expected utility of the information that is gleaned.

It is important to note here that individuals are not only sensitive to the expertise and trustworthiness of different messengers but also factor in additional costs and benefits associated with different social learning choices. Henrich and Henrich (2010) show that Fijian villagers selectively learn from readily available, low-cost messengers, such as family members, because access to those with more expertise is often limited and learning from them carries relatively larger costs. People are also aware that social learning is inherently interpersonal: they recognise that the questions they ask of others may have an impact on their reputation (Brooks et al., 2015) and their relationships (Schwartz et al., 2011). Moreover, theoretical models suggest that family members, close friends, and in-group members should not only be preferentially relied upon by social learners due to greater ease of access but also because they are likely to share the same environment as the learner (and thus have more relevant information) and feel motivated to share useful and reliable information (Boyd & Richerson, 1985; Kendal et al., 2018; Laland, 2004). Again, empirical findings indicate that both humans and non-human animals behave consistently with the prescriptions of these models. For example, Pavlovian threat learning in mice is enhanced by familiarity and relatedness when observing others displaying avoidance responses (Kavaliers et al., 2005), while human advice-takers are sensitive to preference similarity when utilising advice from others on matters of taste and when making self-predictions about their own future or hypothetical actions (Gino et al., 2009; Yaniv et al., 2011). When seeking preference-relevant information, such as when deciding which restaurant to choose or which movie to see, people assume that the opinions of similar others will be correlated with their own and thus rely more heavily on their advice (Yaniv et al., 2011). Likewise, when trying to make self-predictions about how one would act in a future or hypothetical situation people place greater weight on the advice of similar over dissimilar others because they believe the advice provided by those like them is more predictive of how they would act in the same situation (Gino et al., 2009).

The Structure of Impressions

The capacity constraints on human cognition lead to a reliance on simplified cognitive frameworks (Collins & Quillian, 1969), and it is argued that knowledge about others may be stored in memory in the form of hierarchically structured belief networks (Diaconescu et al. 2017; Diaconescu et al. 2014; Hackel et al., 2015; Hastie & Kumar, 1979). These networks represent beliefs as nodes in a hierarchical network, which are connected and influence each other. In a belief network that represents a particular person, the 'person node' (e.g., Jayne) would be at the top. In the level below this would be nodes representing higher-level beliefs about that person, such as their key social characteristics (e.g., competence, warmth). Each of these, in turn, are connected to several lower-level beliefs, such as memories about observed behaviours (e.g., asked an intelligent question at a talk). Thus, observers encode more than just specifics of a particular social interaction, they also infer the abstract and enduring traits of the person. For example, one may spontaneously infer that a person is untrustworthy after watching them lie or infer that they are clever after observing them solve a complex puzzle (Uleman et al., 2012). Categorising others at the trait level, rather than representing them in terms of specific behaviours, tendencies, and skills, allows the mind to organise knowledge in an efficient structure and make predictions about others' future behaviours across contexts (Heider, 1958).

Recent neuroimaging work supports the notion that people learn about behaviours and traits at different levels of a hierarchically structured belief system and suggests that people do so in a manner that approximates Bayes-optimality (Diaconescu et al. 2017; Diaconescu et al. 2014). Participants in these studies received advice from another person whose incentives to provide accurate information varied throughout the task. Thus, to make utility-maximizing choices, participants had to infer both the advisor's expertise and their current intentions (i.e., to help or deceive them) from the feedback they had previously received. Participants' decisions of whether to trust the advisor were best explained by a hierarchical Bayesian model, which represented trial-specific observations of the advisor's accuracy at its lowest level and uncertainty about the advisor's trustworthiness at a higher level. Furthermore, prediction-errors (PEs) at different levels of the hierarchical network were associated with activity in different brain regions. Specifically, low-level PEs, which represented the difference between participants' expectations about the advice accuracy and actual accuracy of the advice, were related to activity in the midbrain, whereas high-level PEs representing the difference between their expectations about the advisor's changing intentions and the actual volatility in the advisor's intentions were associated with activation in the cholinergic basal forebrain (Diaconescu et al. 2017). These findings support the hypothesis that people employ an efficient approximation to ideal Bayesian inference (see Mathys et al., 2011) to learn about both low-level observed behaviours and important higher-level mental states.

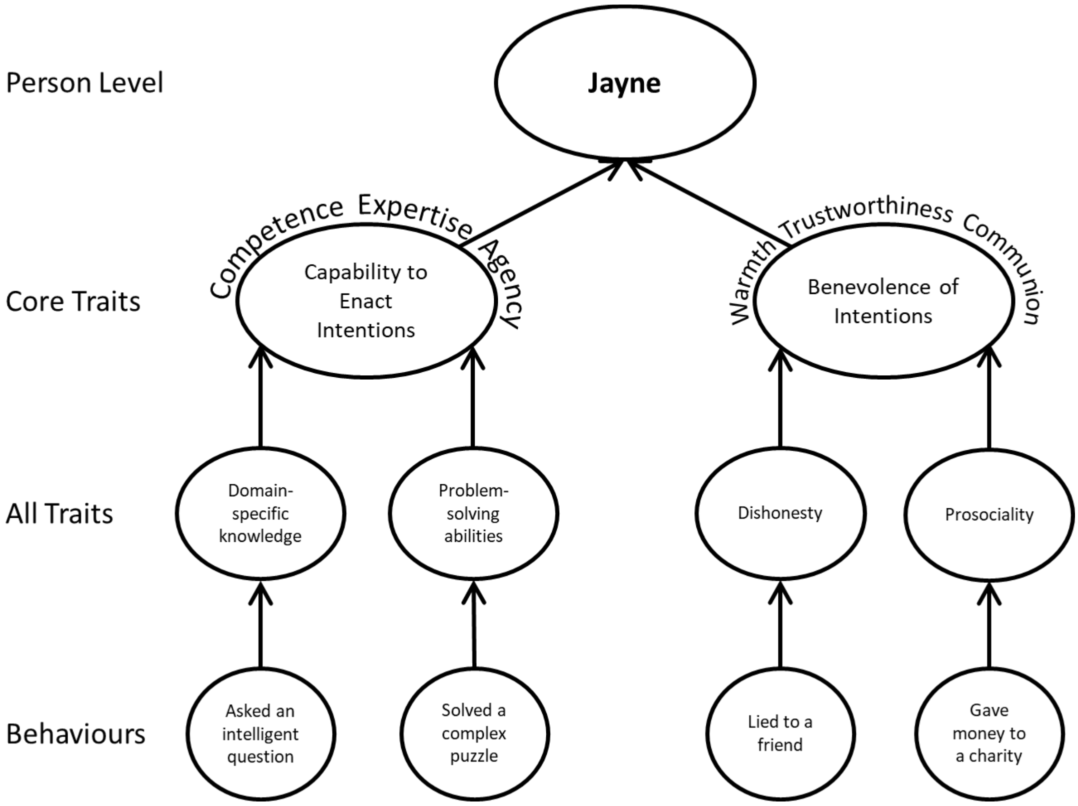
Given the vast number of characteristics that could be inferred from the data we have about others, impression researchers have used dimension reduction techniques, such as principal components analysis, to identify the core traits that underlie social evaluations (e.g., Cuddy et al., 2008; Oosterhof & Todorov, 2008; Rosenberg et al., 1968). These studies are premised on the fact that ratings of different traits (e.g., warm, generous, friendly) tend to cluster together, consistent with the theory that high-level latent factors structure how people store knowledge about others (Figure 2). The findings from this line of research indicate that humans make inferences about a small number of traits that have functional significance.

Different models have proposed that different traits are central – Abele and Wojciszke (2014) list a number of overlapping two-dimensional models – but there is a common core between them. In particular, they suggest that people perceive others in terms of the benevolence of their intentions (e.g., warmth, communion, trustworthiness, social goodness) and their capability to pursue their intentions (e.g., competence, agency, dominance, intellectual goodness) (Bakan, 1966; Koch et al., 2021).

The two core dimensions of person-perception account for the majority of the common variance in ratings of different traits (Abele & Wojciszke, 2007; Oosterhof & Todorov, 2008; Wojciszke et al., 1998) and are related to specific emotional responses and behavioural tendencies (Cuddy et al., 2008). For example, in the terms used by Cuddy et al. (2008), those who are perceived as high on competence and warmth elicit admiration and facilitating behaviours, whereas those perceived to be low on both dimensions elicit contempt and harmful behaviours. The functional explanation for the existence of these core dimensions is that they help people to identify actors who may be relevant to the pursuit of their goals and those who may be willing to help or hinder them in achieving those goals (for a review, see Koch et al., 2021). These dimensions thus align well with the components of the Bayesian source credibility models mentioned earlier (Bovens & Hartmann, 2003; Hahn et al., 2009; Harris et al., 2016), with trustworthiness reflecting the benevolence of a messenger’s intentions and expertise reflecting their capability to enact their intentions. This suggests that the key factors driving people’s impressions of others are the same as those that normative models argue should be relied upon when receiving information from others.

Figure 2

A Hierarchically Structured Belief Network Representing Social Evaluations of a Person, Jayne


Person Level
Jayne
Core Traits
Competence Expertise Agency
Capability to Enact Intentions
Warmth Trustworthiness Communion
Benevolence of Intentions
All Traits
Domain-specific knowledge
Problem-solving abilities
Dishonesty
Prosociality
Behaviours
Asked an intelligent question
Solved a complex puzzle
Lied to a friend
Gave money to a charity

Note. Specific observed behaviours are positioned at the lowest level of the network and influence beliefs about specific traits. Beliefs about individual traits drive beliefs about two core higher-level trait beliefs. The two core higher-level trait beliefs have been given different labels in past works, but one reflects the capability to enact intentions (labelled: competence, expertise, agency, etc.) while the other the benevolence of intentions (labelled: warmth, trustworthiness, communion, etc.).

Social Learning Biases

While there is now a large body of evidence to suggest that learners can effectively monitor and effectively utilise cues of credibility, certain learning biases have also been reported in extant works. Several explanations of why such biases might exist have been proposed and supported by experimental results. Three examples are discussed below. This is by no means an exhaustive list, it merely serves to highlight that biases may arise for many reasons, including a tendency to overweight evidence that confirms prior beliefs (Boorman et al., 2013; Leong & Zaki, 2018), a motivation to preferentially integrate desirable information about similar others into one's beliefs (Hughes et al., 2017), or limited access to information and computational resources (Henrich & Broesch, 2011). Before discussing these examples in more detail, it is worth noting that while deviations from optimality are referred to here as 'biases' for both ease and consistency with the extant literature, Hahn and Harris (2014) caution that most research on bias falls short on at least one of the three criteria that they identify as needing to be met to establish the presence of a costly cognitive bias.

Confirmation Bias in Social Learning

Erie Boorman and colleagues (Boorman et al., 2013) demonstrated that people judge others' expertise according to both the accuracy of their predictions and the degree to which those predictions match their own. In this study, participants were asked to evaluate financial assets while also observing the judgments made by others before receiving feedback. The findings indicated that participants updated their beliefs about the asset (based only on the asset's past performance) in a similar manner to a Bayesian model that took the varying volatility of price changes into account, with the Bayesian model predicting 80% of participants' predictions of the asset's movements. This is consistent with previous work on asocial Bayesian learning in humans when predicting reward likelihoods (Behrens et al., 2007, Boorman et al., 2011). However, when participants updated their beliefs about other people's expertise, they did so in accordance with the predictions of an adapted Bayesian learning model. The adapted model better explained participants' expertise learning than an optimal Bayesian inference model because participants took into account their own judgment about the asset when updating their assessment of the other participant's ability on the task rather than simply relying on the outcome feedback (see also, Hahn et al., 2018). Participants gave considerable credit to people for correct judgements with which they agreed, but barely gave them any credit at all for accurate judgments with which they disagreed. Notably, the participants did not exhibit this bias when they were shown predictions made by an algorithm, suggesting that there is a social specificity to this effect.

Another set of studies adapted the financial advice-taking task used by Boorman et al. (2013) to examine how people learn from and utilise advice from others (Leong & Zaki, 2018). As in the previous study, participants learned about how accurate advisors were when predicting fluctuations in financial assets. And, like the previous study, participants updated their beliefs about others' expertise in accordance with the predictions of an adapted Bayesian learning model. In this study, an adapted model outperformed an optimal Bayesian model because it assumed participants would preferentially learn about others' expertise from evidence that was consistent with their prior expectations. In particular, the likelihood function in the researchers 'confirmation bias' model reflected the weighted combination of the likelihood of the observed outcome (i.e., whether an advisor was actually correct or not) and that of the expected outcome (i.e., whether the participant expected the advisor to be correct or not). Given that participants had high prior beliefs in the advisors' accuracy and underweighted evidence that conflicted with their beliefs, they remained more optimistic about the advisors' expertise than was warranted by the data. After learning about each advisor's accuracy on the task, participants were subsequently given the opportunity to utilise advisors' recommendations when predicting changes in the price of financial assets. Overestimation of the advisors' expertise led them to rely on the advice more heavily than they should have. For example, participants were influenced by an advisor whose advice was non-diagnostic, suggesting that they mistakenly believed the advice provided useful information, even though they had seen considerable evidence that it was not. The combination of optimistic prior beliefs about others' expertise and a confirmation bias in how those beliefs are updated resulted in a tendency for participants to rely on information from others more than they should have (although it is worth noting that there is considerable evidence from other studies to indicate that people prefer to rely on their own information than on socially acquired information: Eriksson & Strimling, 2009; Morgan et al. 2012; Heyes, 2012; Rieucau & Giraldeau, 2011; Toelch et al., 2014; Yaniv & Kleinberger, 2000).

Like many other biases in impression formation, the tendency to have optimistic prior beliefs about others is posited to derive from an adaptive mechanism. Previous research into the psychology of communication suggests that people have a ‘truth bias’ – an inclination to believe and trust others, even though this makes them vulnerable to deception (McCornack & Parks, 1986; but see also, Masip et al., 2009). This is an adaptive strategy in environments where people are honest most of the time, as it facilitates efficient communication, social learning, and cooperation (Baier, 1986; Boseovski, 2010; Hardin, 1993; Levine, 2014). Similar reasoning can be applied to beliefs about others’ expertise. Leong and Zaki (2018) argue that in environments where others tend to possess diagnostic information and a willingness to share knowledge, a general disposition to value and utilise advice would help people to form accurate beliefs.

Similarity Bias in Social Learning

Social learning may also be biased by motivational factors (Ames & Fiske, 2013). The theory of motivated cognition suggests that a person’s goals and needs bias their thinking towards desirable conclusions, because doing so allows them to feel validated, maintain a high sense of control, and reduces cognitive dissonance (Kunda, 1990; Taylor & Brown, 1988). As people’s identities are defined partly by the social groups to which they belong (Tajfel & Turner, 1986), humans appear to not only evaluate themselves more positively than warranted by the evidence before them but also to do the same for their relationship partners, friends, and fellow group members (Brewer, 1999; Dovidio & Gaertner, 2010). This may explain why people negatively update their impressions of out-group members and non-group members (i.e., control targets), but not in-group members, after being presented with a mix of positive and negative information about them (Hughes et al., 2017). In Hughes et al.'s (2017) study, functional Magnetic Resonance Imaging (fMRI) data revealed that weaker activity in brain regions involved in impression updating was associated with reduced learning from negative information about in-group members, suggesting that the motivation to maintain favourable opinions of in-group members results in a failure to encode negative information about them.

This finding is consistent with a vast literature in social psychology documenting the effects of perceived similarities on interpersonal attraction and influence (Byrne, 1969; Cialdini, 2001; Cialdini & Trost, 1998). Early studies exploring the sales process found that the level of similarity between the salesperson and client affected the outcome of the interaction (e.g., Brock, 1965). More recent work has shown that even superficial similarities, such as shared birthdays, nationalities, first names, or favourite sports team can affect how much a person is liked and how much influence they have over another's judgements and behaviour (Burger et al., 2004; Miller et al., 1998; Levine et al., 2005). Researchers in this field argue that the need for belongingness, defined as a "need to form and maintain strong, stable interpersonal relationships" (Baumeister & Leary, 1995, p.497), leads people to judge those whom they think may serve as good coalitional partners especially positively and to generally acquiesce to their requests (Brewer & Caporael, 2006). Cooperation requires trust and reciprocity between group members (Trivers, 1971); thus, cues of similarity and shared group membership are theorised to bias our impressions and interactions with others (Brewer & Caporael, 2006).

In the domain of advice-taking, the level of perceived similarity between an advisor and learner has been shown to induce a momentary subjective feeling of certainty in the learner that makes them more receptive to the advice (Faraji-Rad et al., 2012; Faraji-Rad et al., 2015). Thus, similar advisors have more influence on learners' judgements and decisions than dissimilar advisors (Faraji-Rad et al. 2012). Faraji-Rad et al. (2015) propose that perceived similarity facilitates mentalizing and thus boosts the perceived diagnosticity of the advice being proffered.

Halo Effects in Social Learning

Still other social learning biases are posited to stem from limitations imposed on the learner by cognitive and environmental constraints. People cannot conceivably attend to the behaviour of all the individuals in their community and holding information about others in memory is cognitively taxing (Byrne & Whiten, 1988). The knowledge about others that people do accrue thus fades easily (Hastie & Kumar, 1979). Moreover, information that is acquired second-hand (e.g., through gossip) is often distorted and therefore noisy and unreliable (Boyd & Richerson, 1985; Gilovich, 1987).

When an individual does not have information about another's task-relevant expertise, what should they do? Evolutionary theorists have suggested that they may employ a simpler heuristic whereby they selectively learn from others who show signs of general life success, such as cues of health, status, or reproductive success (Boyd & Richerson, 1985; Laland, 2004). The logic underlying this 'copy-successful-individuals' strategy, or 'prestige bias', is that successful people likely possess utility-enhancing knowledge and skills that are worth learning. This heuristic does not require learners to differentiate between the factors that directly led to the individual's success and those that did not, account for the role of luck in success, or determine whether the individual possesses accurate or useful information in the particular domain of interest. Nonetheless, there is a substantial body of evidence to suggest that people are more likely to listen to and copy 'prestigious' individuals in areas unrelated to those where they achieved success (see below). Theoretical analysis suggests that the 'copy-successful-individuals' social learning strategy will on average help populations to acquire adaptive knowledge, albeit more slowly than a 'copy-task-relevant-experts' strategy, but might also result in neutral or maladaptive information being transmitted through communities (Boyd & Richerson, 1985). Everyday examples of this in the real-world include the outsized influence of celebrities who are famous for singing, dancing or playing sports on their audience's political views, attitudes towards brands, and beliefs about the safety of vaccines (Martin & Marks, 2019).

Seminal research in social psychology attests to the effects that perceived status has on social learning (Berger et al., 1980; Berger et al., 1972; Cialdini, 1984). A classic example is Monroe Lefkowitz's jaywalking experiment. Lefkowitz et al. (1955) arranged for pedestrians waiting at a red light to see a man jaywalk across the road while no cars were crossing. The jaywalker's clothes were experimentally manipulated between conditions to induce different perceptions of status; in some instances, the man wore a suit, in others he wore denim. The results revealed that three times as many pedestrians followed the man across the road when he wore a suit than when he was dressed in more casual clothing. Even though there is unlikely to be a causal relationship between socio-economic status and the ability to safely cross a road at a red light, bystanders were receptive to cues of status when deciding whether to copy the jaywalker's behaviour.

Interestingly, high-status clothing does not only affect higher order decision processes but also influences early-stage cognition. In a study employing an eyetracking device, participants' eye movements were recorded as they were shown pictures of different men and women on a computer screen (Maner et al., 2008). Some of the people were dressed in suits while others were wearing casual clothing. In the first few seconds after the stimulus onset, presumably before participants had had time to consciously decide which pictures to attend to, their eyes were selectively drawn to the high-status men. Participants were no more likely to attend to women in suits than they were women in ordinary clothing; rather their eye movements were biased by the women's attractiveness. Consistent with evolutionary theory, this study supports the notion that individuals possess relatively automatic, lower-order processes that predispose them to acquire information from others based on perceived status characteristics (for similar examples in non-human animals, see Deaner et al., 2005; Shepherd et al., 2006).

These findings are not limited to Western, educated, industrialised, rich and democratic (WEIRD) societies. Data from small and remote communities in Fiji show that people living in small-scale societies – where individuals turn to other members of the community for information rather than books, television or the internet – use perceptions of both task-relevant success and cross-domain success to decide whom to seek information from (Henrich & Broesch, 2011). For example, villagers reported that if they had a question about growing yams, they would ask people whom the researchers had previously noted had a history of success in this area. But being a successful yam grower also increased the likelihood a person would be asked a question about fishing (when fishing knowledge and success were controlled for in a regression model). The same was true for other domains; for example, being a successful yam grower increased one's chances of being asked a question about medicinal plants by 2.5 times.

It thus appears that social learning heuristics that are generally adaptive may carry across to situations where they are not. A plausible explanation for this is that inferences on low-level characteristics in a hierarchically structured belief network are influenced by inferences on superordinate characteristics. Thus, if a person finds reason to believe that a messenger is generally competent, they will probabilistically update their beliefs about that messenger's specific skills and task-relevant expertise. Thus, there would be bidirectional influences between the higher-level and lower-level traits in the hierarchically structured belief network shown in Figure 2. This is consistent with how people represent knowledge about non-social categories (e.g., Gelman, 1988; Rips, 1975; Sloman, 1993; Osherson et al., 1990). It also aligns well with findings in social psychology demonstrating that impressions about others formed in one domain spread to other domains.

The tendency for evaluations of certain traits to influence evaluations of other traits is a well-documented phenomenon, known as the 'halo effect' (Dion et al., 1972, Nisbett & Wilson, 1977, Thorndike, 1920). The term was coined by the behaviourist Edward Thorndike (1920) in a paper in which he analysed managers' ratings of their employees and evaluations of army officers by their superiors. He noted that the managers clearly distinguished good and bad employees but tended to rate their favoured employees positively, and their dispreferred employees negatively, on unrelated traits. For example, if an employee were believed to be highly friendly, managers would tend to view them positively on other positive traits such as industriousness, intelligence and trustworthiness. Thorndike posited that managers were "affected by a marked tendency to think of the person in general as rather good or rather inferior and to color the judgments of the qualities by this general feeling" (Thorndike, 1920, p.25). That is, he believed that the high degree of correlation between different trait ratings reflected managers' proclivity to rely on their general impressions of individuals when assessing them on specific attributes.

While Thorndike merely presumed that the degree of intercorrelation between trait ratings was higher than it should have been, subsequent research has validated this claim in three key ways. First, researchers have experimentally manipulated the order in which different characteristics of a person are presented (Asch, 1946; Gräf & Unkelbach, 2016; Harari & McDavid, 1973; Hendrick & Costantini, 1970). This body of research has demonstrated that impressions formed based on earlier presented traits influence the interpretation or expectation of later presented traits. Second, raters' judgements have been compared to the self-reported ratings (or expert judgements) of the person being evaluated (e.g., Fisicaro & Lance 1990; Segal-Caspi et al., 2012). Evidence for halo effects is reported to emerge when, for example, raters' judgements of others' traits are correlated with the attractiveness but not the self-reported traits of those being evaluated (Segal-Caspi et al., 2012). Third, participants have assessed others on characteristics, such as humorousness, with or without visual access to the person they are rating (e.g., Cowan & Little, 2013; Forgas, 2011). For example, Cowan and Little (2013) found that humorousness was related to attractiveness when participants viewed videos of people talking about which items they would take to a desert island, but the correlation disappeared when participants listened to the audio-only versions of these clips. Humorousness and attractiveness are conceptually unrelated, however these findings demonstrate that physical attractiveness influences how funny people find others, providing evidence of a halo effect.

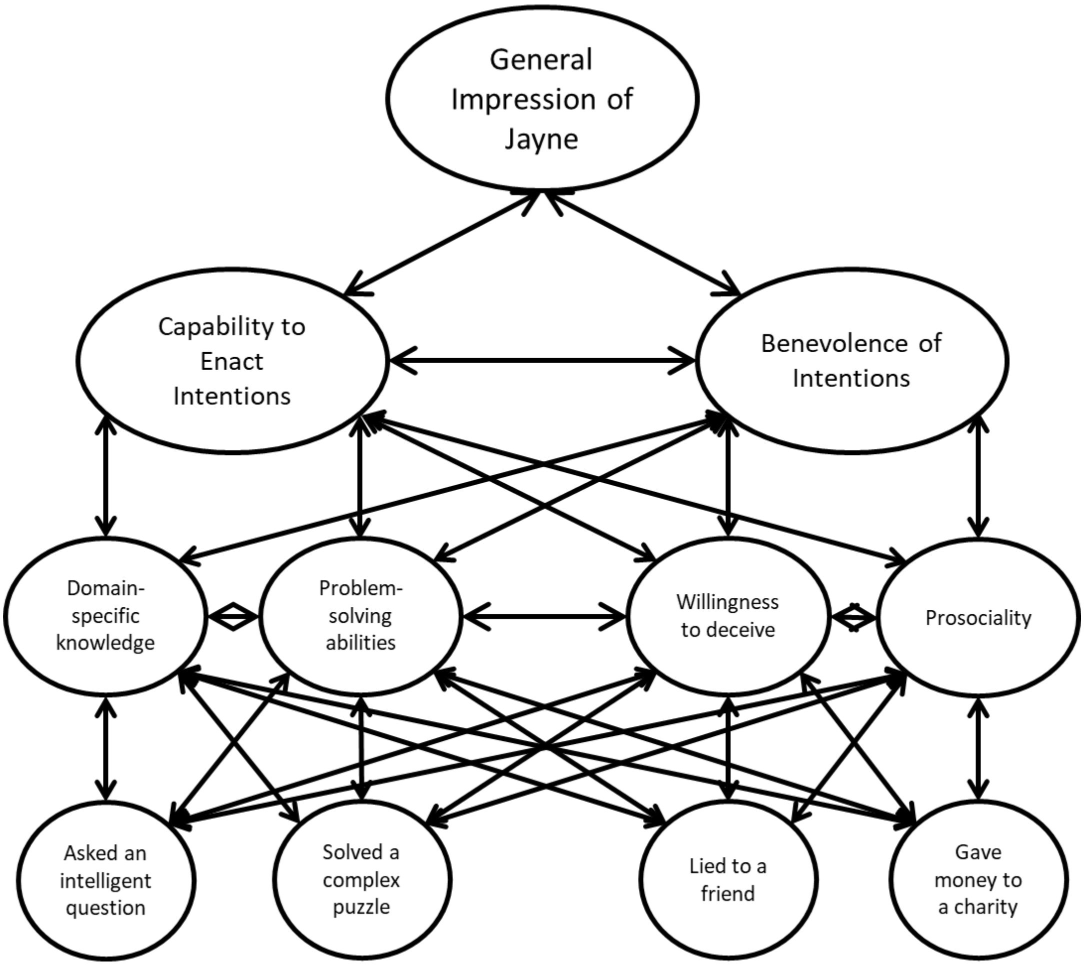
Fisicaro and Lance (1990) show that three causal models of the halo effect are mathematically distinguishable: the 'General Impression' model, the 'Salient Dimension' model, and the 'Inadequate Discrimination' model (see Figure 3). Thorndike's explanation of the halo effect has come to be known as the General Impression model (Fisicaro & Lance, 1990; Gräf & Unkelbach, 2016). According to this model, beliefs about specific characteristics are influenced by the perceiver's general impression of the messenger. Thus, if a perceiver learns that a messenger possesses a particular characteristic, they will form an impression of that messenger’s general character, which in turn influences their beliefs about other, possibly unknown and unrelated, characteristics. In this model, the common causal effect of the perceiver’s general impression of the messenger serves as the basis for the halo effect.

In contrast, in the Salient Dimension model, beliefs about a salient messenger characteristic directly influence beliefs about other, less salient characteristics. This model fits well with the notion that characteristics at higher levels of a hierarchically structured network have a directional influence on those below them. This model also accounts for implicit assumptions about trait co-occurrences, or as Berman and Kenny put it, “how traits and behaviours go together” (Berman & Kenny, 1976, p.263). For example, if a person believes that generous people tend to also be trustworthy, then it is consistent with both associationist and Bayesian accounts for them to update their beliefs about a messenger’s trustworthiness upon observing that messenger act generously (see Figure 3). The degree to which each of the two traits influence each other will depend on the direction and strength of the relationship between them (Orehek et al., 2010). Thus, the salient dimension model argues that halo effects may derive from folk theories, which may or may not be true, about the hierarchical structure and intercorrelations of different traits.

The third model, termed the Inadequate Discrimination model, implies that halo effects may derive from an inability to define the boundaries of conceptually distinct and potentially unrelated traits. According to this model, halo effects occur when beliefs about a messenger characteristic are informed by observed behaviours that should not map onto said characteristic (see Figure 3). The difference between this model and the Salient Dimension is that the Inadequate Discrimination model assumes that the influence of an observed behaviour on an unrelated trait is not mediated by its effect on a related trait. Modelling work that used students’ ratings of lecturers on multiple dimensions and parameter estimation techniques to compare the fit of these three models to participants’ ratings revealed that the General Impression and Salient Dimension models explained a substantial proportion of the variance and that the former provided the most parsimonious explanation of the data (Lance et al., 1994). However, it quickly becomes clear that comprehensively using path analysis to investigate the causal mechanisms underlying halo effects is a tricky business once one starts mapping all the potential pathways through which halo effects may arise, as in Figure 3.

Figure 3

A Hierarchically Structured Belief Network Representing Social Evaluations of a Person, Jayne


General Impression of Jayne
Capability to Enact Intentions
Benevolence of Intentions
Domain-specific knowledge
Problem-solving abilities
Willingness to deceive
Prosociality
Asked an intelligent question
Solved a complex puzzle
Lied to a friend
Gave money to a charity

Note. There are bidirectional relationships between every node in each level of the hierarchy and those in the level below, and bidirectional relationships between the different traits within each level. The graph illustrates that halo effects could emerge through many pathways.

The strength of a halo effect may depend on the context in which beliefs about a messenger are formed or updated. The cognitive effort that goes into evaluating a person influences how likely perceivers are to exhibit halo effects (Hendrick & Costantini, 1970; Jones, 1990). When high attention is maintained while learning about different traits the halo effect is attenuated or eliminated compared to when attention to another's traits progressively decreases as more evidence is observed (Hendrick & Costantini, 1970). The mood of the perceiver also plays a role in determining the strength of halo effects, as positive mood tends to induce a rapid and less attentive style of processing, and consequently increases the prevalence of halo effects, compared to negative mood (Forgas, 2011).

Gräf and Unkelbach (2016) argue that the underlying process by which halo effects emerge also varies according to contextual factors. They show that the valence of the information presented to perceivers influences the degree to which general impressions account for halo effects. Specifically, participants in their studies were more likely to form a positive general impression of a person after receiving one piece of positive information than they were to form a negative general impression after receiving one piece of negative information. The researchers suggest that this is because positive messenger characteristics are more similar to each other – that is, they tend to cluster much more densely – than are negative characteristics, rather than because learning about positive traits boosts the perceiver's mood. They reason that learning about a positive trait consequently tends to activate a broad network of other traits, while learning about a negative trait tends to activate a more localised network.

Halo effects have an influence on important real-world judgements and decisions. For example, they affect how people evaluate each other in various contexts, including the workplace (Frone et al., 1986; Holzbach, 1978; Zysberg & Nevo, 2004), education (Abikoff et al., 1993; Dennis, 2007; Nisbett & Wilson, 1977), and clinical settings (Michelson et al., 1985; Mumma, 2002). Much of the research in social psychology has focused on the influence of physical attractiveness on beliefs about other characteristics. Indeed, there is now a large body of literature indicating that attractive people are perceived as more intelligent, trustworthy, happy, and successful than those that are less attractive (Dion et al., 1972; Eagly et al., 1991; Griffin & Langlois, 2006), and that information is judged more positively if it originates from an attractive than unattractive messenger (Harari & McDavid, 1973; Landy & Sigall, 1974). However, halos are not only granted to people based on their attractiveness (e.g., Deska et al., 2018; Forgas, 2011; Paulhus & Morgan, 1997). The three halo effect models proffered by Fisicaro and Lance (1990) provide potential explanations for why people might prefer to learn from messengers who are generally perceived to possess desirable traits that have no bearing on the actual utility of the information they have to offer (Maestripieri et al., 2017).

Chapter Summary and Aims of This Thesis

Identifying biases in judgements and decisions is a useful scientific practice. If systematic differences between what people should do and what they actually do are observed, we can try to understand why they might exist and thus gain a more complete understanding of how decisions are made. We can also try to quantify the extent of these biases and the harm they are doing to individuals, groups and societies. Further, by understanding why such biases exist, we can try to find ways to improve people's decisions.

As reviewed above, the extant literature suggests that people are remarkably adept at learning about others and using assessments of credibility to inform their social learning decisions. However, due to the capacity constraints on human cognition, and the limited access to and costs of acquiring information about others, the mind stores beliefs about others' characteristics in an efficient structure (e.g., Figure 2) that generally allows for accurate and functionally significant inferences but is also prone to producing systematic errors. In particular, people's beliefs about specific messenger characteristics influence their beliefs about other characteristics to a greater degree than is warranted.

Of particular concern is the impact that halo effects might have on beliefs about others' credibility. To make good use of others' knowledge, individuals must estimate the expected utility of the information that can be obtained through social learning. Only then are they able to decide when to seek information from others, whom to turn to for said information, and how to combine it with their own beliefs. If people who possess desirable characteristics are perceived as more credible in domains that are unrelated to those affected by said characteristics, learners may turn to them for information and advice when they ought not to, as when Fijian villagers seek information about medicinal plants from successful yam growers (Henrich & Broesch, 2011). As social learning informs people's knowledge of the world and guides their choices, systematic errors of this kind may result in the formation of inaccurate beliefs and costly mistakes.

Although there has been a fair amount of research examining how halo effects influence beliefs about others' credibility (e.g., Forgas, 2011; Palmer & Peterson, 2016), less is known about whether such biases are maintained when learners are confronted with disconfirming evidence. Standard theories of learning predict that people update their beliefs upon receiving new information. Thus, after observing evidence to suggest that a competent yam grower knows little about medicinal plants, villagers should update their impressions and avoid seeking such information from them in the future. Moreover, humans can learn about others' competence through word of mouth, so outcome feedback can lead to expertise learning in people who did not observe the evidence first-hand. As gossip is common in human societies (Dunbar, 2004), serves to convey social information about third parties so that group-members can learn about each other in the absence of direct interaction (Feinberg et al., 2012), and facilitates group cooperation (Feinberg et al., 2014), knowledge of others' expertise should be expected to spread through social networks and the impact of the halo effect on perceptions of credibility should therefore be limited. On the other hand, if the halo effect not only biases how people perceive others but also how they learn from social evidence, inaccurate beliefs about others' credibility may be maintained in the face of reality. That impression updating deviates from optimality, in particular when people learn about others who share their beliefs (Boorman et al., 2013) and group status (Hughes et al., 2017), indicates that this may be the case.

Although credibility assessments should comprise beliefs about expertise and trustworthiness, this thesis focuses specifically on expertise learning. Expertise is typically considered more relevant to advice utilisation (for a review, see Bonaccio & Dalal, 2006) and influence (Briñol & Petty, 2009; Chaiken & Maheswaran, 1994; Cialdini, 1984; Dolan et al., 2012; Petty & Cacioppo, 1984), and features more prominently in evolutionary theories of social learning (Boyd & Richerson, 1985; Henrich and Broesch, 2011; Henrich & Gil-White, 2001; Henrich & McElreath 2003; Kendal et al., 2018; Laland, 2004; Schlag, 1998). This may be because a source with no task-relevant knowledge has no instrumental value regardless of how trustworthy they are, as theoretically they are unable to help or hinder others even if they wish to do so (Harris et al., 2016). In contrast, trustworthiness is typically considered more relevant to cooperation and group cohesion (Berg et al., 1995; Camerer & Weigelt, 1998; Rilling & Sanfey, 2011; Rousseau et al., 1998). Moreover, expertise is less personal and subjective than trustworthiness; intentions to help or hinder others are subject to changes in individual motivations whereas knowledge is not (Behrens et al., 2008). This latter point suggests that expertise learning may be less susceptible to learning biases than trustworthiness learning, as learners are not required to infer others' motives, and thus provides a more conservative test of our hypothesis.

There are four main aims of this thesis. The first is to investigate whether learning about messenger characteristics that are unrelated to the task at hand biases expertise learning and, consequently, social learning decisions. To this end, Chapter 2 details the results of two studies that employed a novel paradigm to explore whether learning about others' political beliefs interferes with the ability to learn about and utilise their task-relevant expertise in an unrelated domain. The second aim is to contribute a mechanistic account of the computational processes by which irrelevant messenger characteristics might influence expertise learning and information-seeking choices. Chapter 3 describes two studies in which the experimental paradigm is adapted to allow for the dynamic tracking of participants' beliefs about others’ expertise. Surprisingly, the results of these studies conflicted with those reported in Chapter 2. The third aim is thus to reconcile these contrasting findings. Chapter 4 outlines two studies exploring whether certain adaptations to the paradigm influence the degree to which learning about irrelevant messenger characteristics biases expertise learning and social learning decisions.

Chapter 2

Chapter Overview

On political questions, many people prefer to consult and learn from those whose political views are similar to their own, thus creating a risk of echo chambers or information cocoons. In this chapter, we test whether the tendency to prefer knowledge from the politically like-minded generalises to domains that have nothing to do with politics, even when evidence indicates that politically like-minded people are less skilled in those domains than people with dissimilar political views. Participants had multiple opportunities to learn about others' (1) political opinions and (2) ability to categorise geometric shapes. They then decided to whom to turn for advice when solving an incentivised shape categorisation task. We find that participants falsely concluded that politically like-minded others were better at categorising shapes and thus chose to hear from them. Participants were also more influenced by politically like-minded others, even when they had good reason not to be. These results replicate in two independent samples. The findings demonstrate that knowing about others' political views interferes with the ability to learn about their competency in unrelated tasks, leading to suboptimal information-seeking decisions and errors in judgement. Our findings have implications for political polarisation and social learning in the midst of political divisions.

Introduction

To make good choices, human beings turn to one another for information (Gino et al., 2012; Hofmann et al., 2009; Schrah et al., 2006; Yaniv and Kleinberger, 2000). When selecting a retirement plan or deciding whether to grab an umbrella on the way out, people are motivated to get information from the most accurate source. Obviously, people would prefer to receive a weather report from the weather forecaster whose predictions are 80% correct than from the one who is wrong every other day.

At the same time, people also prefer to receive information from others who are similar to themselves. Democrats are more likely to turn to CNN for their news and

Republicans to Fox News for their daily updates (The Pew Research Center, 2009). This is partly because people assume that like-minded people are more likely to be correct – a phenomenon that can lead to echo chambers (Del Vicario et al., 2016; Sunstein, 2017). But if people had clear and repeated opportunities to learn who is right and who is wrong, would similarity interfere with the ability to learn about accuracy?

It has been suggested that people assess others' expertise based on their own beliefs (Boorman et al., 2013; Faraji-Rad et al., 2012; Faraji-Rad et al., 2015; Schilbach et al., 2013). Our question, however, is whether similarity in one field will generalise to a biased assessment in another field – a kind of epistemic spillover. If we conclude that person X is good at finance simply because he tends to agree with us about the value of stocks, will we then be more likely to conclude he has superior abilities in predicting the weather? Because of the halo effect (Dion et al., 1972; Nisbett & Wilson, 1977; Thorndike, 1920), which is the tendency for an evaluation in one area to influence an evaluation in another area (see Chapter 1), we predicted this to be the case. The likely downstream behavioural consequence is that people will turn to others who think like them in one area for information in another area, even in cases where the evidence in front of them clearly indicates that this is suboptimal.

Overview of Experiments

Here, we ask whether (dis)similarity in political views interferes with the ability to learn about another person's competency in an unrelated task (specifically categorising shapes) in a situation in which it is in people's best interest to learn who excels in the task in order to turn to them for assistance. In the first part of our experiment, participants had an opportunity to learn whether others (i) had similar political opinions to theirs and (ii) how well they did in a task that required learning about shapes. After rating others on these two characteristics, they completed the second part of the experiment, where they decided to whom to turn to for advice when solving the shape task. They were rewarded for accuracy on the task and thus had an economic incentive to turn to the participant who was most skilled at the task.

We find that (dis)similarity in political views interferes with the ability to make an accurate assessment of people’s expertise in the domain of shapes, which leads to two central outcomes. The first is that people chose to hear about shapes from others who are politically like-minded, even when those people are not especially good at the shape task, rather than to hear from people who excel at the shape task but have different political opinions. The second is that people are more influenced by those with similar political opinions, even when they had the opportunity to learn that those by whom they are influenced are not especially good at the task they are solving. The results replicate in two independent samples. We suspect that these findings can be found in the real world, and that they help explain a range of phenomena, including the spread of fake news (Friggeri et al., 2014; Kahne & Bowyer, 2017; Traberg & van der Linden, 2022) conspiracy theories (Del Vicario et al., 2016), polarisation (Druckman et al., 2013; Prior, 2007), and insufficient learning in general (Yaniv & Kleinberger, 2000; Yaniv & Milyavsky, 2007).

Study 1

Method

Participants

American residents over 18 years of age who speak English were recruited on Amazon Mechanical Turk. All participants provided demographic information (Appendix 2). Sample size was determined using a power analysis (G*Power Version 3.1.9.2; Faul et al., 2007), based on the results of a pilot study. The pilot study was run by Eleanor Loh and Tali Sharot and assessed whether people’s (n = 79) choices of whom to hear from in a shape categorisation task are affected by the degree to which others agree with their answers on that task. This revealed that a sample size of 44 participants would achieve 80% power to detect an effect size of d = .43 (the difference between how often participants chose to hear from a source that was prone to agree with them and one that was prone to disagree with them), with an alpha of .05, assuming a correlation among repeated measures of -0.32. However, as this study was conducted online and was the first of this thesis, we chose to collect a larger sample than was suggested by the power analysis.

154 participants completed the first part of the task (Learning Stage). Participants had to pass the learning stage test (see below) in order to continue to the choice stage. 97 participants (34 females and 63 males, aged 20–58 years M = 34.81, SD = 9.59) passed the learning test. Participants who passed the learning test did not differ from those who failed on age, gender, ethnicity, language, education, income, subjective socio-economic-position, political ideology, interest/involvement in US politics, or generalised trust (all P > .12).

All participants were paid a base rate of $2.50. They were told they could earn a bonus between $2.50 and $7.50 based on their performance but were not told exactly how performance would be measured. Unbeknownst to the participants our rule for paying the bonus was as follows: any participants that passed the learning stage test (see details below) and completed the choice stage received $5 bonus.

Study Design

Learning Stage. The goal of the learning stage was to give participants an opportunity to learn about the other participants' (hereafter 'sources') political views and about their competency on the shape task (hereafter 'Blap task'). Before the learning stage, participants completed four practice blap trials and four practice political trials. They were not presented with information from sources on practice trials.

The learning stage consisted of 8 blocks of 20 trials each (10 blap trials and 10 political trials interleaved). Responses from one of the four sources were shown for the duration of a block (each source was used in two blocks), the order of which was randomised across blocks. Qualtrics' loop and merge tool was used to randomise the order of the questions within each block.

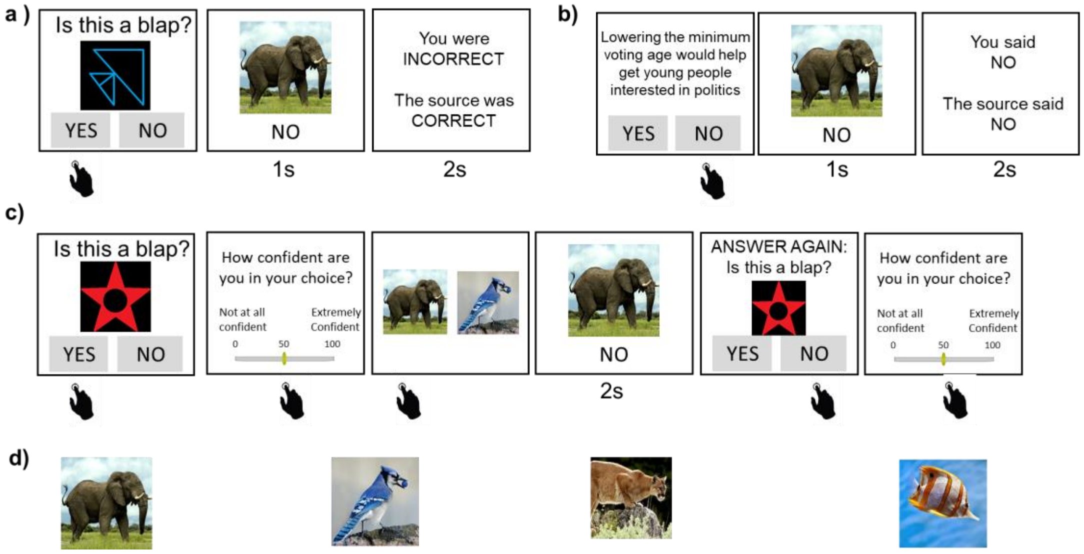
Blap Trials (Figure 4a). On each trial, one of 204 coloured shapes was presented on screen. Participants were required to learn through trial and error to classify shapes as 'blaps' or 'not blaps', ostensibly based on the shape's features. Unbeknownst to the participants, whether a shape was a blap or not was not rule based, but rather randomly determined before the beginning of the task, such that half the stimuli were categorised as 'blaps'. Because participants did not in fact have any means to learn which type of stimulus was a blap, the average performance across participants was around 50% (M = 48%, SD = 10.57). Participants had as much time as they needed to enter their response with a key press indicating either ‘yes’ (the shape is a blap) or ‘no’ (it is not) (M = 2.78s, SD = 9.27). They then observed for 1s the response of one of the four sources. Thereafter they received feedback on whether they and the source were each correct or incorrect (2s).

Political Trials (Figure 4b). On political trials, participants indicated whether they agreed or disagreed with one of 84 social/political cause-and-effect statements (e.g., “Lowering the minimum voting age would help get young people interested in politics”, see full set of statements in Appendix 3). These statements were developed on the basis of various political attitude questionnaires (see Appendix 3). Participants had as much time as they needed to press a key button indicating whether their response was ‘yes’ or ‘no’ (*M* = 5.89s, *SD* = 16.87). They then observed for 1s the response of one of the four sources. Thereafter they were shown their response together with that of the source (2s).

Sources. Participants were told that on each trial, they would be presented with the response of one of four participants (‘sources’) who performed the task earlier. Unbeknownst to the participants, these sources were not in fact other people but algorithms designed to respond in the following pattern. (i) One source agreed with the subject on 80% of the political trials and was correct on 80% of blap trials (Accurate/Similar). (ii) One source agreed with the subject on 80% of the political trials and was correct on only 50% of blap trials (Inaccurate/Similar). (iii) One source agreed with the subject on 20% of the political trials and was correct on 80% of blap trials (Accurate/Dissimilar). (iv) One source agreed with the subject on 20% of the political trials and was correct on 50% of blap trials (Inaccurate/Dissimilar). On blap trials all sources agreed with the participant about half the time on average (M = 50%, SD = 11.52). To avoid gender and racial bias, sources were represented with a picture of an animal (Figure 4d). Pictures assigned to sources were counterbalanced.

Figure 4

Experimental Design of the Task Used in Studies 1 and 2



Note. During the Learning Stage participants learned about the political opinions of four sources (represented by an animal photo) and about the sources' accuracy on a shape task (blap task). (a) Blap trials and (b) political trials were interleaved. (a) On each blap trial a novel shape was presented and the participants had to indicate whether they believed the shape was a blap (yes or no). They then saw the response of one of four sources represented by an animal photo. This was followed by feedback. (b) On political trials a political statement was presented and the participants had to indicate whether they agreed with it (yes or no). They then saw the response of one of four sources represented by an animal photo. This was followed by a reminder of their response and the source's response. (c) During the Choice Stage participants completed blap trials only. On each blap trial a novel shape was presented and the participant had to indicate whether they believed the shape was a blap (yes or no) and enter a confidence rating. They were then presented with two sources and asked to choose whose answer they would like to see. They then saw the response of the chosen source. Finally, they were given a chance to update their initial answer and confidence rating. Responses were self-paced unless otherwise stated. (d) There were four sources represented with animal photos which the participants were led to believe were other participants but were in fact algorithms.

Attention Check. At the end of each block, participants were presented with an attention check in which they were asked one of the following questions regarding the last trial: "Did the source AGREE or DISAGREE with your answer?"; "What was your last response?"; "Which source was shown on the last trial?"; "Was the last question a political or blap question?" For the latter two questions, 98.97% and 93.81% of participants were correct, respectively. Data was mistakenly not saved to report accuracy of the former two questions.

Learning Test. The goal of the study was to assess how similarity affected the ability to assess competence and information-seeking behaviour. We thus tested participants' perception of who was similar to them to determine if the similarity manipulation was successful. Specifically, after the learning stage, participants were presented with 12 trials. On each trial two sources were presented and the subject had to indicate who was more similar to them ("Who is more similar to you?"). Each possible pair of sources (six combinations) was presented twice for a total of twelve trials. Only participants who responded correctly (as determined according to the similarity manipulation described above) on eleven trials or more were considered to have accurately assessed similarity and continued to the choice stage (n = 97).

Ratings of Similarity and Accuracy. Participants then rated each source on (1) how competent they were at determining if each object was a blap ("How competent was the source at figuring out if each object was a blap?" from 0 = "Very incompetent" to 100 = "Very competent") and (2) how similar the source was to them ("How similar do you think this source was to you?" from 0 = "Not at all like me" to 100 = "Exactly like me"). We did not specifically ask about political similarity, as we wanted to avoid artificially focusing subjects' attention on that question. While participants may have construed the question as referring to political similarity and/or similarity on blap performance and/or similarity to the image of the animal, this would have only added noise to the data. As can be observed in the result section, sources who were objectively politically similar to the subjects were rated significantly higher on this scale, as expected.

Choice Stage (Figure 4c). The goal of the choice stage was to assess who the participant wanted to hear from about blaps and how they used the information they received. On each of 120 trials, participants were presented with a novel shape and asked to indicate with a button press whether they thought the shape was a blap ('yes' or 'no') (RT: M = 3.46s, SD = 53.90). They subsequently rated their confidence in this decision (self-paced) on a scale from 0 (not at all confident) to 100 (extremely confident). They were then presented with a pair of sources and asked whose response they wanted to see (self-paced) (RT: M = 2.04s, SD = 79.13). They were then shown the response of the chosen source for 2s. Thereafter the shape was presented again and participants were asked again to indicate with a button press whether they believed the shape was a blap ('yes' or 'no') (RT: M = 1.29s, SD = 9.79). Lastly, participants rated their confidence (self-paced) in their final decision.

The participants were instructed at the beginning of the choice stage that they could alter their answer on this second guess if they wanted to. There were 6 blocks of 20 trials each with the six source pairs pseudo-randomised throughout each block. There were no political trials nor feedback in the choice stage.

Second Attention Check. As in the learning stage, participants were presented with an attention check question at the end of each block in which they were asked one of the following questions: "Which source did you NOT select on the last trial?"; "Which source did you select on the last trial?"; "Did the source AGREE or DISAGREE with your answer?" There was an error in recording these data, thus we cannot provide accuracy rates.

Post-Task Ratings and Debrief. Finally, participants completed a debriefing questionnaire (see Appendix 2). During this debrief, participants were asked once again (1) how competent each source was at determining if each object was a blap ("How competent was the source at figuring out if each object was a blap?" from 0 = "Very incompetent" to 100 = "Very competent") and (2) how similar the source was to them ("How similar do you think this source was to you?" from 0 = "Not at all like me" to 100 = "Exactly like me"). The results remained unchanged if the post-task ratings were used instead of the pre-choice ratings.

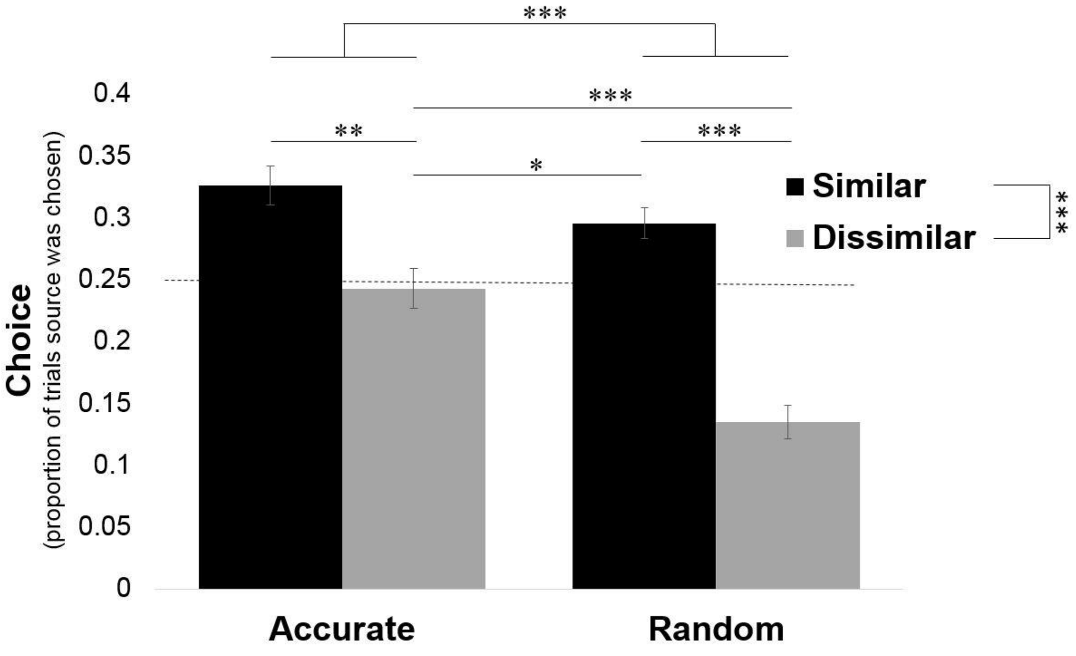
Results

Participants Prefer to Receive Information About Shapes from Politically Like-Minded Sources

We first asked whom participants select to hear from on the blap task. We find that participants sensibly prefer to hear from sources that are more accurate on the blap task, but also prefer to hear from politically like-minded sources even when they were not very good at the blap task (Figure 5).

Figure 5

Proportion of Trials on Which Participants Chose to Seek Information from Each Source



Note. The figure illustrates that participants prefer to receive information about shapes from politically like-minded sources. For each participant we calculated the percentage of times they selected to hear from each source about blaps out of all trials and averaged across participants. As each source was presented as an option an equal number of times, if the participants had no preference each source would be selected on about 25% of trials. A preference (main effect) for both accurate sources over inaccurate sources and for politically similar sources over politically dissimilar sources was found. Error bars represent SEM. *p < .05, **p < .01, ***p < .001.

Specifically, each source was presented as an option out of two sources on 50% of trials. Thus, if participants had no preference they would select each source on 25% of the trials. We found that the Accurate/Similar source was chosen most often (M = 33%, SD = 15.56; significantly greater than chance: t(96) = 4.85, p < .001), followed by the Inaccurate/Similar source (M = 30%, SD = 12.30; significantly greater than chance level: t(96) = 3.65, p < .001), followed by the Accurate/Dissimilar source (M = 24%, SD = 15.93; not different from chance level: t(96) = −0.44, p = .66), and finally by the Inaccurate/Dissimilar source (M = 13%, SD = 13.53; significantly lower than chance: t(96) = −8.34, p < .001). Entering percentage choice into a two (source similarity: similar, dissimilar) by two (source accuracy: accurate, inaccurate) repeated-measures (rm) ANOVA revealed a main effect of source accuracy (F(1,96) = 23.32, p < .001, η_p^2 = 0.20), a main effect of political similarity (F(1,96) = 33.67, p < .001, η_p^2 = 0.26) and an interaction (F(1,96) = 7.22, p = .008, η_p^2 = 0.07). The interaction was due to participants selecting to hear from the Accurate/Dissimilar source over the Inaccurate/Dissimilar source (t(96) = 5.05, p < .001, d = 0.73), but revealing no preference between the two similar sources (t(96) = 1.62, p = .11, d = 0.22). Strikingly, participants preferred to hear from the politically like-minded source that performed randomly on the blap task over the source that was accurate on the blap task but dissimilar politically (t(96) = −2.10, p = .038, d = −0.37).

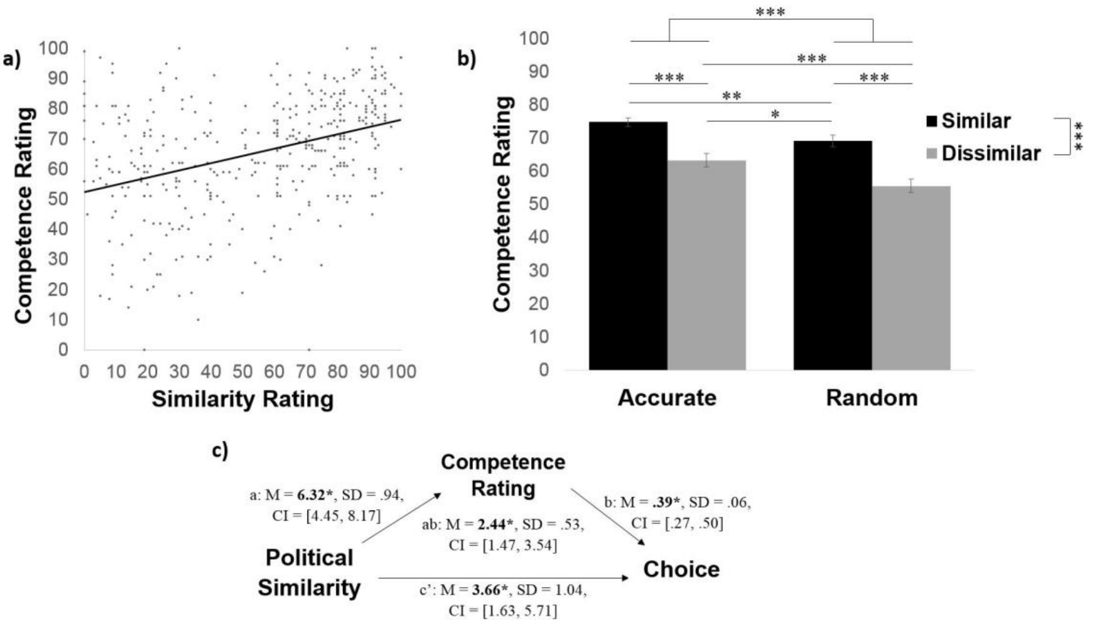
Political Similarity Leads to An Illusory Perception of Competence on The Shape Task

What could explain the tendency to seek information about shapes from others who are politically like-minded? Our hypothesis was that (dis)similarity in political views will interfere with participants’ ability to assess others’ competence on the blap task. The rationale is that political (dis)similarity will generate a (negative) positive view of the source, which will generalise to the unrelated domain of shape categorisation.

To test this hypothesis, we first tested for a correlation between participants' ratings of how similar the sources were to them and how good the sources were on the blap task. The true correlation was zero. Nonetheless, participants had an illusory perception that the more similar the source was to them, the better the source was on the shape task (r = 0.37, p < .001, Figure 6a).

Figure 6

An Illusory Perception of Accuracy Mediates the Relationship Between Political Similarity and Information Seeking Behaviour



Note. (a) The true correlation between how accurate a source was on the blap task and how like-minded they were to the participant was zero. Nevertheless, participants' ratings revealed an illusory perception that the two were related (r = 0.37 p < .001). (b) Participants rated accurate sources as more competent on the blap task, but also rated politically like-minded sources as more competent on the blap task. (c) A mediation model revealed that perceived competence partially mediated the relationship between political similarity and choice of which source to hear from about blaps. Error bars represent SEM. *p < .05, **p < .01, ***p < .001.

Second, we examined how participants rated the four sources on their ability to categorise shapes. Entering these ratings into a two (source similarity: similar, dissimilar) by two (source accuracy: accurate, inaccurate) rmANOVA revealed not only a sensible main effect of source accuracy ($F(1,96) = 22.98$, $p < .001$, $\eta_p^2 = 0.19$), but also an illusory main effect of source political similarity ($F(1,96) = 45.41$, $p < .001$, $\eta_p^2 = 0.32$) and no interaction ($F(1,96) = 0.74$, $p = .39$, $\eta_p^2 = 0.01$). Although both accurate sources were correct 80% of the time, participants rated the Accurate/Similar source as more competent at the blap task ($M = 75\%$, $SD = 12.91$) than the Accurate/Dissimilar source ($M = 63\%$, $SD = 18.83$; comparison between the two $t(96) = 5.52$, $p < .001$, $d = 0.72$). Likewise, although both inaccurate sources were accurate only 50% of the time participants rated the Inaccurate/Similar source as more competent ($M = 69\%$, $SD = 17.24$) than the Inaccurate/Dissimilar source ($M = 56\%$, $SD = 20.26$; comparison between the two $t(96) = 5.89$, $p < .001$, $d = 0.73$; Figure 6b). Interestingly, the source that had different political views but excelled at the blap task (Accurate/Dissimilar) was rated less competent on the blap task than the source that performed randomly but was politically like-minded ($t(96) = -2.58$, $p = .011$, $d = -0.33$).

An Illusory Perception of Competence on Shape Task Mediates the Relationship Between Political Similarity and Information-Seeking Behaviour

The above results suggest that political similarity influenced perceptions of source competence, with more politically similar sources viewed as more competent than their equally accurate counterparts. Does this explain the tendency to turn to politically like-minded people for information on blaps?

To test this possibility formally, we performed a causal mediation analysis (Figure 6c) that asks whether the relationship between objective political similarity and information seeking behaviour is mediated by subjective ratings of competence on the blap task.

A multilevel modelling approach was used (Preacher, 2015), which allows for the appropriate treatment of non-independent observations by nesting trial-level observations within upper-level units (individual participants). Bayesian estimation of the multilevel mediation model was performed in the R programming language, using the open-source software package bmlm (Vuorre & Bolger, 2017). The bmlm package estimates regression models, with individual-level and group-level parameters estimated simultaneously using Markov chain Monte Carlo (MCMC) procedures. The default MCMC sampling procedure was employed, with 4 MCMC chains and 2000 iterations.

The mediation model examined whether perceived competence mediates the relationship between objective political similarity and source chosen with a predictor (X; source political similarity), mediator (M; competence rating), and dependent variable (Y; percentage each source was chosen). Indeed, we found a significant indirect effect of political similarity on choice through subjective competence rating (path ab: $M_{posterior}$ = 2.44, SD = 0.53, CI = [1.47, 3.54]).

The model shows the following. First, objective political similarity predicted how likely the participant was to turn to a source for information about blaps (total effect: $M_{posterior}$ = 6.10, SD = 1.05, CI = [4.06, 8.20]). Politically like-minded sources were, in general, chosen more often. This effect was attenuated, though not eliminated, when controlling for subjective competence ratings (path c': $M_{posterior}$ = 3.66, SD = 1.04, CI = [1.63, 5.71]). Second, objective political similarity was positively related to subjective competence ratings (path a: $M_{posterior}$ = 6.32, SD = 0.94, CI = [4.45, 8.17]); similar sources were perceived as more competent. Third, subjective competence ratings predicted choice when objective political similarity was accounted for (path b: $M_{posterior}$ = 0.39, SD = 0.06, CI = [0.27, 0.50]), suggesting that subjective competence had a unique effect on choice.

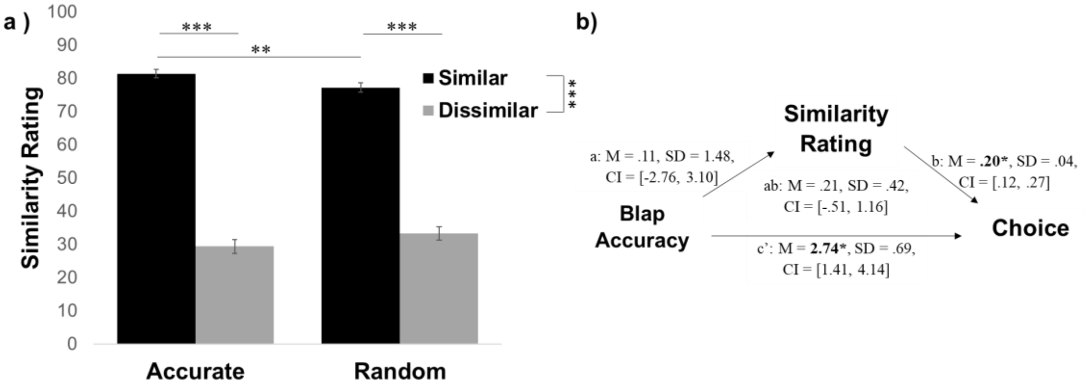
Accuracy on the Blap Task Affects Perception of Similarity

The above results suggest that the effect of political similarity on participants' choice of whom to turn to for information on blaps is partially mediated by their (illusory) subjective perception of the source's competence on the blap task. One may ask, though, whether the reverse relationship is also true. Although less intuitive, could it be that sources that are more accurate on blaps are perceived to be more similar and that this perceived similarity mediates a relationship between objective accuracy and information seeking behaviour?

To answer this question, we first examined how participants rated the sources on similarity. Entering similarity ratings into a 2 (source similarity: similar, dissimilar) × 2 (source accuracy: accurate, inaccurate) rmANOVA revealed a sensible main effect of political similarity (F(1,96) = 648.76, p < .001, η_p^2 = 0.87) and no significant main effect of accuracy (F(1,96) = 0.013, p = .91, η_p^2 < 0.01). An interaction also emerged (F(1,96) = 7.23, p = .008, η_p^2 = 0.07). The interaction was due to the fact that while both politically similar sources agreed with the participant 80% of the time on political trials, there was an illusory perception that the more accurate source on blaps (Accurate/Similar) was significantly more similar to the subject (M = 81%, SD = 11.81) than the source that performed randomly on the blap task (Inaccurate/Similar, M = 77%, SD = 14.15, difference between the two: t(96) = 2.48, p = .015, d = 0.33). The two politically dissimilar sources were not rated as significantly different on similarity (Accurate/Dissimilar M = 29%, SD = 20.44; Inaccurate/Dissimilar M = 33%, SD = 20.03; comparison between the two t(96) = −1.50, p = .140, d = −0.19; Figure 7a).

Figure 7

Accuracy On Blap Task Partially Enhances Sense of Similarity



Note. (a) Politically similar sources were rated as more similar by participants. Interestingly the politically like-minded source that was also more accurate on blaps was rated as more similar than the politically like-minded source that was random on blaps. This suggests that accuracy on blap task partially affected perceived similarity. (b) The reverse mediation to that tested in Figure 6 – by which perceived similarity mediates the effect between source accuracy and information seeking behaviour – was not significant. Error bars represent SEM. **p < .01, ***p < .001.

The above results reveal an illusion by which a source that is more accurate on the blap task is viewed as more similar to the self than a less accurate source that is equally similar on political questions, perhaps revealing a motivation to associate the self with successful, similar others. We therefore conducted a second mediation analysis, using the same procedure as above, to examine whether perceived similarity mediates the relationship between objective accuracy and source chosen, with a predictor (X; source accuracy), mediator (M; similarity rating), and dependent variable (Y; percentage each source was chosen).

Our mediation model showed that it was not the case that subjective similarity mediated a relationship between objective accuracy on the blap task and information seeking behaviour (Figure 7b). We did not find a significant effect of source accuracy on similarity rating, nor did we find evidence of an indirect effect.

In particular, the mediation showed that objective accuracy on the blap task predicted how likely the participant was to turn to a source for information about blaps (total effect: $M_{posterior} = 2.96$, $SD = 0.81$, $CI = [1.42, 4.57]$), showing that accurate sources were chosen more often. The effect was not, however, reduced when subjective similarity was controlled (path c’: $M_{posterior} = 2.74$, $SD = 0.69$, $CI = [1.41, 4.14]$), suggesting that the accuracy-related variance in source choice is not shared with subjective similarity. Although subjective similarity ratings predicted choice when objective blap accuracy was accounted for (path b: $M_{posterior} = 0.20$, $SD = 0.04$, $CI = [0.12, 0.27]$), suggesting that subjective similarity had a unique effect on choice, objective accuracy was not predictive of subjective similarity ratings (path a: $M_{posterior} = 0.11$, $SD = 1.48$, $CI = [-2.76, 3.10]$), and the indirect effect of accuracy on the blap task on choice through subjective similarity rating was not significant (path ab: $M_{posterior} = 0.21$, $SD = 0.42$, $CI = [-0.51, 1.16]$).

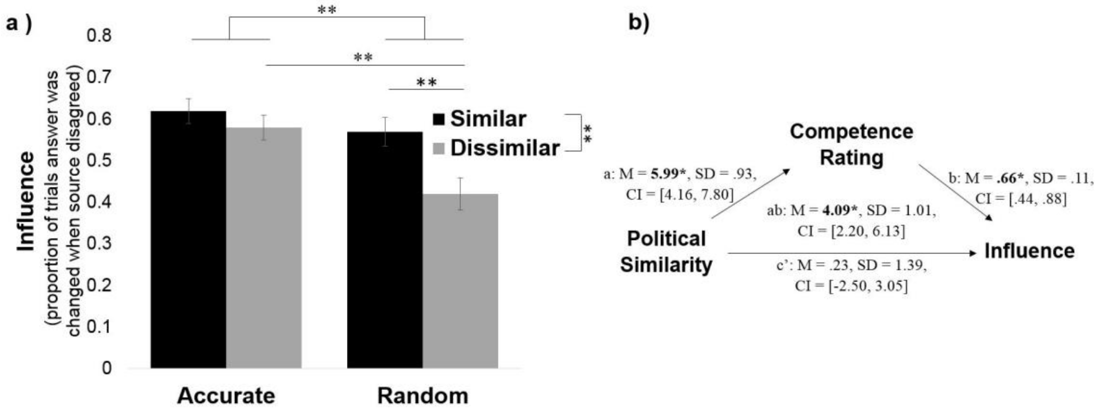
Participants' Shape Judgments Are More Influenced by Sources That Are Politically Like-Minded

Thus far we find that participants are inclined to turn to sources that are like-minded politically to receive information on blaps. Are they also more likely to be influenced by them? We quantified the extent to which participants were influenced by a source by calculating the percentage of times the participant changed their answer when a source disagreed with them (only participants who chose to hear from each source at least once could be included in this analysis, N included = 70).

We find that after choosing whom to listen to participants are more influenced by the sources that are politically like-minded and more accurate on the blap task (Figure 8a). Participants changed their decisions on disagreement trials most often in response to information from the Accurate/Similar source (M = 62%, SD = 30.02; significantly greater than chance: t(93) = 4.74, p < .001) followed by the Accurate/Dissimilar source (M = 58%, SD = 29.59; significantly greater than chance: t(93) = 3.31, p = .001), followed by the Inaccurate/Similar source (M = 57%, SD = 29.94; significantly greater than chance: t(95) = 3.10, p = .003) and finally by the Inaccurate/Dissimilar source (M = 42%, SD = 34.72; not different from chance level: t(76) = −1.70, p = .093).

Figure 8

Participants' Blap Judgments Are More Influenced by Sources That Are Politically Like-Minded



Note. (a) Participants were more likely to change their minds about blaps when sources that were (i) more accurate at the blap task and (ii) more politically like-minded disagreed with their blap judgment than when sources that were less accurate on blaps and/or politically different disagreed with their blap judgement. (b) A mediation model revealed that the relationship between political similarity and source influence was mediated by perceived competence on the blap task. Error bars represent SEM. **p < .01.

Entering percentage of answers changed on disagreement trials into a two (source similarity: similar, dissimilar) by two (source accuracy: accurate, inaccurate) rmANOVA revealed a main effect of source accuracy (F(1,69) = 8.90, p = .004, η_p^2 = 0.11), a main effect of political similarity (F(1,69) = 7.14, p = .009, η_p^2 = 0.09) and a marginal interaction effect (F(1,69) = 3.98, marginal p = .050, η_p^2 = 0.06). Post-hoc t-tests showed that the interaction was due to the Accurate/Dissimilar source having greater influence than the Inaccurate/Dissimilar source (t(73) = 3.24, p = .002, d = 0.53) while there was no difference in influence between the two similar sources (t(92) = 1.43, p = .16, d = 0.29).

In the Choice Stage participants rated how confident they were in both their initial answer and final answer. It was therefore possible to assess source influence based not only on the participant's decision to keep or change their answer in response to new information but also according to how much confidence in their initial judgement was affected. To incorporate confidence ratings into the analysis of source influence, we used the Change of Mind (COM) measure developed by Edelson et al. (2014). This measure computes the total amount of change in the participant's confidence in their answer after observing a source's answer, using the following equations:

$$COM_{Stay} = (\alpha * Final\ Confidence) - (\alpha * Initial\ Confidence) \tag{13}$$

For agree trials α = 1; for disagree trials α = –1;

$$COM_{Change} = (\alpha * Final\ Confidence) + (\alpha * Initial\ Confidence) \qquad (14)$$

For agree trials α = –1; for disagree trials α = 1;

Thus, if the participant does not change their answer after seeing the source's answer, then COM will simply reflect the difference between their initial and final confidence ratings. However, if the participant changes their answer in light of the information they received from the source, then COM is computed by summing the difference between their initial confidence rating and zero and the difference between zero and their final confidence rating. COM is positive when the participant's confidence moves in the direction of the source's answer and negative when their confidence moves in the opposite direction of the source's answer.

We calculated average COM scores for each participant. Where there was missing data due to the source never being chosen we imputed zeros, as COM was unaffected by these sources. We found that politically similar sources and sources that were accurate on the blap task had a greater effect on COM. Participants changed their minds most when receiving information from the Accurate/Similar source (M = 40.64, SD = 29.45), followed by the Accurate/Dissimilar source (M = 35.92, SD = 28.04), followed by the Inaccurate/Similar source (M = 34.91, SD = 25.06) and finally by the Inaccurate/Dissimilar source (M = 16.71, SD = 31.49).

Entering COM into a 2 (source similarity: similar, dissimilar) × 2 (source accuracy: accurate, inaccurate) rmANOVA revealed a main effect of source accuracy (F(1,96) = 22.74, p < .001, $\eta_{\mathbf{p}}^2$ = .19), a main effect of political similarity (F(1,96) = 15.13, p < .001, η_p^2 = .14) and an interaction effect (F(1,96) = 8.42, p = .005, η_p^2 = .08). The interaction was due to the Inaccurate/Similar source having greater influence than the Inaccurate/Dissimilar source (t(96) = 4.72, p < .001, d = .64) while there was no difference in influence between the two accurate sources (t(96) = 1.30, p = .20, d = .16).

The results suggest that both accuracy on the blap task and political similarity exert an effect on how influenced participants are by the sources. This finding held when confidence ratings were incorporated into our measure of source influence. We next conducted a mediation model to test whether the effect of political similarity on influence (we used the measure of source influence that did not include confidence ratings – i.e., whether the participant changed their answer when a source disagreed with them) was mediated by perceived accuracy on the blap task. Results of the multilevel mediation showed that objective political similarity predicted source influence (total effect: $M_{posterior} = 4.32$, SD = 1.44, CI = [4.32, 1.54]) and was also positively related to the subjective ratings of competence (path a: $M_{posterior} = 5.99$, SD = 0.93, CI = [4.16, 7.80]), which in turn predicted source influence when source similarity was accounted for (path b: $M_{posterior} = 0.66$, SD = 0.11, CI = [0.44, 0.88]). The indirect effect of political similarity on source influence through competence rating was significant (path ab: $M_{posterior} = 4.09$, SD = 1.01, CI = [2.20, 6.13]) and once subjective competence rating was controlled for political similarity no longer predicted source influence (path c': $M_{posterior} = 0.23$, SD = 1.39, CI = [−2.50, 3.05]). These results demonstrate that the effect of political similarity on influence is fully mediated by the perceived competence of the source (Figure 8b).

Note that the conceptually reverse mediation model, with objective source accuracy as the predictor, subjective political similarity as the mediator and source influence as the dependent variable was not significant (no significant effect of objective accuracy on subjective similarity nor an indirect effect on source influence).

In particular, the model shows that objective accuracy on the blap task predicted source influence (total effect: $M_{posterior} = 4.34$, SD = 1.37, CI = [1.65, 7.09]), showing that accurate sources had more influence. The effect was still significant when controlling for subjective similarity (path c': $M_{posterior} = 4.56$, SD = 1.35, CI = [1.91, 7.31]), suggesting that objective accuracy had a unique effect on source influence. Again, although subjective similarity ratings predicted source influence when objective blap accuracy was accounted for (path b: $M_{posterior} = 0.19$, SD = 0.05, CI = [0.08, 0.30]), suggesting that subjective similarity had a unique effect on source influence, objective accuracy was not predictive of subjective similarity ratings (path a: $M_{posterior} = -1.39$, $SD = 1.53$, $CI = [-4.49, 1.62]$), and the indirect effect of accuracy on the blap task on source influence through subjective similarity rating was not significant (path ab: $M_{posterior} = -0.22$, $SD = 0.38$, $CI = [-0.98, 0.54]$).

Discussion

The results of Study 1 suggested that knowledge of another's political views interferes with the ability to learn about that person's competence in an unrelated task. Politically like-minded sources were more likely to be chosen and the information they provided had a greater influence on participants' decisions. Our mediation analyses suggest that participants preferred to hear from, and were more influenced by, politically similar sources because they falsely believed these sources were better at categorising blaps than politically dissimilar sources.

Study 2

In Study 2 we test whether the findings of Study 1 replicate with minor adjustments to the methods (see below).

Method

Participants

The recruitment procedure was the same as for Study 1. In Study 2, 186 participants completed the Learning Stage. 101 (47 females and 54 males, aged 18–63 years $M = 37.59$, $SD = 10.92$) passed the learning test and proceeded to the Choice Stage.

Participants who passed the learning test did not differ from those who failed on age, gender, ethnicity, language, political ideology, interest/involvement in US politics, or generalised trust (all $P > .18$). Unlike in Study 1, participants that passed tended to have higher income ($t(184) = 2.06$, $p = .041$), education ($t(184) = 2.59$, $p = .010$) and subjective socio-economic-position ($t(184) = 4.83$, $p < .001$).

There was a strong positive correlation between performance on the attention check and accuracy on the learning test ($r = 0.44$, $p < .001$), suggesting that participants who passed the learning test (by answering at least eleven out of twelve trials correctly) were more attentive than those who failed. Participants who passed the learning test were correct on a greater number of the attention check questions (M = 92%, SD = 11.80) than those who failed the learning test (M = 78%, SD = 18.54; comparison between the two t(184) = 6.31, p < .001, d = 0.91). As in Study 1, participants who failed the learning test were not progressed to the choice stage and thus did not complete the main experimental task. In the choice stage, participants answered 74% of the attention checks correctly (SD = 19.71).

Study Design

The methods of Study 2 were the same as in Study 1 except for the following changes:

1) Contrary to Study 1, we did not determine in advance which stimuli were blaps. Rather, feedback was given regardless of stimulus shown such that all participants were told they were correct on exactly 50% of the blap trials and incorrect on exactly 50% of blap trials. In contrast, in Study 1 participants' accuracy rates depended on whether a stimulus was in fact coded to be a blap or not.
2) The percentage of times the sources gave the same answer to the participant's answer on blap trials was held constant at exactly 50% for each subject and source. In contrast, in Study 1 the percentage of times the sources gave the same answer as the participant on blap trials was not hard-coded and normally distributed around 50% (SD = 11.52).
3) The wording of one of the post-task questions was changed slightly to read "How politically similar do you think this source was to you?" This was done to ensure that participants knew the question referred to political similarity and not similarity on the blap task.
4) We added the following post-task question "How competent do you think you were at figuring out if each object was a blap?" 0 = "Very incompetent" to 100 = "Very competent".
5) Attention-check data was successfully recorded.

Changes 1–3 enables us to test for replication under slightly different conditions. Change 4 was to test for participants' perception of their own ability.

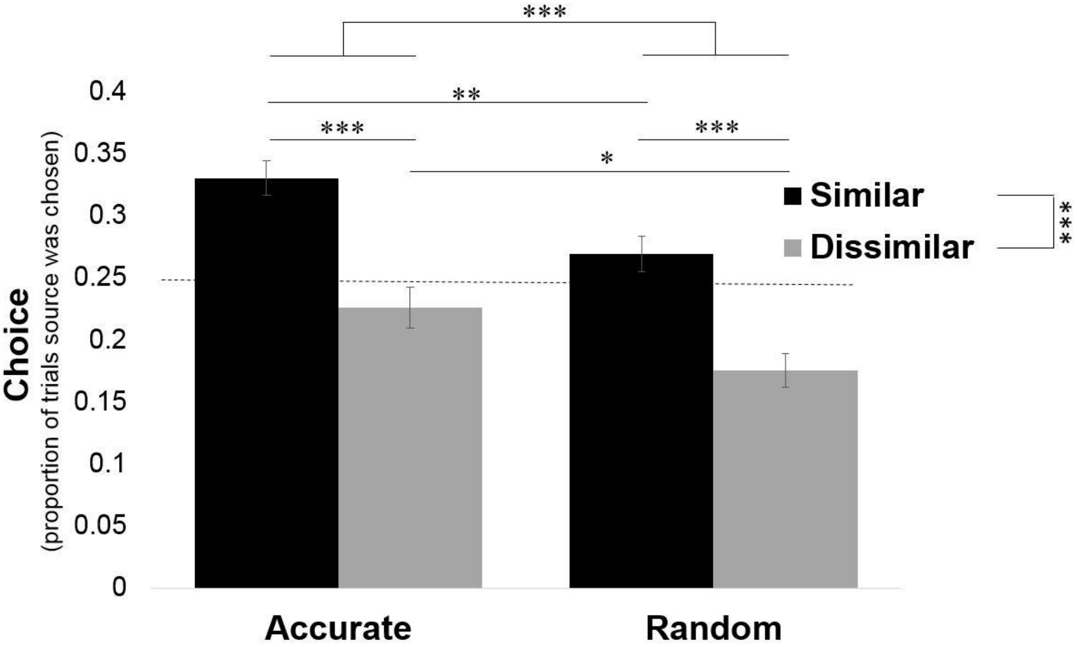
Results

Participants Prefer to Receive Information About Shapes from Politically Like-Minded Sources

As in Study 1, we find that participants sensibly prefer to hear from sources that are more accurate on the blap task, but also prefer to hear from politically like-minded sources even when they were not very good at the blap task (Figure 9). Specifically, the Accurate/Similar source was chosen most often (M = 33%, SD = 14.18; significantly greater than chance: t(100) = 5.70, $p < .001$), followed by the Inaccurate/Similar source (M = 27%, SD = 14.16; not different from chance: t(100) = 1.32, p = .19), followed by the Accurate/Dissimilar source (M = 23%, SD = 16.72; not different from chance: t(100) = −1.45, p = .15) and finally by the Inaccurate/Dissimilar source (M = 18%, SD = 13.66; significantly lower than chance: t(100) = −5.51, $p < .001$). Entering the percentage of times each participant selected each source into a two (source similarity: similar, dissimilar) by two (source accuracy: accurate, inaccurate) rmANOVA revealed a main effect of source accuracy (F(1,100) = 14.09, $p < .001$, $\eta_p^2 = 0.12$) and political similarity (F(1,100) = 25.07, $p < .001$, $\eta_p^2 = 0.20$) with no interaction (F(1,100) = 0.13, p = .720, $\eta_p^2 = 0.001$). Participants were not more likely to choose the source that was accurate on the blap task but dissimilar politically (Accurate/Dissimilar) over the politically like-minded source that performed randomly on the blap task (Inaccurate/Similar) (t(100) = −1.61, p = .11, d = −0.28).

Figure 9

Proportion of Trials on Which Participants Chose to Seek Information from Each Source



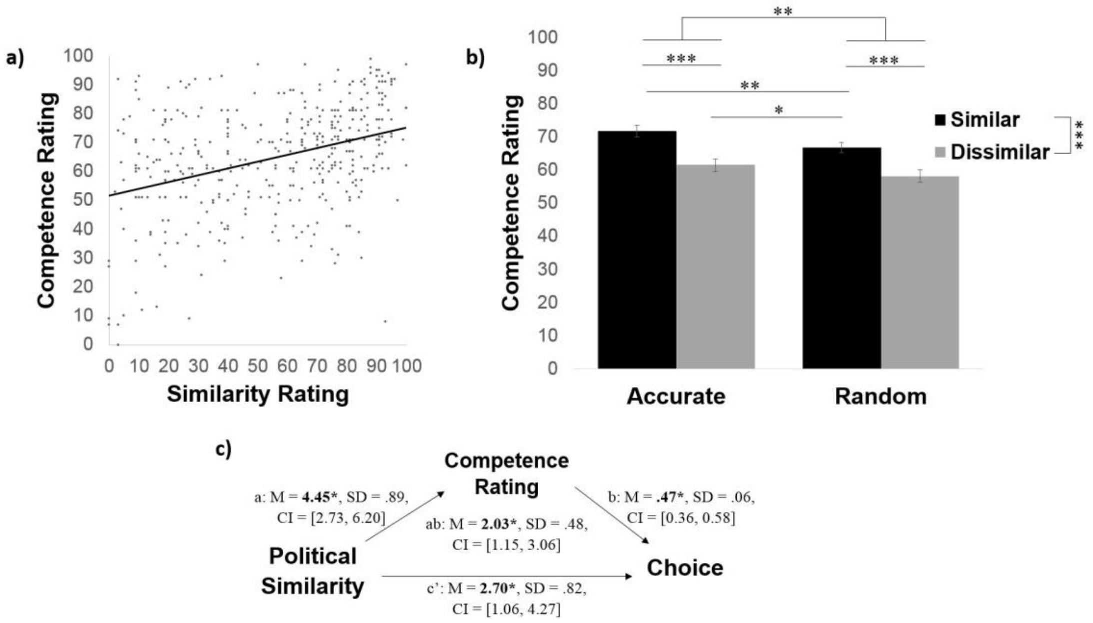
Note. For each participant we calculated the percentage of times they selected to hear from each source about blaps out of all trials and averaged across participants. As each source was presented as an option an equal number of times if the participants had no preference each source would be selected about 25% of trials. A preference (main effect) for both accurate sources over inaccurate sources and for politically similar sources over politically dissimilar sources was found. Error bars represent SEM. *p < .05, **p < .01, ***p < .001.

Political Similarity Leads to An Illusory Perception of Competence on The Shape Task

As in Study 1, we find that participants' ratings of how similar the sources were to them correlated with their ratings of how competent they thought the sources were at the blap task. Specifically, participants had an illusory perception that the more similar the source was to them, the better the source was on the shape task (r = 0.36, p < .001, Figure 10a).

Figure 10

An Illusory Perception of Accuracy Mediates the Relationship Between Political Similarity and Information Seeking Behaviour



Note. (a) The true correlation between how accurate a source was on the blap task and how like-minded they were to the participant was zero. Nevertheless, participants' ratings revealed an illusory perception that the two were related (r = 0.36 p < .001). (b) Participants rated accurate sources as more competent on the blap task, but also rated politically like-minded sources as more competent on the blap task. (c) A mediation model revealed that perceived competence partially mediated the relationship between political similarity and choice of which source to hear from about blaps. Error bars represent SEM. *p < .05, **p < .01, ***p < .001.

We then assessed how participants rated the four sources on their ability to categorise blaps, entering these ratings into a two (source similarity: similar, dissimilar) by two (source accuracy: accurate, inaccurate) rmANOVA. This revealed a sensible main effect of source accuracy (F(1,100) = 7.67, p = .007, η_p^2 = 0.07), an illusory main effect of source political similarity (F(1,100) = 27.88, p < .001, η_p^2 = 0.22) and no interaction (F(1,100) = 39, p = .530, η_p^2 < 0.01).

Participants rated the Accurate/Similar source as more competent at the blap task (M = 72%, SD = 17.12) than the Accurate/Dissimilar source (M = 62%, SD = 20.13; comparison between the two t(100) = 4.26, p < .001, d = 0.55). Likewise, participants rated the Inaccurate/Similar source as more competent (M = 67%, SD = 15.19) than the Inaccurate/Dissimilar source (M = 58%, SD = 18.43; comparison between the two t(100) = 4.18, p < .001, d = 0.51; Figure 10b). The source that was politically like-minded but poor on the blap task (Inaccurate/Similar) was rated as more competent at the blap task than the source that performed well but had different political views (t(100) = −2.18, p = .031, d = −0.29).

An Illusory Perception of Competence on Shape Task Mediates the Relationship Between Political Similarity and Information-Seeking Behaviour

We next test whether participants chose to hear from the politically similar sources because they believed they were more competent at the blap task. That is, we ask whether the relationship between objective political similarity and information seeking behaviour is mediated by subjective ratings of competence on the blap task. We used the same procedure as in Study 1 to perform this mediation analysis.

The model shows that objective political similarity predicted how likely the participant was to turn to a source for information about blaps (total effect: $M_{posterior}$ = 4.73, SD = 0.90, CI = [2.92, 6.46]), with politically like-minded sources chosen more often. This effect was attenuated, though not eliminated, when controlling for subjective competence ratings (path c’: $M_{posterior}$ = 2.70, SD = 0.82, CI = [1.06, 4.27]). Objective political similarity was positively related to subjective competence ratings (path a: $M_{posterior}$ = 4.45, SD = 0.89, CI = [2.73, 6.20]); similar sources were perceived as more competent. Subjective competence ratings predicted choice when objective political similarity was accounted for (path b: $M_{posterior}$ = 0.47, SD = 0.06, CI = [0.36, 0.58]), suggesting that subjective competence had a unique effect on choice. Finally, we find a significant indirect effect of political similarity on choice through subjective competence rating (path ab: $M_{posterior}$ = 2.03, SD = 0.48, CI = [1.15, 3.06]).

Accuracy on the Blap Task Affects Perception of Similarity

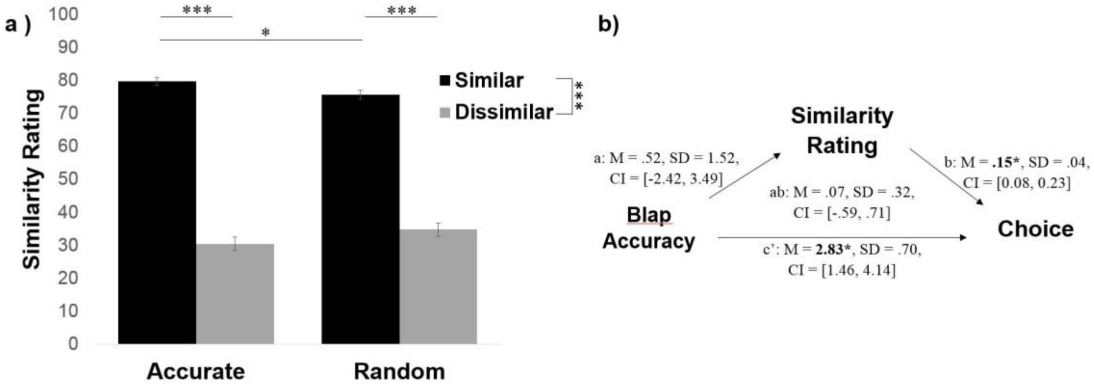
We next test whether sources that are more accurate on blaps are perceived as more similar and whether this increase in perceived similarity mediates the relationship between objective accuracy and information seeking behaviour.

We examined how participants rated the four sources on similarity, entering similarity ratings into a 2 (source similarity: similar, dissimilar) × 2 (source accuracy: accurate, inaccurate) rmANOVA. The results revealed a main effect of political similarity ($F(1,100) = 596.46$, $p < .001$, $\eta_p^2 = 0.86$), no main effect of accuracy ($F(1,100) = 0.01$, $p = .94$, $\eta_p^2 < 0.01$) and an interaction effect ($F(1,100) = 6.94$, $p = .010$, $\eta_p^2 = 0.07$).

As in Study 1, for politically similar sources participants believed that the more accurate source on blaps (Accurate/Similar) was significantly more similar (M = 80%, SD = 12.24) than the source that performed randomly on the blap task (Inaccurate/Similar, M = 76%, SD = 14.57, difference between the two: $t(100) = 2.30$, $p = .024$, $d = 0.30$). The politically dissimilar sources were not rated as significantly different on similarity (Accurate/Dissimilar M = 30%, SD = 20.09; Inaccurate/Dissimilar M = 35%, SD = 20.69; comparison between the two $t(100) = -1.60$, $p = .11$, $d = -0.21$; Fig. 8). Thus our finding from Study 1 that sources that are both politically similar and accurate on the blap task are viewed as more similar to the self than sources that are politically similar but less accurate on the blap task was replicated.

Figure 11

Accuracy on Blap Task Partially Enhances Sense of Similarity



Note. (a) Politically similar sources were rated as more similar by participants. Interestingly the politically like-minded source that was also more accurate on blaps was rated as more similar than the politically like-minded source that was random on blaps. This suggests that accuracy on blap task partially affected perceived similarity. (b) The reverse mediation to that tested in Figure 10c – by which perceived political similarity mediates the effect between source accuracy and information seeking behaviour – was not significant. Error bars represent SEM. *p < .05, ***p < .001.

We conducted another mediation analysis to examine whether perceived political similarity mediates the relationship between objective accuracy and source chosen, with a predictor (X; source accuracy), mediator (M; similarity rating), and dependent variable (Y; percentage each source was chosen).

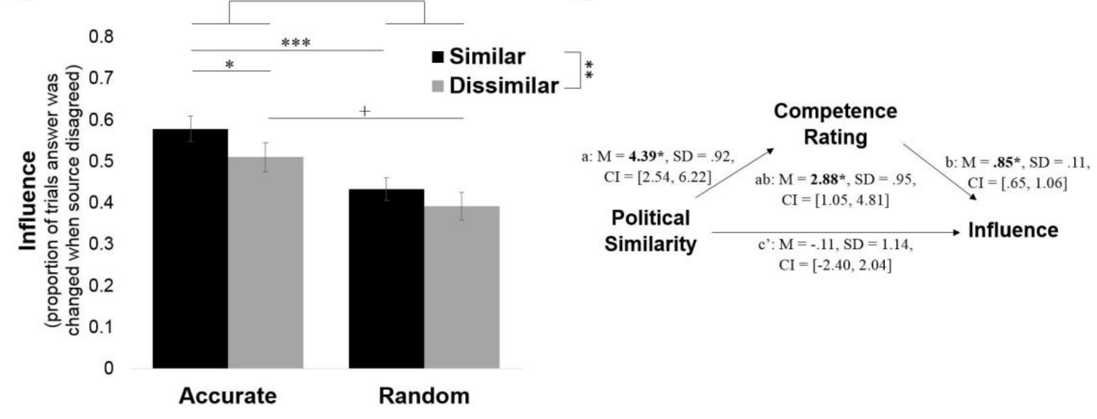
Again, we did not find a significant effect of source accuracy on similarity rating nor did we find evidence of an indirect effect. The mediation showed that objective accuracy on the blap task predicted how likely the participant was to turn to a source for information about blaps (total effect: $M_{posterior} = 2.90$, SD = 0.77, CI = [1.33, 4.40]), showing that accurate sources were chosen more often. The effect was not, however, reduced when subjective similarity was controlled (path c': $M_{posterior} = 2.83$, SD = 0.70, CI = [1.46, 4.14]), suggesting that the accuracy-related variance in source choice is not shared with subjective similarity. Although subjective similarity ratings predicted choice when objective blap accuracy was accounted for (path b: $M_{posterior}$ = 0.15, SD = 0.04, CI = [0.08, 0.23]), suggesting that subjective similarity had a unique effect on choice, objective accuracy was not predictive of subjective similarity ratings (path a: $M_{posterior}$ = 0.52, SD = 1.52, CI = [−2.42, 3.49]), and the indirect effect of accuracy on the blap task on choice through subjective similarity rating was not significant (path ab: $M_{posterior}$ = 0.07, SD = 0.32, CI = [−0.59, 0.71]).

Participants' Shape Judgments Are More Influenced by Sources That Are Politically Like-Minded

As in Study 1, participants were more influenced by the politically similar sources as well as those that were more accurate on the blap task (N included = 75; Figure 12a). Participants changed their answer most after hearing that the Accurate/Similar source disagreed with them (M = 58%, SD = 31.07; significantly different from chance: t(97) = 2.59, p = .011) followed by the Accurate/Dissimilar source (M = 52%, SD = 34.39; not different from chance: t(91) = 0.51, p = .61), followed by the Inaccurate/Similar source (M = 45%, SD = 27.59; not different from chance: t(94) = −1.73, p = .088) and finally by the Inaccurate/Dissimilar source (M = 41%, SD = 34.24; significantly lower than chance: t(92) = −2.43, p = .017).

Figure 12

Participants' Blap Judgments Are More Influenced by Sources That Are Politically Like-Minded



Note. (a) Participants were more likely to change their minds about blaps when sources that were (i) more accurate at the blap task and (ii) more politically like-minded disagreed with them than when sources that were random at the blap task or dissimilar disagreed with them. (b) A mediation model revealed that the relationship between political similarity and source influence was mediated by perceived accuracy on blap task. Error bars represent SEM. +p < .10, *p < .05, **p < .01, ***p < .001.

Entering percentage of answers changed out of trials in which the source disagreed with the participants' blap judgment into a two (source similarity: similar, dissimilar) by two (source accuracy: accurate, inaccurate) rmANOVA revealed a main effect of source accuracy (F(1,74) = 11.27, p = .001, η_p^2 = 0.13), a main effect of political similarity (F(1,74) = 7.36, p = .008, η_p^2 = 0.09) and no interaction (F(1,74) = 0.72, p = .40, η_p^2 = 0.01).

To confirm the robustness of these results, we ran a second analysis in which participants' confidence ratings were incorporated into the measure of source influence. Specifically, we calculated a COM score (see Study 1 for more details) for each source per participant. We found that the Accurate/Similar source had the greatest effect on COM (M = 38.29, SD = 26.70), followed by the Accurate/Dissimilar source (M = 30.27, SD = 27.42), followed by the Inaccurate/Similar source (M = 26.00, SD = 27.31) and finally by the Inaccurate/Dissimilar source (M = 25.25, SD = 27.23). The results of a 2 (source similarity: similar, dissimilar) × 2 (source accuracy: accurate, inaccurate) rmANOVA revealed a main effect of source accuracy (F(1,100) = 13.81, p < .001, η_p^2 = .12), a marginal effect of political similarity (F(1,100) = 3.25, p = .074, η_p^2 = .03), and no interaction effect (F(1,100) = 2.48, p = .11, η_p^2 = .02).

We next conducted a mediation model to test whether the effect of political similarity on influence (as in Study 1, we used the measure of source influence that did not include confidence ratings here – i.e., whether the participant changed their answer when a source disagreed with them) was mediated by perceived accuracy on the blap task (Figure 12b). Results of the multilevel mediation showed that objective political similarity predicted source influence (total effect: $M_{posterior} = 2.77$, SD = 1.32, CI = [0.15, 5.35]) and was also positively related to the subjective ratings of competence (path a: $M_{posterior} = 4.39$, SD = 0.92, CI = [2.54, 6.22]), which in turn predicted source influence when source similarity was accounted for (path b: $M_{posterior} = 0.85$, SD = 0.11, CI = [0.65, 1.06]). The indirect effect of political similarity on source influence through competence rating was significant (path ab: $M_{posterior} = 2.88$, SD = 0.95, CI = [1.05, 4.81]) and once subjective competence rating was controlled for political similarity no longer predicted source influence (path c': $M_{posterior} = -0.11$, SD = 1.14, CI = [−2.40, 2.04]). These results demonstrate that the effect of political similarity on influence is fully mediated by the perceived competence of the source.

Note that the conceptually reverse mediation model, with objective source accuracy as the predictor, subjective political similarity as the mediator and source influence as the dependent variable was not significant (no significant effect of objective accuracy on subjective political similarity nor an indirect effect on source influence).

In particular, the model shows that objective accuracy on the blap task predicted source influence (total effect: $M_{posterior} = 5.32$, SD = 1.22, CI = [2.90, 7.71]), showing that accurate sources had more influence. The effect was still significant when controlling for subjective similarity (path c': $M_{posterior} = 5.25$, SD = 1.20, CI = [2.86, 7.63]), suggesting that objective accuracy had a unique effect on source influence. Subjective similarity ratings did not predict source influence when objective blap accuracy was accounted for (path b: $M_{posterior} = 0.09$, SD = 0.05, CI = [−0.01, 0.20]), objective accuracy was not predictive of subjective similarity ratings (path a: $M_{posterior} = 0.53$, SD = 1.49, CI = [−2.42, 3.55]), and the indirect effect of accuracy on the blap task on source influence through subjective similarity rating was not significant (path ab: $M_{posterior} = 0.07$, SD = 0.28, CI = [−0.49, 0.65]).

Discussion

The central results of Study 1 were replicated in Study 2, demonstrating the robustness of the findings. We can therefore conclude that the differences between the two studies, such as whether participants' accuracy on the blap task was normally distributed around 50% or hard-coded to be exactly 50%, were trivial and cannot explain the observed effects of political similarity on perceptions of competence, information-seeking decisions, and advice utilisation.

General Discussion

The current studies offer three central findings. The first is that people choose to hear from those who are politically like-minded on topics that have nothing to do with politics (like geometric shapes) in preference to those who have different political views. The second is that all else being equal, people are more influenced by politically like-minded others on non-political issues such as shape categorisation than they are by those who disagree with them on political issues. The third is that people are biased to believe that others who share their political opinions are better at tasks that have nothing to do with politics, even when they have all the information they need to make an accurate assessment about who is the expert in the room. Our mediation analysis suggests that it is this illusion that underlies participants' tendency to seek and use information from politically like-minded others.

A great deal of attention has recently been paid to what sources of political information people choose (Prior, 2007; Sunstein, 2017), how algorithms affect what they see (Garimella et al., 2018; Hannak et al., 2013; Sîrbu et al., 2018; Sunstein, 2017), and how people are affected by encountering diverse information on political issues (Colleoni et al., 2014; Druckman et al., 2013; Kahan, 2016; Tappin et al., 2017). There is also growing interest in how political affiliations affect people's affective responses to those with different affiliations (Iyengar et al., 2012; Iyengar & Westwood, 2015).

Our focus here has been on epistemic spillovers – on whether and how a sense of shared political convictions influences people's desire to consult and to use people's views on a task that is entirely unrelated to politics. The most striking finding is that people consult and are influenced by the judgments of those with shared political convictions even when they had observed evidence suggesting that those with different convictions are far more likely to offer the right answer.

While we manipulated similarity on political views, we hypothesise that similar findings may be observed when similarity is manipulated along other dimensions that are significant to people (e.g., personal values, hobbies etc.), a hypothesis that warrants empirical testing. Moreover, it would be of interest to test whether people are also more influenced by the like-minded when they receive information from sources passively (i.e., without first choosing to seek information from a given source) rather than after making an active choice of whom to hear from (as in our study).

What accounts for our findings? We have referred to the halo effect: If people think that products or people are good along some dimension, they tend to think that they are good along other dimensions as well (Dion et al., 1972; Nisbett & Wilson, 1977; Thorndike, 1920). If people have an automatic preference for those who share their political convictions, their positive feelings may spill over into evaluation of other, unrelated characteristics (including their ability to identify blaps). This would be a consequence of political tribalism. A related explanation is that people use a heuristic, or mental shortcut, which often works well but which can lead to severe and systematic errors (Kahneman & Frederick, 2002). That is, if people generally believe that politically like-minded people are particularly worth consulting, they might extend that belief to contexts in which the belief does not make much sense. The current studies do not provide any insight into the causal model underlying halo effects (see Chapter 1; for more details, see Fisicaro & Lance, 1990; Lance et al., 1994), nor can they distinguish between these affective and cognitive explanations.

Our findings have implications for the spread of false news, for political polarisation, and for social divisions more generally. A great deal of false news is political (Kuklinski et al., 2000; Kull et al., 2003) and it is spread by and among like-minded people (Del Vicario et al., 2016). But our findings suggest that among the politically like-minded, false news will spread even if it has little or nothing to do with politics, or even if the connection to politics is indirect and elusive. Suppose, for example, that someone with congenial political convictions spreads a rumour about a coming collapse in the stock market, a new product that supposedly cures cancer or baldness, cheating in sports, an incipient epidemic, or a celebrity who has shown some terrible moral failure. Even if the rumour is false, and even if those who hear it have reason to believe that it is false, they may well find it credible (and perhaps spread it).

The results help identify both a cause and a consequence of political polarisation. If people trust like-minded others not only on political questions (Nyhan & Reifler, 2010) but also on questions that have nothing at all to do with politics, the conditions are ripe for sharp social divisions, potentially leading people to live in different epistemic universes.

Chapter 3

Chapter Overview

In Chapter 2, we demonstrated that information about others' political opinions interferes with the ability to learn about and utilise others' subject-matter expertise in a shape categorisation task. In this chapter, we sought to formalise this learning bias using computational modelling. We hypothesised that participants in the previous studies learned more from evidence indicating that politically similar sources were accurate and politically dissimilar sources were inaccurate on the blap task than evidence indicating the reverse. In an adapted version of the experimental paradigm used previously, in the two studies presented here we had participants bet on how the sources would answer questions during the learning stage rather than having participants answer the questions themselves. This allowed us to dynamically track participants' estimates of each source's competence and probe the trial-by-trial dynamics of learning using computational models. A second aim of the current studies was to assess whether knowledge of others' generosity influences the ability to learn about expertise in unrelated domains. People prefer to cooperate with and are more influenced by those whom they perceive as warm and likeable. We hypothesised that people misperceive those who share their political views as more competent at non-political tasks than those with different political opinions because their positive evaluations of them generalise to unrelated domains, therefore the same effects should be observed when people form positive evaluations on other dimensions. In Study 3, participants learned about others' generosity, under the pretence that those others were given multiple opportunities to donate money to charities, as well as their ability to correctly answer general knowledge questions. In Study 4 participants learned about others' political opinions and their ability to correctly answer general knowledge questions. Surprisingly, in our adapted version of the experimental paradigm, we did not find effects of generosity or political similarity on participants' expertise learning, information-seeking decisions, or advice utilisation. Participants correctly concluded which sources were accurate and which were not and thus chose to hear from accurate sources regardless of how generous or politically like-minded they were. These null results suggest that the effects observed in Chapter 2 are less robust than first thought. It remains an open question as to which of the adaptations made to the studies in this chapter altered participants' judgements and behaviour.

Introduction

The studies reported in the previous chapter demonstrate that politically like-minded messengers are perceived as more competent in non-political domains. While the ratings and choice data indicated that participants emerged from the learning stage with biased beliefs about the sources' competence, we did not directly measure how participants updated their beliefs in response to feedback and were therefore unable to provide an account of *how* learning about others' political beliefs biases expertise learning.

Computational modelling allows researchers to test competing 'algorithmic hypotheses' about how behaviour is generated and therefore offers a useful approach for delineating the cognitive processes underlying how decision-making biases emerge (e.g., Lefebvre et al., 2017; for a primer on computational modelling, see Wilson & Collins, 2019). In the field of social cognition, it is becoming increasingly common to use computational models to compare impression updating in humans to that of an optimal Bayesian agent (Behrens et al., 2008; Boorman et al., 2013; Diaconescu et al., 2014, 2017; Leong & Zaki, 2018) and to contrast how much people learn from feedback about different types of people (Chang et al., 2010; Delgado et al., 2005; Freeman et al., 2010; Hackel et al., 2015; Hackel et al., 2020; Yu et al., 2014; for reviews, see Hackel & Amodio, 2018; Mende-Siedlecki, 2018). Here, we sought to use modelling techniques to provide a computational account of the effects described in Chapter 2.

As participants in the previous studies did not make trial-by-trial judgements or predictions before or after receiving feedback, it was not possible to employ computational models that could provide a mechanistic account of how political similarity biases expertise learning. In the two studies described in this chapter, we alter the experimental paradigm to allow us to model the data and thus probe the trial-by-trial dynamics of learning. Specifically, we had participants bet on how the sources would answer on each trial of the learning stage before seeing feedback. This provided us with a behavioural trace from which to infer how people update their beliefs about different messengers in response to observed evidence.

We hypothesised that people learn more from congruent feedback (e.g., feedback indicating that a politically like-minded source is competent at categorising shapes) than incongruent feedback (e.g., feedback indicating that a politically like-minded source is incompetent at categorising shapes). That is, evidence that suggests that messengers who possess desirable characteristics in unrelated domains are competent at the task at hand will be weighted relatively more than evidence to the contrary. Likewise, evidence that suggests that messengers who possess undesirable characteristics in unrelated domains are incompetent in the relevant domain will be overweighted compared to evidence suggesting they are competent. A congruence bias in belief updating of this nature could explain why people perceived politically similar messengers as more competent at categorising shapes in the studies reported in Chapter 2. It would also align well with previous findings indicating that people discount undesirable information about in-group relative to out-group members (Hughes et al., 2017) and underweight information about others that is inconsistent with their expectations (Leong & Zaki, 2018).

A second aim of the studies reported in this chapter was to assess whether the effects reported in Chapter 2 generalise to other characteristics and contexts. As mentioned in Chapter 1, there are various routes by which beliefs about specific characteristics might influence beliefs about other characteristics. In particular, people might infer the characteristics of other people either from salient unrelated traits or behaviours, or from a general impression of the person (Fisicaro & Lance, 1990; Gräf & Unkelbach, 2016). Based on the 'Salient Dimensions' model of the halo effect (Fisicaro & Lance, 1990), which suggests that halo effects depend on the salience of observed traits and behaviours, it stands to reason that an observed characteristic should have a stronger influence on perceptions of another characteristic if the former tends to be predictive of the latter (Gräf & Unkelbach, 2018; Orehek et al., 2010). It is possible that people hold implicit assumptions about the relationship between intellectual competence and political ideology. For example, the motivation to maintain a positive social identity may lead people to think that those who think like them tend to be more intelligent (Tajfel & Turner, 1979). If these two characteristics are linked in a person's mind, one might expect a direct halo effect from inferences about political similarity to inferences about general competence. In contrast, perceptions of warmth-related traits (e.g., generosity) and competence-related traits (e.g., intelligence) are theoretically orthogonal (Abele & Wojciszke, 2014; Bakan, 1966; Cuddy et al., 2008; Koch et al., 2021; Oosterhof & Todorov, 2008) and, according to some works, compensatory (Aaker et al., 2010; Judd et al., 2005; Yzerbyt, 2018; Yzerbyt et al., 2005; Yzerbyt et al., 2008). If the effects we observed in our previous studies were driven by a more general tendency to expect messengers who possess desirable characteristics to perform well on unrelated tasks, we would expect learning about generosity to also interfere with the ability to assess and utilise others' expertise (via an indirect halo effect on general impressions). On the other hand, if knowing others' political opinions influences expertise learning due to an implicit association between political like-mindedness and competence, then learning about a desirable trait that is not seen to be related to competence should not produce the same effects.

The evidence to date suggests that both possibilities are plausible. Experimental work demonstrates that people who share money with others are perceived as warmer but no more competent than those who keep money that is given to them for themselves (Klein & Epley, 2014). This suggests that perceptions of warmth do not influence beliefs about others' general competence. However, other research shows that learning about others' generosity in a monetary task has an impact on decisions of whom to cooperate with in a subsequent non-monetary intellect-based task (Hackel et al., 2015). Moreover, research in healthcare settings has demonstrated that patients experience more positive outcomes when healthcare providers, such as therapists and doctors, act warmly towards them than when they act coldly (Ambady et al., 2002; Howe et al., 2017; Rogers, 1957). In a particularly striking study, Howe et al. (2017) examined how social inferences about a health-care provider influenced expectations about the efficacy of a treatment they were prescribing. Participants in this study met with a doctor who administered a histamine skin prick test to induce a mild allergic reaction. The doctor was trained to appear either competent or incompetent, and either warm or cold, while conducting the procedure. After inducing an allergic reaction, the doctor rubbed a cream onto the reaction site, which they told the participant would reduce its size. In reality, the cream was an unscented hand lotion with no medicinal active ingredient (i.e., a placebo). Remarkably, participants who interacted with a warm doctor experienced a larger placebo effect than those who interacted with a cold doctor, as evidenced by a greater reduction in allergy symptoms in response to the (non-medicinal) treatment. This suggests that perceptions of others' warmth affect judgements of their competence in unrelated contexts, although it is unclear whether warmth assessments influenced expertise beliefs per se from these results. In Study 3, we test these competing hypotheses directly by manipulating observed generosity in a charity donation task.

In addition to assessing whether the effects generalise to a warmth-related characteristic, we also tested whether we could replicate the findings in a different setting. The previous studies were run online using a US-based sample of participants. The US has experienced a particularly large increase in both affective polarisation (Boxell et al., 2020; Iyengar et al., 2012; Iyengar et al., 2019) and ideological polarisation (Draca & Schwarz, 2020) over the last few decades. It was thus important to replicate our findings in a non-US context. The studies presented here were run in the UK, with a student sample. Replicating the effects in this context would provide evidence to suggest that the tendency to believe those with desirable characteristics are better at unrelated tasks is a feature of human cognition rather than a US-specific cultural quirk.

A third aim was to replicate the effects with a more naturalistic task than the one used in the previous studies. The benefit of having participants learn about "blaps" is that it is impossible for a statistical association between task performance and political ideology to exist in their minds prior to participating in the study. However, the benefits of using a task that does not evoke prior associations must be weighed against the costs that arise from reduced external validity (Markman, 2018). The trade-off between internal and external validity is somewhat inevitable, as achieving greater experimental control typically requires researchers to abstract away the complexity that is present in real world settings. Nonetheless, a lack of external validity is concerning, especially considering that lab-based experimental paradigms sometimes fail to predict behaviour in naturalistic settings (e.g., Galizzi & Navarro-Martinez, 2019; Schonberg et al., 2011), and steps should be taken to address this where possible. One real-world setting in which people observe evidence pertaining to others' expertise on a trial-by-trial basis is when taking part in quizzes. Much like the Blap task, in quizzes groups are asked questions, are presented with possible answers by others, and receive feedback indicating whose answers were correct. Thus, they provide a more naturalistic setting for learning about others' expertise. Indeed, there are objectively right and wrong answers to quizzes; by substituting blap for quiz questions we were therefore also able to remove one element of deception from the experimental paradigm.

As noted above, one concern with using a quiz task is that participants might have prior beliefs about the types of people who are competent on such tasks. However, as we have already demonstrated that political similarity biases expertise learning on an abstract task, we reasoned that this was not a major issue. Another concern was that quiz questions have a ground truth and therefore we would lose experimental control of participants' accuracy. For this reason, we did not ask participants to provide answers to the quiz questions during the learning stage. Rather, they were instructed to bet on who they thought would answer questions correctly and incorrectly on each trial. As in the previous experiments, participants provided confidence ratings in the choice stage to indicate how certain they were of their answers. Thus, we were able to statistically control for participants' knowledge when examining their information-seeking and advice-utilisation. A related concern was that participants might look up the answers to quiz questions. However, as the studies in this chapter were conducted in person, we were able to ensure that this did not happen.

A final aim of these studies was to reduce the amount of time it took participants to complete the task. The motivation for this was twofold: First, reducing the task length makes the paradigm more scalable. Researchers with budgetary constraints may not want or be able to pay participants to complete an hour and a half long experiment. Shortening the task will therefore facilitate further research efforts on this topic from other labs. Second, shorter tasks place less cognitive demands on participants than longer tasks. Humans find it challenging to maintain focus and remain alert to stimuli for long periods of time, especially when performing a repetitive task (Langner & Eickhoff, 2013). Sustained attention requires mental effort and when people become bored during a task they exhibit increased absentmindedness and mind-wandering, allocating their attention away from the task at hand and thus conserving mental resources (Warm et al., 2008; Pattyn et al., 2008). If participants in our studies do not focus on the information presented to them, they may fail to discriminate between evidence pertaining to expertise and that relating to other characteristics. To reduce the task length in the studies reported here, we considerably cut down the number of trials in the learning and choice stages of the task, removed the learning test (as participants' bets gave us an indication of how well they learned about the sources), and presented the responses of all four sources jointly during the learning stage. The downside of this latter change is that it increases the amount of information that needs to be processed on each trial. It therefore may not serve to reduce the cognitive burden placed on participants but does reduce the time it takes for participants to complete the experimental tasks.

Overview of Experiments

The present chapter aims to replicate and extend the findings from Chapter 2. The two studies presented here (i.e., Study 3 and 4) were run in immediate succession, as part of an Introductory Psychology lab class. In Study 3, we ask whether learning about others' generosity (rather than political similarity, as in Study 1 and 2) interferes with the ability to learn about and utilise expertise in an unrelated (general knowledge quiz) task. In Study 4, we manipulated political similarity (as in Study 1 and 2) rather than generosity, so that if we found null results in Study 3, as turned out to be the case, we could assess whether this failure to replicate the findings of Chapter 2 was due to the substitution of the political similarity manipulation for the generosity manipulation or the other changes to the experimental paradigm and setting that were made.

In both studies, we use a modelling approach to explore the computational mechanisms underlying how people learn from social evidence. In the first stage of the experiment, participants bet on how others would answer different types of questions: In Study 3 they bet on whether others would donate part of a monetary endowment to each of a series of charities, whereas in Study 4 they bet on whether others would give the same answers as them on political questions. In both studies, they also bet on how accurately others would answer general knowledge questions. In both studies, participants were subsequently presented with general knowledge questions and asked to decide who to seek information from. They earned points for correct answers and lost points for incorrect answers in both experimental stages. The participant in each study with the most points won a £40 cash bonus, thus providing an incentive to answer questions accurately throughout the experiment.

Study 3

Method

Participants

This study was conducted as part of an Introductory Psychology course at University College London. All participants were first-year undergraduate students enrolled in a BSc Psychology degree. The sample size was therefore not based on a power analysis but was determined according to the number of students enrolled on the course. However, a power analysis using data from Study 1 revealed that a sample size of 50 participants would achieve 95% power to detect the smallest effect size of interest (the effect of political similarity on source influence, η_p^2 = .09) with an alpha of .05, assuming a correlation among repeated measures of 0.269 (the observed correlation among repeated measures in Study 1). The class was split into two groups: the first group (n = 53) completed this study (i.e., Study 3), while the second group (n = 51) completed Study 4. Data from one participant who took part in Study 3 had to be excluded from the analysis due to a technical error, so the final sample size was n = 52 (43 females, 9 males; mean age = 18.79, SD = 0.75).

Participants were not compensated for taking part in this study, as it was part of a learning exercise; after the study was completed, the data was pre-processed, anonymised, and sent to the students enrolled on the course. The students analysed the data and wrote up the results in a lab report. The lab reports were graded and the students received course credits and feedback on their coursework. However, to incentivize good performance on the task, a bonus payment was given to the top-performing participant. Participants were informed before taking part that whoever earned the most points in the experiment would be given a £40 cash bonus in the next lecture. Ethical approval was granted from UCL (SHaPS_2015_AH_017).

Study Design

Participants were told that their task was to learn to predict how four previous participants responded when asked general knowledge questions and when given the opportunity to donate money to charities. As in the previous studies, these "previous participants" were not in fact other people but algorithms designed to respond probabilistically according to the trial type. The study consisted of three stages: a learning stage, a ratings stage, and a choice stage.

Learning Stage. The goal of the learning stage was to give participants an opportunity to learn about the sources' charity donation behaviour and about their expertise on general knowledge questions. On each trial of the learning stage, participants were shown either a general knowledge question ('quiz trials') or the name of a charity ('charity trials') and told that the four previous participants (hereafter, "sources") were shown the same piece of information. Participants were told that on quiz trials the sources were shown two possible answers and asked to identify which one was correct. They were told that on each charity trial the sources were given £1 and asked to decide whether to give half of this money (50p) to the specified charity or keep the whole amount for themselves. The quiz and charity trials were interleaved; the trial type presented first was counterbalanced across participants.

Participants were told that their task in this stage was to try to guess who answered the general knowledge questions correctly and who answered them incorrectly, and to guess who gave money to each charity and who did not. They were informed that they would gain or lose points for each of their predictions. For each accurate prediction they gained 10 points and for each inaccurate prediction they lost 10 points. The points acted as an incentive to perform well on the task because the participant with the most points at the end of the experiment received a £40 cash bonus.

The learning stage consisted of one block of 80 trials (40 quiz trials and 40 charity trials). Responses from all four sources were shown on each trial. Qualtrics' loop and merge tool was used to randomise the order of the questions within each block.

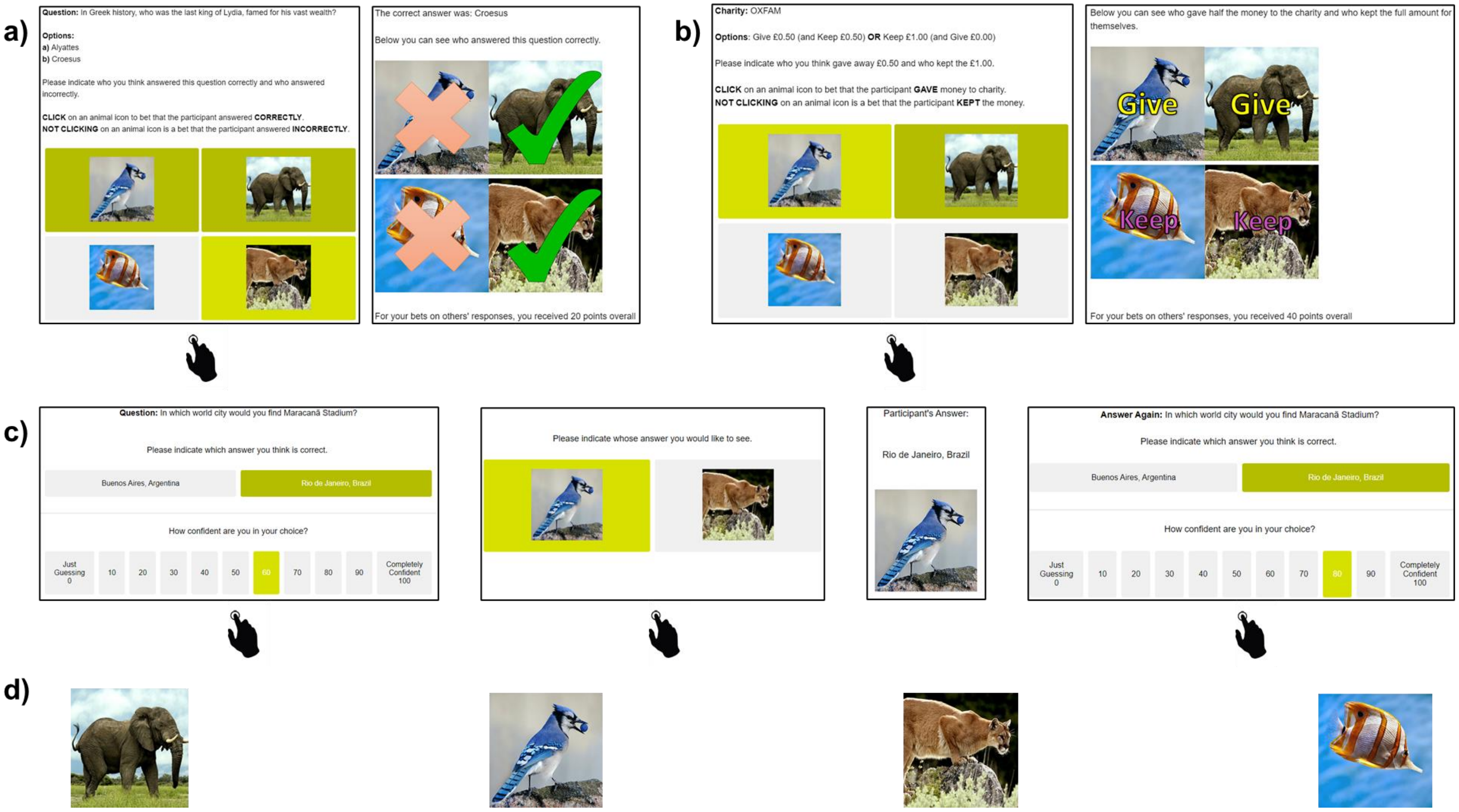
Before starting the learning stage, participants completed four practice trials: two quiz trials and two charity trials. In one practice quiz trial all four sources answered the general knowledge question correctly, while in the other all four sources answered the general knowledge question incorrectly. Likewise, in one practice charity trial all four sources gave money to charity, while in the other none of the sources gave money to charity. Participants were then asked three attention check questions (e.g., "How many participants from our previous study will you learn about during this task?") to make sure they had understood the instructions. If a participant answered one of these attention check questions incorrectly they were told that they had given the wrong answer and asked to answer again until they answered correctly. They were also told that they should not make notes during this task to aid their memory or look up answers on the internet.

Quiz Trials. On each quiz trial, one of 40 general knowledge questions was presented on screen, along with two possible answers (Figure 13a). The quiz questions were taken from various pub quiz websites (e.g., https://pubquizquestionshq.com) and designed for participants to be difficult to answer. Animal pictures representing the four sources were presented in a 2x2 matrix below the answer options. Participants were asked to indicate who they thought answered the quiz question correctly and who answered incorrectly (self-paced). They were told that clicking on a source would indicate that they thought the source would be correct and not-clicking on a source would indicate that they thought the source would be incorrect. After making their predictions, they were presented with a feedback screen, showing them the correct answer, whether each of the four sources answered correctly or not, and how many points they earned overall on the trial (self-paced). The feedback was manipulated so that two (‘accurate’) sources answered correctly with 80% probability on each quiz trial and two (‘inaccurate’) sources answered correctly with 50% probability on each quiz trial (i.e., performed at chance).

Charity Trials. On each charity trial, one of 40 charities was presented on screen, along with two options (“Give £0.50 (and Keep £0.50) **OR** Keep £1.00 (and Give £0.00)”) (Figure 13b). The charities were taken from a list of Britain's top 1,000 charities, ranked by donations (https://www.theguardian.com/news/datablog/2012/apr/24/top-1000-charities-donations-britain). The animal pictures representing the four sources were presented in a 2x2 matrix below this. Participants were asked to indicate who they thought gave money to the charity and who kept the money for themselves (self-paced). Participants were randomly assigned to either a congruent clicking condition or an incongruent clicking condition, using the block randomisation feature in Qualtrics. Those assigned to the congruent clicking condition were told to click on a source to bet that they gave money to charity and not to click on them to bet that they kept all the money. Those assigned to the incongruent clicking condition were told to click on an animal icon to bet that the source would keep all the money and not to click on them to bet that they would donate half of the money to the charity. This counterbalancing was performed to cancel out any influence of habitual (“model-free”) betting behaviour across the quiz and charity trial types. After making their predictions, participants were presented with a feedback screen, showing them whether each of the four sources gave money to charity or not, and how many points they earned overall on the trial (self-paced). The feedback was manipulated so that two (‘generous’) sources (one that was accurate and one that was inaccurate on quiz trials) gave money with 80% probability on each charity trial and two (‘selfish’) sources (one that was accurate and one that was inaccurate on quiz trials) gave money with 20% probability on each charity trial.

Figure 13

Experimental Design of the Task Used in Study 3



Question: In Greek history, who was the last king of Lydia, famed for his vast wealth?

Options:
a) Alyattes
b) Croesus

Please indicate who you think answered this question correctly and who answered incorrectly.

CLICK on an animal icon to bet that the participant answered **CORRECTLY**.
NOT CLICKING on an animal icon is a bet that the participant answered **INCORRECTLY**.

The correct answer was: Croesus

Below you can see who answered this question correctly.

For your bets on others' responses, you received 20 points overall

Charity: OXFAM

Options: Give £0.50 (and Keep £0.50) **OR** Keep £1.00 (and Give £0.00)

Please indicate who you think gave away £0.50 and who kept the £1.00.

CLICK on an animal icon to bet that the participant **GAVE** money to charity.
NOT CLICKING on an animal icon is a bet that the participant **KEPT** the money.

Below you can see who gave half the money to the charity and who kept the full amount for themselves.

For your bets on others' responses, you received 40 points overall

Question: In which world city would you find Maracanã Stadium?

Please indicate which answer you think is correct.

Buenos Aires, Argentina | Rio de Janeiro, Brazil

How confident are you in your choice?

Just Guessing 0 | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | Completely Confident 100

Please indicate whose answer you would like to see.

Participant's Answer:

Rio de Janeiro, Brazil

Answer Again: In which world city would you find Maracanã Stadium?

Please indicate which answer you think is correct.

Buenos Aires, Argentina | Rio de Janeiro, Brazil

How confident are you in your choice?

Just Guessing 0 | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | Completely Confident 100

Note. During the Learning Stage participants learned about the sources' accuracy on a quiz task and generosity on a charity donation task. (a) Quiz trials and (b) charity trials were interleaved. (a) On each quiz trial a novel general knowledge question was presented and participants had to bet on which sources answered it correctly. They then saw which of four sources answered correctly and incorrectly, as well as how many points they earned for their bets. (b) On charity trials, the name of a charity was presented and participants had to bet on which sources gave half of an endowment of £1 to the charity. Whether participants clicked or did not click to bet that a source would agree with them on these trials was counterbalanced to cancel out any influence of habitual ("model-free") betting behaviour across the quiz and charity trial types (c) During the Choice Stage participants completed quiz trials only. On each quiz trial a novel question was presented and the participant had to indicate which answer they thought was correct and enter a confidence rating. They were then presented with two sources and asked to choose whose answer they would like to see. They then saw the response of the chosen source. Finally, they were given a chance to update their initial answer and confidence rating. Responses were all self-paced unless otherwise stated. (d) There were four sources represented with animal photos which the participants were led to believe were other participants but were in fact algorithms.

Sources. The same animal pictures used in the previous studies were used here to represent the sources (Figure 13d). The source condition assigned to each animal picture (i.e., Accurate/Generous, Accurate/Selfish, Inaccurate/Generous, Inaccurate/Selfish) was randomised using the block randomisation feature in Qualtrics, however the order that the animal pictures were presented on screen remained constant. Therefore, the bird always appeared in the top left corner, for example, but whether the bird was generous or selfish in the charity trials, and accurate or inaccurate on quiz trials, was randomised across participants.

Ratings Stage. Participants then rated each source on: (1) how competent they were at answering quiz questions ("How competent was the source at answering general knowledge questions?" on a 6-point scale from "Very Incompetent" to "Very Competent") and (2) how generous they were about giving money to charity ("How generous was the source on the charity rounds?" on a 6-point scale from "Very Ungenerous" to "Very Generous"). The order in which each source was presented in the ratings stage was randomised, using the block randomisation feature in Qualtrics.

Choice stage. On each of 24 trials, participants were presented with a novel general knowledge question, along with two possible answers, and asked to indicate which they thought was correct (self-paced) (Figure 13c). They subsequently rated their confidence in this decision (self-paced) on a scale from 0 (Just Guessing) to 100 (Completely Confident). Next, participants were presented with a pair of sources and asked whose response they wanted to see (self-paced). Qualtrics' choice randomisation feature was used to randomise which two sources were presented on each trial. Unfortunately, this feature balanced the number of times each source was presented across rather than within participants, meaning that some participants were able to hear from a source on more than 50% of trials while others could hear from that same source on less than 50% of trials. We accounted for this by using a generalised linear mixed model to analyse the data rather than averaging the choices made by each participant per condition and performing a repeated-measures (rm) ANOVA, as was done in Study 1 and 2 (see below for more details). Participants were then shown the response of the chosen source (self-paced). The source's response was programmed to be correct with 50% probability. This should not have affected participants' estimates of the sources' competence because no feedback was provided in the choice stage. Thereafter, the general knowledge question was presented again, and participants were asked again to indicate which of the two answers they thought was correct (self-paced). Lastly, participants rated their confidence in their final decision (self-paced). Participants gained 10 points for each correct answer and lost 10 points for each incorrect answer they gave during this stage, although this feedback was not displayed to them.

Results

Manipulation Checks

As the percentage of quiz trials on which the sources gave correct answers was not hard-coded, but rather based on the probability of each source answering each question correctly, we first examined whether (i) our accuracy manipulation was successful and (ii) source accuracy did not vary as a function of source generosity. We did this by entering the percentage of accurate answers made by each source observed by each participant into a 2 (source accuracy condition: accurate, inaccurate) x 2 (source generosity condition: generous, selfish) rmANOVA.

On average, the accurate sources answered approximately 80% of the general knowledge questions correctly (Accurate/Generous: M = 80.19, SD = 6.77. Accurate/Selfish: M = 81.15, SD = 5.52) and the inaccurate sources answered approximately 50% correctly (Inaccurate/Generous: M = 50.96, SD = 7.67. Inaccurate/Selfish: M = 52.93, SD = 8.38). An rmANOVA revealed that sources that were programmed to be more accurate did indeed give accurate responses more often than sources programmed to be inaccurate (F(1,51) = 751.79, p < .001, η_p^2 = .94). There was no main effect of source generosity (F(1,51) = 2.58, p = .114, η_p^2 = .048), suggesting that, overall, generous sources were no more accurate than selfish sources. The interaction between source accuracy and generosity was not significant (F(1,51) = 0.26, p = .613, η_p^2 = .01).

Likewise, we examined whether the algorithm used to manipulate generosity produced the desired pattern of source responses in the charity trials of the learning stage. To this end, we entered the percentage of charitable donations made by each source observed by each participant into a 2 (source accuracy condition: accurate, inaccurate) x 2 (source generosity condition: generous, selfish) rmANOVA.

On average, the generous sources gave money on approximately 80% of charity trials (Accurate/Generous: M = 79.90, SD = 5.88. Inaccurate/Generous: M = 80.43, SD = 5.64) and the selfish sources gave money on approximately 20% of these trials (Accurate/Selfish: M = 20.53, SD = 6.29. Inaccurate/Selfish: M = 20.00, SD = 6.38). An rmANOVA revealed that sources that were programmed to be more generous gave money to charities more often than sources programmed to be selfish (F(1,51) = 5279.26, p < .001, η_p^2 = .99). There was no main effect of source accuracy (F(1,51) < 0.01, p > .999, η_p^2 < .01), suggesting that, overall, accurate sources were no more generous than inaccurate sources. The interaction between source accuracy and generosity was not significant (F(1,51) = 0.38, p = .54, η_p^2 = .01).

Participants' Information-Seeking Choices Were Unaffected by Generosity

We next examined who participants chose to hear from when answering general knowledge questions in the choice stage. As the mean accuracy and generosity of each source, as well as the number of times each source was presented during the choice stage, varied between participants, we used a generalised linear mixed-effects model (GLME) to analyse the choice data. Specifically, we entered participants' choices of whom to seek information from in the choice stage as the dependent variable in a GLME with a binomial response variable distribution. Source choice was coded as 0 if the participant chose the source presented on the left and as 1 if the participant chose the source presented on the right. The difference in source accuracy and the difference in source generosity between the source presented on the right and left were included as both fixed and random factors. The interaction between the source accuracy difference and the source generosity difference was included as a fixed and as a random factor. The difference in source accuracy was calculated by subtracting the percentage of quiz questions the left-hand source answered correctly from the percentage the right-hand source answered correctly. To illustrate, if the participant was given the choice between a source who was accurate on 50% of quiz questions in the learning stage (presented on the left) and another who was accurate on 80% (presented on the right), the difference in source accuracy would be coded as 30 (or -30 if the order of presentation were reversed). Likewise, the difference in source generosity was calculated by subtracting the percentage of times the left-hand source gave money to charity from the percentage of times the right-hand source gave money to charity. If the participant was given the choice between a source who gave money on 20% of the charity trials (presented on the left) and another who gave money on 80% (presented on the right), the difference in source generosity would be coded as 60 (or -60 if the order of presentation were reversed). These predictor variables were standardised (z-scored) before being entered into the GLME. Subject ID was entered as a random factor (grouping variable).

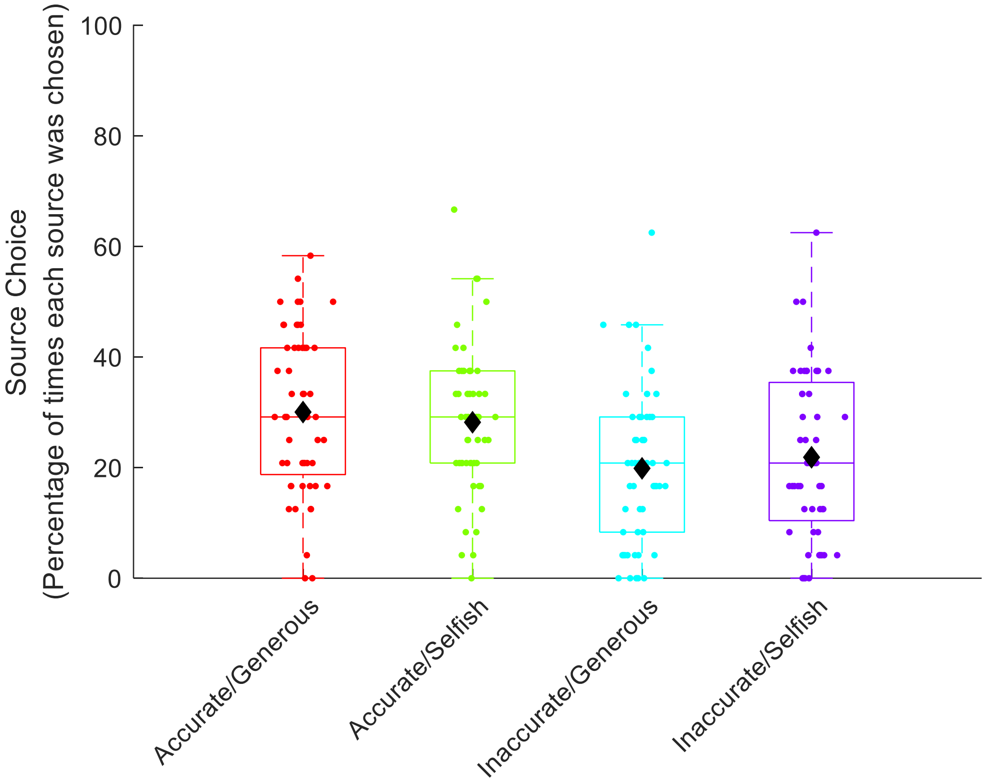
The GLME revealed that the probability of choosing to hear from the source presented on the right-hand side increased with the (standardized) difference in source accuracy (β = 0.66, SE = 0.19, 95% CIs = [0.29, 1.03], t(1244) = 3.51, p < .001). That is, participants preferred to seek information on general knowledge questions in the choice stage from sources that were more accurate on quiz trials in the learning stage (Figure 14). In contrast, the (standardized) difference in source generosity did not affect participants' information-seeking decisions (β = 0.08, SE = 0.16, 95% CIs = [-0.22, 0.39], t(1244) = 0.53, p = .595). The interaction between source accuracy and source generosity was not significant (β = -0.03, SE = 0.09, 95% CIs = [-0.20, 0.15], t(1244) = -0.28, p = .778). The intercept of the model was on the cusp of significance (β = 0.16, SE = 0.08, 95% CIs = [0.00, 0.31], t(1244) = 1.96, p = .050), suggesting that participants had a general tendency to choose sources that were presented on the right-hand side.

To ensure that the results were not confounded by participants knowing the answers to some quiz questions and not others, we re-ran the above analysis with participants' confidence in their initial answer included in the model. In particular, the initial confidence ratings, the interaction between initial confidence and the source accuracy difference, the interaction between initial confidence and the source generosity difference, and the three-way interaction between initial confidence, the source accuracy difference and the source generosity difference were all included as both fixed and random factors in the GLME. Controlling for participants' confidence in their initial answers did not alter the main results. The effect of source accuracy was still significant (β = 0.57, SE = 0.22, 95% CIs = [0.14, 1.00], t(1240) = 2.60, p = .010) and the effect of source generosity was still not significant (β = 0.03, SE = 0.18, 95% CIs = [-0.31, 0.37], t(1240) = 0.15, p = .879).

We also calculated the percentage of times each participant chose to hear from each source and entered these values into a 2 (source accuracy condition: accurate, inaccurate) x 2 (source generosity condition: generous, selfish) rmANOVA, in order to facilitate direct comparison with the results reported in Chapter 2 and graphically display the data in a format that was easily interpretable. The rmANOVA similarly revealed a main effect of source accuracy (F(1,51) = 10.42, p = .002, η_p^2 = .17), no main effect of source generosity (F(1,51) < 0.01, p = .973, η_p^2 < .01), and no interaction between source accuracy and source generosity (F(1,51) = 1.12, p = .295, η_p^2 = .02).

Figure 14

Percentage of Trials on Which Participants Chose to Seek Information from Each Source



Note. Participants preferred to receive information on general knowledge questions from the more accurate sources, regardless of how generous they were. To facilitate interpretation, in this figure we have plotted the percentage of times each participant selected to hear from each source (coloured dots). The black diamonds represent the mean of these percentages. The box plots show the distribution of these percentages: boxes indicate 25–75% interquartile range, whiskers extend from the first and third quartiles to most extreme data point within

1.5 × interquartile range, and the median is shown as a horizontal line within this box.

Participants' Change of Mind Did Not Vary According to Whom They Received Information From

The above analysis was performed to determine who participants chose to seek information from during the choice stage. Next, we explored whether the impact of said information on participants' final answers, and their confidence in those answers, varied according to who it came from.

To quantify the sources' influence on a participants' judgments, we used the Change of Mind (COM) measure described in more detail in Study 1. A linear mixed model was used to assess whether the chosen sources' accuracy on quiz trials and generosity on charity trials affected COM. As we used a linear mixed model, and therefore did not enter average COM scores per source per participant into the analysis and instead entered trial-level COM scores, there was no missing data and thus no need to impute zeros, as in Chapter 2. COM was entered as the dependent variable; source accuracy (i.e. the percentage of times the chosen source answered general knowledge questions correctly in the learning stage, z-scored), source generosity (i.e. the percentage of times the chosen source gave money to charity, z-scored), a variable indicating whether the source agreed or disagreed with the participant's answer on each trial, and their interactions were all entered as fixed and random factors; and Subject ID was entered as a random (grouping) factor.

The linear mixed model revealed that COM did not vary according to the accuracy of the chosen source (β = 2.26, SE = 1.50, 95% CIs = [-0.68, 5.19], t(1240) = 1.51, p = .132) or the generosity of the chosen source (β = -0.06, SE = 1.04, 95% CIs = [-2.11, 1.99], t(1240) = -0.06, p = .956). The interaction between source accuracy and source generosity was not significant (β = 1.06, SE = 1.28, 95% CIs = [-1.46, 3.58], t(1240) = 0.83, p = .409). This suggests that participants' beliefs about how accurate their initial answers to general knowledge questions were did not update as a function of which source they received information from. As the intercept of the model was significant (β = 24.05, SE = 1.95, 95% CIs = [20.23, 27.87], t(1240) = 12.35, p < .001), we can conclude that participants' beliefs about how likely each answer was to be correct were influenced by the sources' answers. Participants exhibited a greater COM when sources disagreed with their initial answers than when they agreed (β = 8.16, SE = 1.58, 95% CIs = [5.05, 11.26], t(1240) = 5.15, p < .001). All other interactions were non-significant (all p-values > .10).

Generosity Learning Did Not Affect Expertise Learning

Competence Ratings. The above results suggest that participants sensibly chose to hear from accurate sources and were not influenced by the sources' generosity when receiving information from others on general knowledge questions. Is this because beliefs about generosity have no effect on expertise learning? To test this, we first examined participants' ratings of the sources' competence on quiz trials (Figure 15). Competence ratings were entered as the dependent variable into a linear mixed-effects model. Source accuracy (i.e., the percentage of times the source answered general knowledge questions correctly), source generosity (i.e., the percentage of times the source gave money to charities), and their interactions were entered as fixed and random factors, and Subject ID was entered as a random factor (grouping variable).

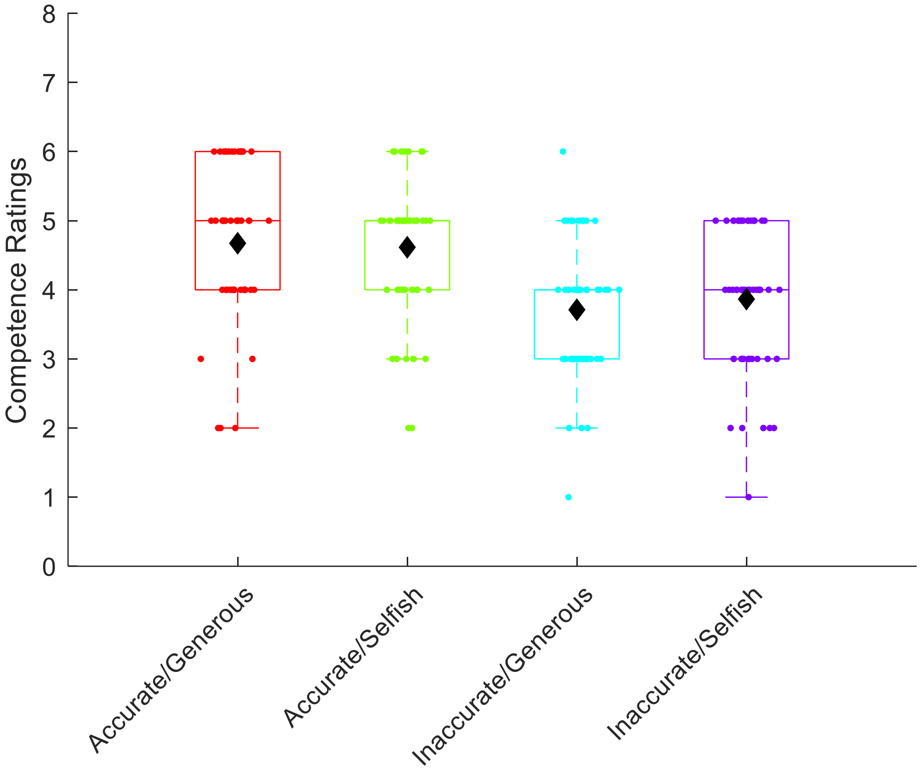
The linear mixed-effects model revealed a significant effect of source accuracy on participants' perceptions of which sources were competent on the quiz trials (β = 0.40, SE = 0.08, 95% CIs = [0.23, 0.56], t(204) = 4.78, p < .001) and no effect of source generosity (β = -0.02, SE = 0.07, 95% CIs = [-0.15, 0.11], t(204) = -0.30, p = .767). The interaction was also not significant (β = 0.06, SE = 0.05, 95% CIs = [-0.04, 0.16], t(204) = 1.24, p = .218).

To facilitate direct comparison with the results reported in Chapter 2, we also entered participants' ratings of source competence into a 2 (source accuracy condition: accurate, inaccurate) x 2 (source generosity condition: generous, selfish) rmANOVA. This analysis produced the same pattern of results. Specifically, there was a significant effect of source accuracy on participants' competence ratings (F(1,51) = 22.94, p < .001, η_p^2 = .31), no effect of source generosity (F(1,51) =

0.11, p = .744, $\eta_p^2 < .01$), and no interaction between source accuracy and generosity (F(1,51) = 0.92, p = .341, $\eta_p^2 = .02$).

Figure 15

Participants' Ratings of Each Source's Competence



Note. Participants rated sources that answered more quiz questions correctly as more competent at answering general knowledge questions, regardless of how generous they were on the charity trials. The coloured dots represent each participant's rating of each source. The black diamonds represent the mean of these ratings. The box plots show the distribution of the competence ratings: boxes indicate 25–75% interquartile range, whiskers extend from the first and third quartiles to most extreme data point within 1.5 × interquartile range, and the median is shown as a horizontal line within this box.

Betting Behaviour on Quiz Trials. We subsequently examined whether participants' bets in the quiz trials of the learning stage were influenced by source generosity in the charity trials. If participants believed that sources that gave money to charity were more likely to answer general knowledge questions correctly than those who tended not to donate money, we would expect to observe a greater percentage of bets on generous sources being accurate than on selfish sources being accurate in the quiz trials of the learning stage. To test this, we entered the percentage of trials on which each participant bet on each source in the quiz trials of the learning stage into a 2 (source accuracy condition: accurate, inaccurate) x 2 (source generosity condition: generous, selfish) rmANOVA. Note, we did not see the need to confirm the robustness of the results using a linear mixed model because we modelled the betting data using reinforcement-learning and Bayesian models, as discussed below).

The rmANOVA revealed that participants bet that accurate sources would be correct (Accurate/Generous: M = 71.88, SD = 20.09. Accurate/Selfish: M = 69.76, SD = 17.41) more often than inaccurate sources (Inaccurate/Generous: M = 56.63, SD = 19.83. Inaccurate/Selfish: M = 59.28, SD = 19.54; F(1,51) = 21.11, p < .001, η_p^2 = .29). The sources' generosity in the charity trials did not affect participants' betting behaviour in the quiz trials (F(1,51) = 0.01, p = .909, η_p^2 < .01), and the interaction between source accuracy and source generosity was not significant (F(1,51) = 1.05, p = .311, η_p^2 = .02).

Computational Modelling. As the percentage of bets placed on the sources ignores the temporal dynamics of how participants' estimates of each source's competence change as they see more evidence, we also fit four sets of computational models to the betting data from the quiz trials. Modelling the data allowed us to test whether participants learned more about the sources when observing congruent evidence (i.e., feedback suggesting that generous [selfish] sources were more [less] competent on quiz trials) than incongruent evidence (i.e., feedback suggesting that selfish [generous] sources were more [less] competent on quiz trials).

We built two sets of reinforcement-learning (RL) models and two sets of Bayesian (beta-binomial) models (Table 3). In one set (of each class of model), we assessed whether an unbiased model fit the betting data better than a model that assumed participants learned differently about the competence of generous and selfish sources, where generosity was defined according to the underlying probability of the source giving money to charity. The congruence bias models in this set assumed that participants categorised the sources by generosity from the outset to determine how much to update their beliefs about the sources' competence in light of the evidence with which they were presented on quiz trials. In reality, however, participants learned about the sources' generosity throughout the learning stage and could therefore not have had a differential updating rule from the first trial (unless they saw a charity trial first). Furthermore, over the course of the learning stage, it was possible for their beliefs about which sources were relatively generous to change. The second set of models were designed to deal with this problem. That is, these models assumed that what participants learned about each source's generosity in the charity trials affected how much they updated their beliefs in response to the evidence they saw in the quiz trials. This latter set of models was parameterised to allow learning to differ depending on whether the model inferred that a participant believed that a source was relatively generous or selfish, compared to the other sources, on the current trial.

Reinforcement Learning Models. The RL models assume that participants update their beliefs about each source using a temporal difference learning rule (Sutton & Barto, 1998). On each trial *t*, the models compute a prediction error for each source (δ_t^s), the difference between their belief about the expertise of source s (Q_t^s) and the evidence they observe (r_t^s) on the trial.

$$\delta_t^s = r_t^s - Q_t^s \tag{15}$$

The ('prior') expected accuracy before the first trial (Q_0) was set at 0.5 for all sources. The participant's estimate of a source's accuracy is updated by adding the product of the prediction error and a learning rate, α, to the participant's estimate from the previous trial.

$$Q_{t+1}^s = Q_t^s + \alpha \cdot \delta_t^s \tag{16}$$

In the congruence bias models, the learning rate was estimated separately for congruent and incongruent evidence (see also, Kuzmanovic et al., 2018; Palminteri et al., 2017). That is, one learning rate, α_1, is used when a generous source is correct on a quiz trial or a selfish source is incorrect (congruent feedback), while a second learning rate, α_2, is used when a generous source is incorrect or a selfish source is correct (incongruent feedback). Therefore, the congruence bias models have an additional free parameter, which may improve the model fit but also increases the model complexity.

Bayesian Models. We conducted an equivalent model comparison to the one above using Bayesian (beta-binomial) models. As noted in Chapter 1, an ideal learner would update their beliefs about a source's characteristics using Bayes theorem. Thus, while RL models have been shown to explain behavioural and neural data in a wide range of social learning tasks (Burke et al., 2010; Chang et al., 2010; Hackel et al., 2015; Hampton et al., 2008; King-Casas, 2005; Suzuki et al., 2012) and are often used to explore biases in learning (e.g., Kuzmanovic et al., 2018; Lefebvre et al., 2017; Palminteri et al., 2017), we opted to corroborate the results of the RL models with a Bayesian approach. Using a Bayesian model additionally allowed us to estimate participants' uncertainty in their beliefs about how accurate or generous the sources were.

The beta-binomial model can be used to compute posterior beliefs from a prior belief and some observed evidence. It does this by combining a binomial likelihood function with a (prior) beta distribution. When assessing source competence on quiz trials, the likelihood of observing a source answer a particular number of questions correctly (k_t^s) is determined by the source's expertise (Q_t^s) and the number of questions that have been asked, t.

$$k_t^s | t, Q_t^s \sim Binomial(Q_t^s, t) \tag{17}$$

The participant's prior belief about the source's expertise is described by a beta distribution.

$$Q_t^s|\beta1_t^s, \beta2_t^s \sim Beta(\beta1_t^s, \beta2_t^s) \tag{18}$$

The beta distribution has two parameters, β1 and β2 (the first is typically denoted by an α, however, to avoid confusing this parameter with the learning rate in the RL models we chose to denote the parameters using two beta symbols). These parameters dictate the shape of the distribution (i.e., its skew, mean, and variance).

The parameters in the beta distribution are updated every time the participant sees new evidence pertaining to the source's competence using the following equation:

$$Q_t^s|k_t^s, t, \beta1_t^s, \beta2_t^s \sim Beta(\beta1_t^s + k_t^s, \beta2_t^s + t - k_t^s) \tag{19}$$

That is, the probability of a source answering quiz questions correctly can be estimated by adding the number of observed correct answers to the β1 parameter associated with a given source and adding the number of incorrect answers observed to the β2 parameter associated with said source.

To allow models to assume that participants learned more from some types of evidence than others, we added one or more scaling parameter(s), γ. Specifically, the posterior beta distribution parameters were computed by adding the product of the observed evidence and the scaling parameter to the prior parameters.

$$Q_t^s|k_t^s, t, \beta1_t^s, \beta2_t^s \sim Beta(\beta1_t^s + \gamma \cdot k_t^s, \beta2_t^s + \gamma \cdot (t - k_t^s)) \tag{20}$$

In the congruence bias models, the scaling parameter was estimated separately for congruent and incongruent evidence. That is, one scaling parameter, γ_1, was used to determine how much participants learn when a generous source is accurate or a selfish source is inaccurate (congruent feedback) on a quiz trial, while a second scaling parameter, γ_2, was used to determine how much participants learn when a generous source is incorrect or a selfish source is correct (incongruent feedback). As in the RL congruence bias models, these models thus have an additional free parameter, which may improve the model fit but also increases the model complexity and thus the BIC and AIC.

Note, even though Q_t^s is a probability here (e.g., in the case of an accurate source, P(0.8)), the beta distribution actually describes a probability density function. The mean estimate (denoted below by $\hat{\beta}_t^s$) of this density function is calculated as follows:

$$\hat{\beta}_t^s = \frac{\beta 1_t^s}{\beta 1_t^s + \beta 2_t^s} \tag{21}$$

Thus, the mean estimate of a source's competence will be 0.5 when $\beta 1_t^s = \beta 2_t^s$. If $\beta 1_t^s > \beta 2_t^s$, then the distribution has greater mass on the right, indicating that the participant believes that the source will answer quiz questions correctly more often than not. If $\beta 1_t^s < \beta 2_t^s$, then the distribution has greater mass on the left, indicating that the participant believes that the source will answer quiz questions correctly with a probability less than 0.5.

In the Bayesian models, we estimated the parameters that dictated each participant's prior belief about the sources' expertise (i.e., $\beta 1_{t=0}, \beta 2_{t=0}$) by fitting the models outlined in Table 3 to the betting data. Participants were assumed to hold the same prior beliefs about all four sources.

Table 3

Model Specifications

No.	**Model**
RL Models	
1. RL Unbiased	$Q_{t+1}^s = Q_t^s + \alpha \cdot \delta_t^s$
2. RL Congruence Bias	If r_t^s = 1 $Q_{t+1}^{s\_Generous} = Q_t^{s\_Generous} + \alpha_1 \cdot \delta_t^s$ $Q_{t+1}^{s\_Selfish} = Q_t^{s\_Selfish} + \alpha_2 \cdot \delta_t^s$ If r_t^s = 0 $Q_{t+1}^{s\_Generous} = Q_t^{s\_Generous} + \alpha_2 \cdot \delta_t^s$ $Q_{t+1}^{s\_Selfish} = Q_t^{s\_Selfish} + \alpha_1 \cdot \delta_t^s$
Bayesian Models	

3. BB Unbiased	$Beta(\beta 1^{s}_{t+1}, \beta 2^{s}_{t+1}) = Beta(\beta 1^{s}_{t} + k^{s}_{t}, \beta 2^{s}_{t} + n - k^{s}_{t})$
4. BB with 1 Scaling Parameter	$Beta(\beta 1^{s}_{t+1}, \beta 2^{s}_{t+1}) = Beta(\beta 1^{s}_{t} + \gamma \cdot k^{s}_{t}, \beta 2^{s}_{t} + \gamma \cdot (n - k^{s}_{t}))$
5. BB Congruence Bias	$Beta\big(\beta 1^{s\_Generous}_{t+1}, \beta 2^{s\_Generous}_{t+1}\big) = Beta(\beta 1^{s\_Generous}_{t} + \gamma_1 \cdot k^{s}_{t}, \beta 2^{s\_Generous}_{t} + \gamma_2 \cdot (n - k^{s}_{t}))$ $Beta\big(\beta 1^{s\_Selfish}_{t+1}, \beta 2^{s\_Selfish}_{t+1}\big) = Beta(\beta 1^{s\_Selfish}_{t} + \gamma_2 \cdot k^{s}_{t}, \beta 2^{s\_Selfish}_{t} + \gamma_1 \cdot (n - k^{s}_{t}))$

Note. RL = Reinforcement-learning, BB = Beta-binomial. Each model in this table was fit to the betting data from the quiz trials of the learning stage (Set 1). In Set 1, the generosity of the sources was classified according to the underlying probability of the source giving money to charity. Each model was also fit to the betting data from the quiz and charity trials in the learning stage simultaneously (Set 2). In Set 2, source generosity was updated in light of the observed evidence using either the unbiased RL or unbiased beta-binomial model. The source(s) that were estimated to be relatively generous (i.e., more generous than the mean generosity of the four sources) on a given trial were classified as 'generous', while those that were estimated to be relatively selfish (i.e., less generous than the mean generosity of the four sources) were classified as 'selfish'.

Model Fitting and Comparison. We used each participant's bets in the learning stage to fit the RL models and find the individual-level best-fit values of the model parameters. As done by Leong and Zaki (2018), we assumed that the relationship between a participant's estimate of a source being correct or generous and their betting behaviour is described by a logistic function:

$$p(bet_{t+1} = Correct) = \frac{1}{1 + e^{-\tau(E_t - 0.5)}} \tag{22}$$

where τ is a subject-specific free parameter that represents the gain of the logistic function. Each of the models in Table 3 were fit to participants' bets using a maximum-likelihood estimation procedure. The fmincon function in Matlab (version 2019a) was used to find the optimal set of model parameters (i.e., the parameter values that minimized the negative log likelihood). Fit was assessed using the Bayesian Information Criterion (BIC; Schwarz, 1978) and Akaike Information Criterion (AIC; Akaike, 1974):

$$AIC = -2\log lik + 2k \tag{23}$$

$$BIC = -2\log lik + klog(n) \tag{24}$$

where *lik* denotes the maximum likelihood of the data given the model, k the number of free parameters, and *n* the total number of data points (i.e., the product of the number of trials and sources). The lower the BIC and AIC, the better the fit. We calculated the BIC and AIC for each model per participant and then derived a total BIC and AIC score through summation. A Pseudo R-Squared statistic was also calculated (using McFadden's Pseudo R-Squared formula and a null model assuming participants bet on each source with 50% probability on each trial) to assess which model explained the most variance in each participant's choices. Lastly, to help readers interpret the Pseudo R-Squared statistics, we ran 1000 simulations using the best-fit parameters from the unbiased beta-binomial models and assessed the in-sample accuracy of the model's predictions by calculating the average percentage of times the model-predicted betting behaviour matched the participant's actual betting behaviour.

Modelling Results. In summary, in the first model comparison, we compared the performance of an unbiased RL model against a congruence bias RL model, using only the data from the quiz trials. In the second, we also compared an unbiased RL model against a congruence bias RL model, however we modelled how participants learned about source competence in the quiz trials and source generosity in the charity trials in unison. Here, the confirmation bias model was designed so that participants' beliefs about the relative generosity of the sources on a given trial (as inferred from the model) dictated which learning rate was employed when they observed evidence pertaining to source competence. The third and fourth model comparisons were analogous to the former two but employed Bayesian rather than RL models. Notably, because the beta-binomial model does not include a learning rate, we compared three Bayesian models in each set: a standard beta-binomial model; a beta-binomial model with a scaling parameter, which modulated how much the participant was assumed to update their beliefs in light of observed evidence; and a beta-binomial model with two scaling parameters, one of which was applied when the participant observed congruent evidence while the other was applied when the participant observed incongruent evidence. The model statistics are presented below (Table 4).

Table 4

Expertise Learning Model Comparison Results

Model No.	**BIC**	**AIC**	**Mean Pseudo R-Squared**	**% of participants for whom model fit best (BIC)**	**% of participants for whom model fit best (AIC)**
RL models using data from quiz trials only (Set 1)					
1. RL Unbiased	**10478**	10159	0.14	81%	52%
2. RL Congruence Bias	10483	**10004**	0.16	19%	48%
Bayesian models using data from quiz trials only (Set 1)					
3. BB Unbiased	**10482**	10002	0.16	73%	46%
4. BB with 1 Scaling	10689	10050	0.16	6%	12%

Parameter					
5. BB Congruence Bias	10680	**9880**	0.19	21%	42%
RL models using data from quiz and charity trials (Set 2)					
1. RL Unbiased	**19567**	18979	0.19	81%	42%
2. RL Congruence Bias	19686	**18902**	0.20	19%	58%
Bayesian models using data from quiz and charity trials (Set 2)					
3. BB Unbiased	**19743**	**18763**	0.21	85%	50%
4. BB with 1 Scaling Parameter	19952	18776	0.21	10%	23%
5. BB Congruence Bias	20360	18793	0.22	6%	27%

Note: RL = Reinforcement-learning, BB = Beta-binomial.

The model comparisons indicated that unbiased models generally provided a better fit to participants' bets in the quiz trials than congruence bias models (Table 4). As is typical, the AIC tended to favour complex models while the BIC favoured simpler models (Vrieze, 2012). Of note, though, in all but one of the model comparisons, data from a plurality of participants was best explained by an unbiased model when model fit was assessed using the AIC.

In the RL [beta-binomial] congruence bias models, the learning rates [scaling parameters] were estimated separately for congruent and incongruent evidence. Performing Wilcoxon signed rank tests on the learning rates [scaling parameters] revealed that there was no significant difference between them in any of the four congruence bias models, suggesting that there was not a systematic bias to learn more from one type (e.g., congruent) of evidence than the other (Table 5). This is consistent with the results of the model comparison, as well as those we observed when analysing the percentage of bets on each source and when analysing participants' post-learning competence ratings.

Table 5

Wilcoxon Signed Rank Tests Comparing the Magnitude of the Learning Rates [Scaling Parameters] Included in the Congruence Bias Models

Congruence Bias Model	**Median α_1 [γ_1]**	**Median α_2 [γ_1]**	**Z**	**p**
RL, using data from quiz trials only (Set 1)	0.05	0.05	0.14	.891
BB, using data from quiz trials only (Set 1)	0.42	0.29	1.11	0.267
RL, using data from quiz and charity trials (Set 2)	0.04	0.03	-0.49	.623
BB, using data from quiz and charity trials (Set 2)	0.61	0.68	1.02	.308

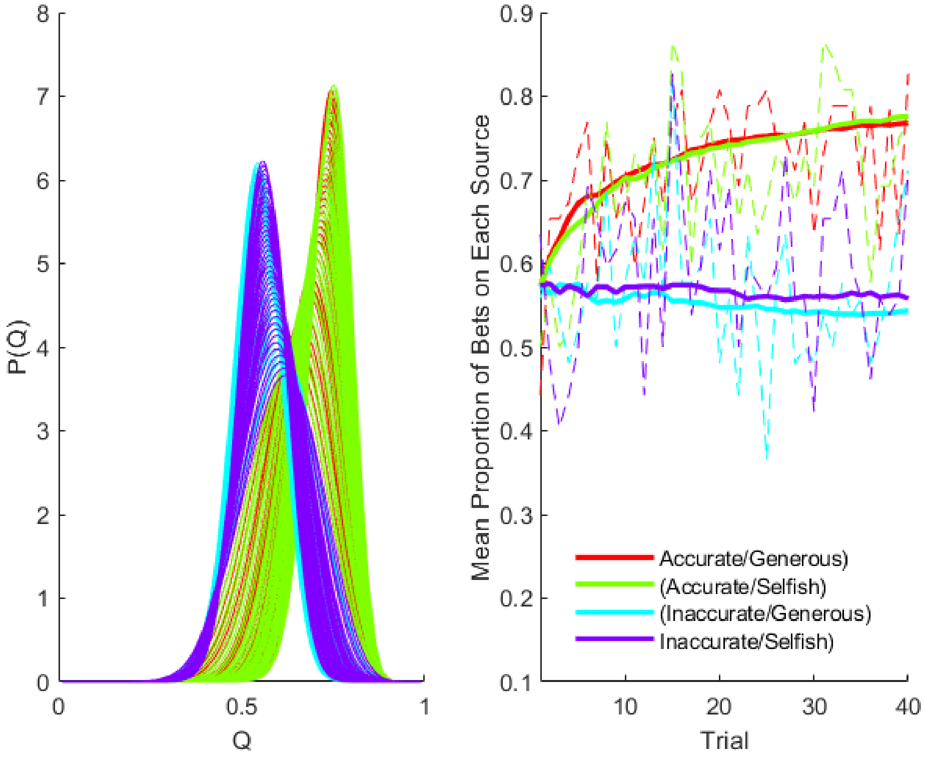
Note. RL = Reinforcement-learning, BB = Beta-binomial. α_1 is the learning rate parameter applied to congruent feedback in the RL models; α_2 is the learning rate parameter applied to incongruent feedback in the RL models; γ_1 is the scaling parameter applied to congruent feedback in the BB models; γ_2 is the scaling parameter applied to incongruent feedback in the BB models.

Notably, participants' bets in the quiz trials were well-described by a standard beta-binomial model (see Figure 16). The Pseudo R-squared statistics may not appear too large, however it is important to bear in mind that a model will not be able to predict a participant's decisions at better than chance levels if that participant estimates a source's accuracy or generosity at P(0.5), as they should have learned to do for the inaccurate sources.

To determine on how many trials the standard beta-binomial model accurately predicted each participant's bets, we used the best-fit parameters to simulate their decisions. For each participant, we simulated the bets that the model would make on each trial 1000 times, given the participant's best-fit values of $\beta1_{t=0}$, $\beta2_{t=0}$, and τ and the evidence they observed throughout the learning stage. For each simulation, we measured the percentage of trials on which the model accurately predicted the participant's betting behaviour. We then computed the mean accuracy of the model by averaging across the simulations. This revealed that, overall, the standard beta-binomial model accurately predicted participants' bets on 60% of quiz trials. Specifically, the model predicted 64.92% of participant's bets on the Accurate/Generous source, 63.35% of bets on the Accurate/Selfish source, 55.61% on the Inaccurate/Generous source, and 55.09% on the Inaccurate/Selfish source.

Figure 16

A Standard Beta-Binomial Model Fit to Participants' Bets on Each Source in the Quiz Trials



Note. Left-hand side: The probability distributions illustrate how participants' beliefs about each source's expertise (Q) evolved over the course of the learning stage. Each distribution was calculated by averaging the model parameters for each trial across participants. The distributions from each trial are plotted one on top of the other. Before observing any evidence pertaining to the sources' competence at answering quiz questions, the prior distribution did not vary by source. Participants' beliefs about each source were updated on each trial in light of the evidence they saw. Over the course of the learning stage, the model suggests that participants became less uncertain in their beliefs – as evidenced by the increasing height of the distributions – and learned which sources were more and less accurate – as evidenced by the leftward movement for inaccurate sources and rightward movement for accurate sources. Right-hand side: Solid lines show the mean model-predicted probability of betting on each source on every quiz trial of the learning stage. Dotted lines show the proportion of participants that actually bet on each source on each quiz trial.

Expertise Learning Did Not Affect Generosity Learning

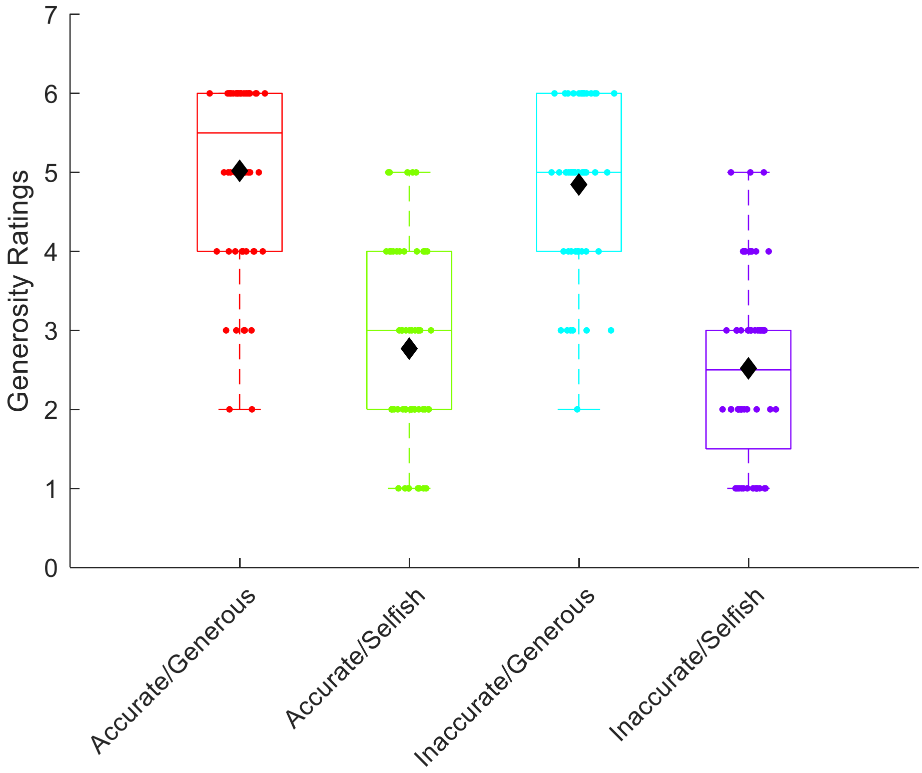
Generosity Ratings. While learning about the sources' generosity on charity trials did not affect participants' ability to learn about the sources' expertise on quiz trials, there may have been an effect in the reverse causal direction. That is, the sources' accuracy on the quiz trials may have influenced participants' perceptions of their generosity. To test this, we entered the generosity ratings as the dependent variable in a linear mixed-effects model. Source accuracy (i.e., the percentage of times the source answered general knowledge questions correctly), source generosity (i.e., the percentage of times the source gave money to charities), and their interactions were all entered as both fixed and random factors, and Subject ID was entered as a random factor (grouping variable).

The linear mixed model revealed a significant effect of source generosity on participants' perceptions of which sources were generous (β = 0.72, SE = 0.06, 95% CIs = [0.60, 0.84], t(204) = 12.13, p < .001), no effect of source accuracy (β = 0.07, SE = 0.04, 95% CIs = [-0.01, 0.16], t(204) = 1.71, p = .088), and no interaction between source accuracy and generosity ($\beta < 0.01$, SE = 0.04, 95% CIs = [-0.08, 0.09], $t(204) = 0.12$, $p = .908$) (Figure 17).

To facilitate direct comparison with the results reported in Chapter 2, we also entered participants' ratings of source generosity on charity trials into a 2 (source accuracy condition: accurate, inaccurate) x 2 (source generosity condition: generous, selfish) rmANOVA. The rmANOVA replicated the above pattern of results; there was a significant main effect of source generosity on participants' generosity ratings ($F(1,51) = 121.12$, $p < .001$, $\eta_p^2 = .70$), no main effect of source accuracy ($F(1,51) = 2.54$, $p = .117$, $\eta_p^2 = .05$), and no interaction between source accuracy and generosity ($F(1,51) = 0.06$, $p = .801$, $\eta_p^2 < .01$).

Figure 17

Participants' Ratings of Each Source's Generosity



Note. Participants rated sources that tended to give money to charities as more generous than those who tended not to, regardless of how accurate they were on the quiz trials. The coloured dots represent each participant's rating of each source.

The black diamonds represent the mean of these ratings. The box plots show the distribution of the generosity ratings: boxes indicate 25–75% interquartile range, whiskers extend from the first and third quartiles to most extreme data point within 1.5 × interquartile range, and the median is shown as a horizontal line within this box.

Betting Behaviour on Charity Trials. If participants believed that sources that tended to answer quiz questions correctly were more likely to donate to charity than those who were less competent on quiz trials, we would expect to observe a greater percentage of bets on accurate sources being generous than on inaccurate sources being generous in the charity trials. To test this, we entered the percentage of trials on which each participant bet that each source would donate money in the charity trials into a 2 (source accuracy condition: accurate, inaccurate) x 2 (source generosity condition: generous, selfish) rmANOVA.

The rmANOVA revealed that participants bet that generous sources would give money (Accurate/Generous: M = 73.37, SD = 18.31. Inaccurate/Generous: M = 70.34, SD = 18.60) more often than selfish sources (Accurate/Selfish: M = 33.89, SD = 18.23. Inaccurate/Selfish: M = 33.03, SD = 21.38; F(1,51) = 72.84, p < .001, η_p^2 = .59). Participants were no more likely to bet on accurate than inaccurate sources (F(1,51) = 1.93, p = .171, η_p^2 = .04). The interaction between source accuracy and source generosity was not significant (F(1,51) = 0.45, p = .507, η_p^2 = .01).

Computational Modelling. We next examined the trial-by-trial dynamics of generosity learning, using the same procedure that was used to explore how participants learned about source competence. As above, we ran four separate model comparisons. Here, models were fit to each participant's betting data in the charity trials and the congruence bias models allowed learning to differ according to the expertise of the source and whether they gave money to the charity or not on a given charity trial. The models were thus parameterised as shown in Table 6.

Table 6

Model Specifications

No.	Model
RL Models	
1. RL Unbiased	$Q_{t+1}^{s} = Q_{t}^{s} + \alpha \cdot \delta_{t}^{s}$
2. RL Congruence Bias	If r_{t}^{s} = 1 $Q_{t+1}^{s\_Accurate} = Q_{t}^{s\_Accurate} + \alpha_1 \cdot \delta_{t}^{s}$ $Q_{t+1}^{s\_Inaccurate} = Q_{t}^{s\_Inaccurate} + \alpha_2 \cdot \delta_{t}^{s}$ If r_{t}^{s} = 0 $Q_{t+1}^{s\_Accurate} = Q_{t}^{s\_Accurate} + \alpha_2 \cdot \delta_{t}^{s}$ $Q_{t+1}^{s\_Inaccurate} = Q_{t}^{s\_Inaccurate} + \alpha_1 \cdot \delta_{t}^{s}$
Bayesian Models	
3. BB Unbiased	$Beta(\beta 1_{t+1}^{s}, \beta 2_{t+1}^{s}) = Beta(\beta 1_{t}^{s} + k_{t}^{s}, \beta 2_{t}^{s} + n - k_{t}^{s})$
4. BB with 1 Scaling Parameter	$Beta(\beta 1_{t+1}^{s}, \beta 2_{t+1}^{s}) = Beta(\beta 1_{t}^{s} + \gamma \cdot k_{t}^{s}, \beta 2_{t}^{s} + \gamma \cdot (n - k_{t}^{s}))$
5. BB Congruence Bias	$Beta\left(\beta 1_{t+1}^{s\_Accurate}, \beta 2_{t+1}^{s\_Accurate}\right) = Beta(\beta 1_{t}^{s\_Accurate} + \gamma_1 \cdot k_{t}^{s}, \beta 2_{t}^{s\_Accurate} + \gamma_2 \cdot (n - k_{t}^{s}))$ $Beta\left(\beta 1_{t+1}^{s\_Inaccurate}, \beta 2_{t+1}^{s\_Inaccurate}\right) = Beta(\beta 1_{t}^{s\_Inaccurate} + \gamma_2 \cdot k_{t}^{s}, \beta 2_{t}^{s\_Inaccurate} + \gamma_1 \cdot (n - k_{t}^{s}))$

Note. RL = Reinforcement-learning, BB = Beta-binomial. Each model in this table was fit to the betting data from the charity trials of the learning stage (Set 1). In Set 1, the accuracy of the sources was classified according to the underlying probability of the source answering quiz questions correctly. Each model was also fit to the betting data from the charity and quiz trials in the learning stage simultaneously (Set 2). In Set 2, source accuracy was updated in light of the observed evidence using either the unbiased RL or unbiased beta-binomial model. The source(s) that were estimated to be relatively generous (i.e., more generous than the mean generosity of the four sources) on a given trial were classified as 'generous', while those that were estimated to be relatively selfish (i.e., less generous than the mean generosity of the four sources) were classified as 'selfish'.

We used the same model fitting and model comparison procedures here as we did when modelling participants' expertise learning. The results of the model comparisons are displayed in Table 7.

Table 7

Generosity Learning Model Comparison Results

Model number and name	**BIC**	**AIC**	**Mean Pseudo R-Squared**	**% of participants for whom model fit best (BIC)**	**% of participants for whom model fit best (AIC)**
RL models using data from charity trials only (Set 1)					
1. RL Unbiased	**9179**	8860	0.25	79%	62%
2. RL Congruence Bias	9283	**8803**	0.26	21%	38%
Bayesian models using data from charity trials only (Set 1)					
3. BB Unbiased	**9241**	**8761**	0.27	90%	69%
4. BB with 1 Scaling Parameter	9468	8829	0.27	2%	8%
5. BB Congruence Bias	9598	8798	0.28	8%	23%
RL models using data from charity and quiz trials (Set 2)					

1. RL Unbiased	**19567**	**18979**	0.19	94%	69%
2. RL Congruence Bias	19764	18981	0.20	6%	31%
Bayesian models using data from charity and quiz trials (Set 2)					
3. BB Unbiased	**19743**	**18763**	0.21	94%	60%
4. BB with 1 Scaling Parameter	19997	18822	0.21	6%	23%
5. BB Congruence Bias	20550	18982	0.21	0%	17%

Note: RL = Reinforcement-learning, BB = Beta-binomial.

The model comparisons indicated that the unbiased models tended to outperform the congruence bias models (Table 7). Here, the unbiased models fit best to the majority of participants in every model comparison, according to both the BIC and AIC. Performing Wilcoxon signed rank tests on the learning rates [scaling parameters] of the RL [beta-binomial] congruence bias models revealed that participants did not systematically learn more from one type of evidence than the other, as the learning rates [scaling parameters] for congruent and incongruent evidence were not significantly different (Table 8). This is consistent with the null main effect of source accuracy on the percentage of bets placed on sources in the charity trials and on participants' generosity ratings.

Table 8

Wilcoxon Signed Rank Tests Comparing the Magnitude of the Learning Rates [Scaling Parameters] Included in the Congruence Bias Models.

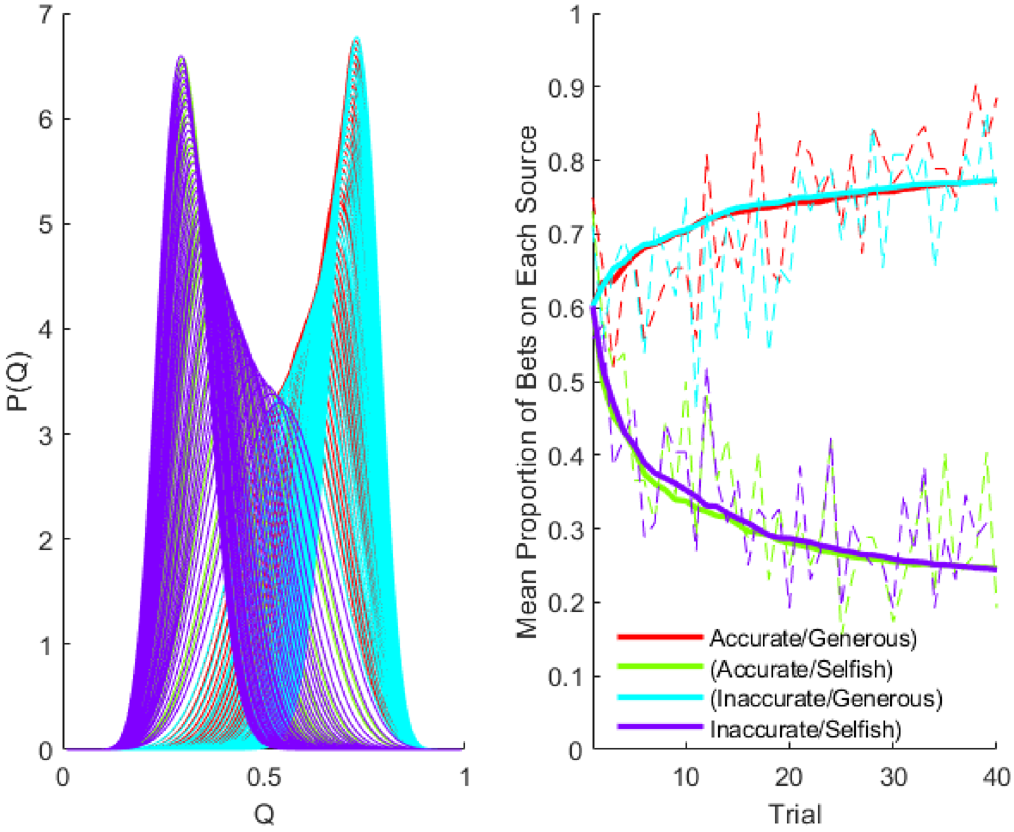
Congruence Bias Model	**Median α_1 [γ_1]**	**Median α_2 [γ_1]**	**Z**	**p**
RL, using data from charity trials only (Set 1)	0.05	0.05	0.35	.729
BB, using data from charity trials only (Set 1)	1.01	0.68	1.38	.169
RL, using data from charity and quiz trials (Set 2)	0.03	0.03	-0.26	.799
BB, using data from charity and quiz trials (Set 2)	0.99	1.09	1.17	.240

Note: RL = Reinforcement-learning, BB = Beta-binomial. α_1 is the learning rate parameter applied to congruent feedback in the RL models; α_2 is the learning rate parameter applied to incongruent feedback in the RL models; γ_1 is the scaling parameter applied to congruent feedback in the BB models; γ_2 is the scaling parameter applied to incongruent feedback in the BB models.

As in the quiz trials, bets in the charity trials were well-described by a standard beta-binomial model (Figure 18). Using the same procedure as above, we used the best-fit parameters from this model to simulate participants' decisions 1000 times and computed the mean accuracy of the model across the simulations. Overall, the model accurately predicted participants' bets on 65% of charity trials. Specifically, the model predicted 67.54% of participant's bets on the Accurate/Generous source, 63.57% of bets on the Accurate/Selfish source, 65.88% on the Inaccurate/Generous source, and 64.44% on the Inaccurate/Selfish source. Again, it is important to note that these figures represent in-sample, rather than out-of-sample, prediction accuracy, as the model was trained on the same data it was then used to predict.

Figure 18

A Standard Beta-Binomial Model Fit to Participants' Bets on Each Source in the Charity Trials



Note. Left-hand side: The probability distributions illustrate how participants' beliefs about each source's generosity (Q) evolved over the course of the learning stage. Each distribution was calculated by averaging the model parameters for each trial across participants. The distributions from each trial are plotted one on top of the other. Before observing any evidence pertaining to the sources' generosity, the prior distribution did not vary by source. Participants' beliefs about each source were updated on each trial in light of the evidence they saw. Over the course of the learning stage, the model suggests that participants became less uncertain in their beliefs – as evidenced by the increasing height of the distributions – and learned which sources were more generous and which were more selfish– as evidenced by the leftward movement for selfish sources and rightward movement for generous sources. Right-hand side: Solid lines show the mean model-predicted probability of betting on each source on every charity trial of the learning stage. Dotted lines show the proportion of participants that actually bet on each source on each charity trial.

Discussion

The results of Study 3 suggest that learning about others' generosity does not bias how people learn about their competence at answering general knowledge questions. Participants were more likely to choose sources that had a history of answering quiz questions correctly when seeking information, regardless of how generous those sources were. They did not perceive generous sources as more competent on quiz trials than selfish sources, nor did they perceive sources that were competent on general knowledge questions as more generous in the charity trials. Computational modelling provided further evidence that participants did not exhibit a systematic congruence bias when updating their beliefs about source expertise or source generosity.

Study 4

Method

Participants

This study was conducted immediately after Study 3, with a separate group of first-year BSc Psychology undergraduate students, under the same conditions as in the previous experiment. 51 participants completed Study 4, however data from one participant had to be excluded from the analysis due to a technical error, leaving a final sample size of n = 50 (45 females, 5 males; mean age = 18.78, SD = 0.68). The compensation procedure was the same as in Study 3. Ethical approval was granted from UCL (SHaPS_2015_AH_017).

Study Design

The design of this study is comparable with that used in Study 3, except here participants were told that their task was to learn to predict how four previous participants responded when asked general knowledge and political questions. Rather than manipulating source generosity, we experimentally manipulated political similarity, as in Study 1 and 2, although here the political stimuli were relevant to UK rather than US politics. Unless otherwise stated, the methods were the same as in Study 3.

Political Trials in the Learning Stage. On every other trial of the learning stage ('political trials'), participants were shown two opposing political statements (e.g., "a) The EU should impose a quota of migrants per country b) The EU should not impose a quota of migrants per country") and informed that the four sources were shown the same piece of information. As political opinions are classed as a 'special category of personal data' in the Data Protection Act 2018 (DPA) and the General Data Protection Regulation (GDPR), we did not record participants' responses to the political questions that were asked in this study. Instead, we randomised whether different viewpoints were assigned to option "a" or "b", using a custom JavaScript function, and asked participants to indicate which option they agreed with more ("a" or "b"). We could not recover information of which viewpoint was presented on which side to each participant. We therefore did not know which political opinion the participant chose, but merely which letter they selected. On the Information screen presented at the beginning of the study, participants were informed that their political opinions would not be recorded in this study.

The political statements were adapted from questions on the website https://uk.isidewith.com/political-quiz. The stimuli covered a broad range of topics, including economic, social, criminal, domestic policy, foreign policy, education, electoral, environmental, healthcare, immigration, science, and transportation issues. All participants saw the same political statements in the same pairings, however the order in which the pairings were presented was randomised using Qualtrics' loop and merge tool. After indicating which answer was more consistent with their political views, participants were asked to bet on which of the sources gave the same answer as them and which answered differently (self-paced). Participants who were randomly assigned to a congruent clicking condition were told to click on an animal icon to bet that the source gave the same answer as them and not to click on an animal icon to bet that the source answered differently to them. Those assigned to the incongruent clicking condition were told to click on an animal icon to bet that the source would disagree with them and not to click on an animal icon to bet that the source would agree with them. After making their bets, participants were presented with a feedback screen, showing them whether each of the four sources agreed or disagreed with their answer, and how many points they earned overall on the trial (self-paced). The feedback was manipulated so that two ('similar') sources (one that was accurate and one that was inaccurate on quiz trials) agreed with the participant's answer with 80% probability on each political trial and two ('dissimilar') sources (one that was accurate and one that was inaccurate on quiz trials) agreed with the participant's answer with 20% probability on each political trial.

Ratings Stage. After completing the learning stage, participants rated each source on: (1) how competent they were at answering quiz questions ("How competent was the source at answering general knowledge questions?" on a 6-point scale from "Very Incompetent" to "Very Competent") and (2) how similar they were to the participant on political issues ("How similar was the source to you in terms of their political views?" on a 6-point scale from "Very Dissimilar" to "Very Similar").

Results

The analyses used in Study 4 are the same as those used in Study 3, except political similarity replaces generosity.

Manipulation Checks

The accuracy manipulation was successful. Sources that were programmed to be accurate answered approximately 80% of the general knowledge questions correctly (Accurate/Similar: M = 81.25, SD = 6.87. Accurate/Dissimilar: M = 80.50, SD = 6.12) while those programmed to be inaccurate answered approximately 50% of the quiz questions correctly (Inaccurate/Similar: M = 50.20, SD = 8.73. Inaccurate/Dissimilar: M = 48.25, SD = 6.93). Entering the percentage of accurate answers for each source per participant into a 2 (source accuracy condition: accurate, inaccurate) x 2 (source similarity condition: similar, dissimilar) rmANOVA revealed a main effect of source accuracy (F(1,49) = 915.93, p < .001, η_p^2 = .95), no main effect of source similarity (F(1,49) = 1.65, p = .205, η_p^2 = .03), and no interaction between source accuracy and similarity (F(1,49) = 0.38, p = .541, η_p^2 = .01).

Likewise, the sources that were programmed to be similar on political questions agreed with the participants on approximately 80% of politics trials (Accurate/Similar: M = 82.05, SD = 5.95. Inaccurate/Similar: M = 78.60, SD = 7.32) while those programmed to be dissimilar agreed with participants on approximately 20% of these trials (Accurate/Dissimilar: M = 20.30, SD = 5.82. Inaccurate/Dissimilar: M = 20.20, SD = 6.35). Entering the percentage of political questions on which each source agreed with each participant into a 2 (source accuracy condition: accurate, inaccurate) x 2 (source similarity condition: similar, dissimilar) rmANOVA revealed a main effect of source similarity (F(1,49) = 4005.41, p < .001, η_p^2 = .99), no main effect of source accuracy (F(1,49) = 3.53, p = .066, η_p^2 = .07), and no interaction between source accuracy and similarity (F(1,49) = 3.73, p = .059, η_p^2 = .07).

Participants' Information-Seeking Choices Were Unaffected by Political Similarity

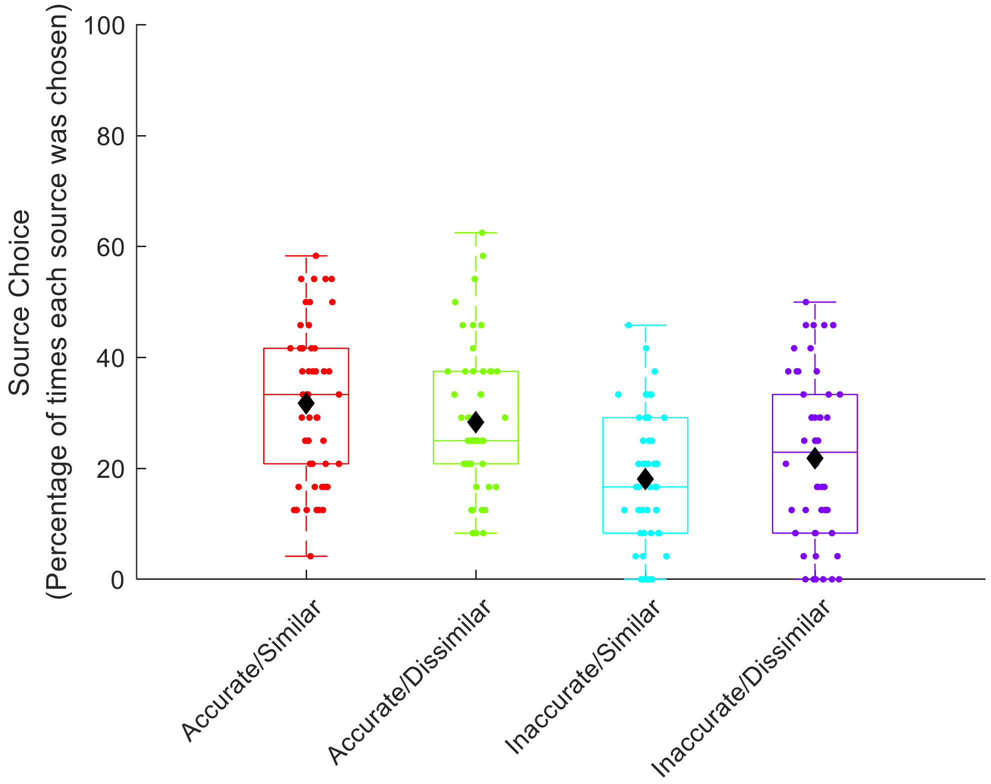
Entering participants' information-seeking decisions in the choice stage into a GLME, with the difference in accuracy and similarity between the two sources presented on each trial, along with their interaction, all entered as both fixed and random factors, revealed that the probability of choosing to hear from the source presented on the right-hand side increased with the (standardized) difference in source accuracy between the two sources presented on each trial (β = 0.78, SE = 0.17, 95% CIs = [0.45, 1.11], t(1196) = 4.61, p < .001). The (standardized) difference in source similarity between the two sources presented on each trial did not influence participants' information-seeking decisions (β = -0.05, SE = 0.12, 95% CIs = [-0.29, 0.19], t(1196) = -0.39, p = .697). The interaction between source accuracy and source similarity was not significant (β = 0.05, SE = 0.09, 95% CIs = [-0.13, 0.23], t(1196) = 0.56, p = .574) and neither was the intercept of the model (β = 0.06, SE = 0.08, 95% CIs = [-0.09, 0.22], t(1196) = 0.82, p = .411).

Controlling for participants' confidence in their initial answers did not alter the main results. When these confidence ratings were included in the model, the effect of the source accuracy difference was still significant (β = 0.84, SE = 0.18, 95% CIs = [0.48, 1.20], t(1192) = 4.58, p < .001) and the effect of the source similarity difference was not (β = -0.05, SE = 0.14, 95% CIs = [-0.32, 0.22], t(1192) = -0.39, p = .696).

To allow for a more direct comparison with the results presented in Chapter 2, we also entered the percentage of times each source was chosen by each participant into a 2 (source accuracy: accurate, inaccurate) by 2 (source similarity condition: similar, dissimilar) rmANOVA (Figure 19). This revealed a main effect of source accuracy (F(1,49) = 21.60, p < .001, η_p^2 = .31), no main effect of source similarity (F(1,49) < 0.01, p = .973, η_p^2 < .01), and no interaction between source accuracy and source similarity (F(1,49) = 2.57, p = .115, η_p^2 = .05).

Figure 19

Percentage of Trials on Which Participants Chose to Seek Information from Each Source



Note. Participants preferred to receive information on general knowledge questions from the more accurate sources, regardless of how similar their political views were to those of the participant. To facilitate interpretation, in this figure we have plotted the percentage of times each participant selected to hear from each source (coloured dots). The black diamonds represent the mean of these percentages. The box plots show the distribution of these percentages: boxes indicate 25–75% interquartile range, whiskers extend from the first and third quartiles to most extreme data point within 1.5 × interquartile range, and the median is shown as a horizontal line within this box.

Participants' Change of Mind Did Not Vary According to Whom They Received Information From

A linear mixed model revealed that COM did not vary according to the accuracy of the chosen source (β = 2.54, SE = 1.31, 95% CIs = [-0.03, 5.11], t(1192) = 1.94, p = .053) or the similarity of the chosen source (β = -0.09, SE = 1.09, 95% CIs = [-3.08, 1.21], t(1192) = -0.86, p = .391). The interaction between source accuracy and source similarity was not significant (β = -0.29, SE = 1.24, 95% CIs = [-2.72, 2.15], t(1192) = -0.23, p = .818). The intercept of the model was significant (β = 21.59, SE = 1.95, 95% CIs = [17.76, 25.41], t(1192) = 11.08, p < .001), indicating that participants were positively influenced by the sources' answers on quiz questions in the choice stage. COM was greater when the chosen source disagreed with the participant's initial answers than when they agreed (β = 8.40, SE = 1.47, 95% CIs = [5.51, 11.28], t(1192) = 5.72, p < .001). All the interactions included in the model were non-significant (all p-values > .20).

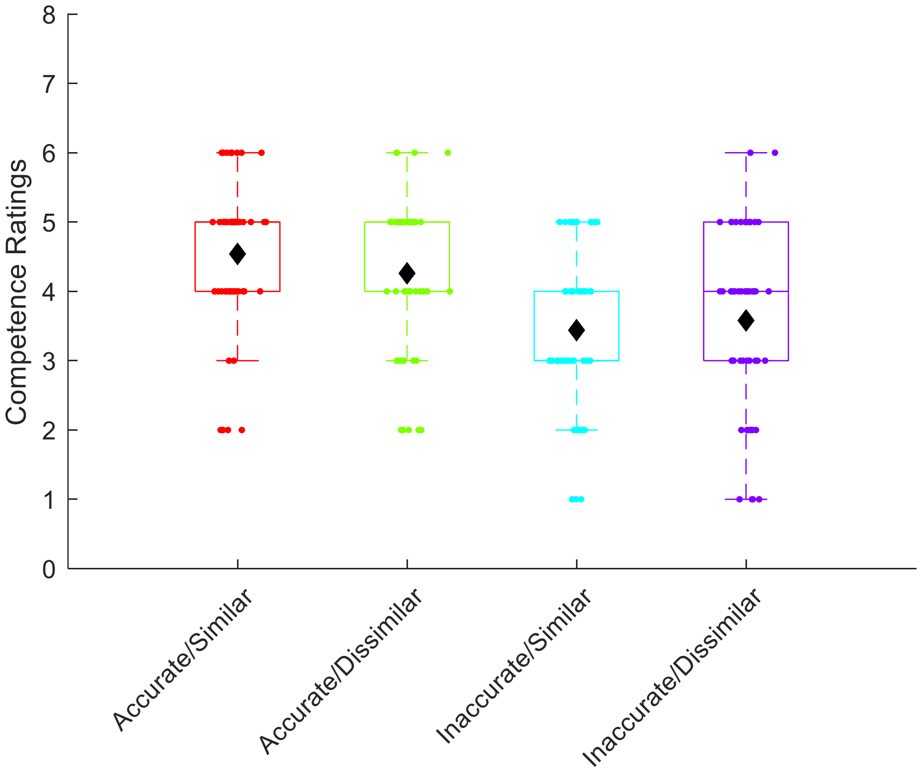
Similarity Learning Did Not Affect Expertise Learning

Competence Ratings. Entering participants' rating of each source's competence on the quiz trials into a linear mixed-effects model, with the source's objective accuracy (i.e., the percentage of times they answered general knowledge questions correctly), objective political similarity (i.e., the percentage of times they agreed with the participant's answer), and their interaction all entered as both fixed and random factors, revealed a significant effect of source accuracy on participants' perceptions of which sources were competent on the quiz trials (β = 0.37, SE = 0.06, 95% CIs = [0.25, 0.49], t(196) = 6.14, p < .001) and no effect of source similarity (β = 0.02, SE = 0.08, 95% CIs = [-0.14, 0.18], t(196) = 0.21, p = .836). The interaction between source accuracy and similarity was not significant (β = 0.06, SE = 0.06, 95% CIs = [-0.07, 0.19], t(196) = 0.89, p = .373).

A 2 (source accuracy condition: accurate, inaccurate) x 2 (source similarity condition: similar, dissimilar) rmANOVA on the competence ratings likewise revealed a significant main effect of source accuracy (F(1,49) = 35.84, p < .001, η_p^2 = .42), no main effect of source similarity (F(1,49) = 0.11, p = .737, η_p^2 < .01), and no interaction between source accuracy and similarity (F(1,49) = 1.88, p = .177, η_p^2 = .04) (Figure 20).

Figure 20

Participants' Ratings of Each Source's Competence



Note. Participants rated accurate sources as more competent at answering general knowledge questions, regardless of how politically similar they were to the participant. The coloured dots represent each participant's competence rating for each source. The black diamonds represent the mean of these ratings. The box plots show the distribution of the competence ratings: boxes indicate 25–75% interquartile range, whiskers extend from the first and third quartiles to most extreme data point within 1.5 × interquartile range, and the median is shown as a horizontal line within this box.

Betting Behaviour on Quiz Trials. A 2 (source accuracy condition: accurate, inaccurate) x 2 (source similarity condition: similar, dissimilar) rmANOVA revealed that participants bet that accurate sources would be correct (Accurate/Similar: M = 71.65, SD = 17.75. Accurate/Dissimilar: M = 71.20, SD = 15.83) on a greater percentage of quiz trials than inaccurate sources (Inaccurate/Similar: M = 56.00, SD = 16.97. Inaccurate/Dissimilar: M = 54.95, SD = 19.58; $F(1,49) = 33.99$, $p < .001$, $\eta_p^2 = .41$). Source similarity in the politics trials did not affect participants' betting behaviour in the quiz trials ($F(1,49) = 0.14$, $p = .711$, $\eta_p^2 < .01$). The interaction between source accuracy and source similarity was not significant ($F(1,49) = 0.03$, $p = .860$, $\eta_p^2 < .01$).

Computational Modelling. We next assessed whether participants learned more about source competence when observing congruent evidence (i.e., information suggesting that similar [dissimilar] sources were more [less] competent on quiz trials) than incongruent evidence (i.e., feedback suggesting that dissimilar [similar] sources were more [less] competent on quiz trials). We did this using the same model comparison procedure used in Study 3. The model statistics are presented below (Table 9).

Table 9

Expertise Learning Model Comparison Results

Model No.	BIC	AIC	Mean Pseudo R-Squared	% of participants for whom model fit best (BIC)	% of participants for whom model fit best (AIC)
RL models using data from quiz trials only (Set 1)					
1. RL Unbiased	**10255**	9947	0.12	74%	50%
2. RL Congruence Bias	10291	**9829**	0.14	26%	50%
Bayesian models using data from quiz trials only (Set 1)					
3. BB Unbiased	**10210**	9748	0.15	82%	50%
4. BB with 1 Scaling Parameter	10391	9776	0.15	6%	10%
5. BB Congruence Bias	10445	**9677**	0.17	12%	40%
RL models using data from quiz and political trials (Set 2)					
1. RL Unbiased	**20630**	20065	0.11	84%	52%
2. RL Congruence Bias	20769	**20015**	0.12	16%	48%
Bayesian models using data from quiz and political trials (Set 2)					
3. BB Unbiased	**20765**	19822	0.13	80%	52%
4. BB with 1 Scaling	20940	19809	0.13	12%	26%

Parameter					
5. BB Congruence Bias	21113	**19794**	0.14	8%	22%

Note: RL = Reinforcement-learning, BB = Beta-binomial.

The results of the model comparisons differed depending on which information criterion was used to assess the model fit. In every comparison, the BIC suggested that the unbiased model outperformed the congruence bias model, while the AIC indicated the reverse. However, both criteria suggest that the unbiased models fit best to a greater percentage of participants than the congruence bias models. Additionally, Wilcoxon signed rank tests revealed that the two learning rates [scaling parameters] in the RL [beta-binomial] congruence bias models did not differ significantly (Table 10), suggesting that there was not a systematic bias to learn more from congruent or incongruent evidence in the quiz trials. This is consistent with our analysis of participants' competence ratings and their average betting behaviour in these trials.

Table 10

Wilcoxon Signed Rank Tests Comparing the Magnitude of the Learning Rates [Scaling Parameters] Included in the Congruence Bias Models

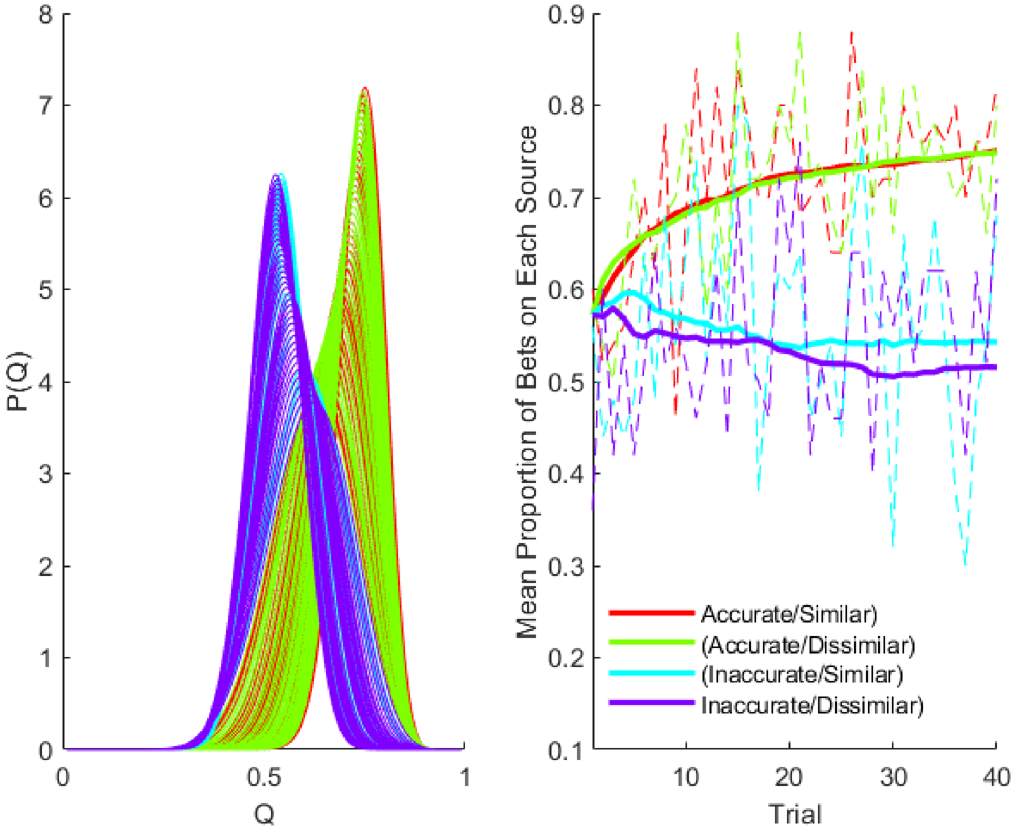
Congruence Bias Model	Median α_1 [γ_1]	Median α_2 [γ_1]	Z	p
RL, using data from quiz trials only (Set 1)	0.07	0.07	-1.02	.309
BB, using data from quiz trials only (Set 1)	0.44	0.34	0.95	.342
RL, using data from quiz and political trials (Set 2)	0.03	0.04	-0.95	.342
BB, using data from quiz and political trials (Set 2)	1.13	0.77	1.67	.094

Note: RL = Reinforcement-learning, BB = Beta-binomial. α_1 is the learning rate parameter applied to congruent feedback in the RL models; α_2 is the learning rate parameter applied to incongruent feedback in the RL models; γ_1 is the scaling parameter applied to congruent feedback in the BB models; γ_2 is the scaling parameter applied to incongruent feedback in the BB models.

Participants' betting behaviour on the quiz trials was reasonably well-described by a standard beta-binomial model (see Figure 21). Using the best-fit parameters from this model (Model 3), we assessed how accurately it predicted each participant's bets by simulating the bets of each participant on each trial 1000 times and computing the mean accuracy (i.e., the percentage of times the model's prediction matched the participant's behaviour) across these simulations. Overall, the model accurately predicted participants' bets on 59% of quiz trials. In particular, it predicted 63.48% of participant's bets on the Accurate/Similar source, 62.55% of bets on the Accurate/Dissimilar source, 53.79% on the Inaccurate/Similar source, and 54.28% on the Inaccurate/Dissimilar source.

Figure 21

A Standard Beta-Binomial Model Fit to Participants' Bets on Each Source in the Quiz Trials



Note. Left-hand side: The probability distributions illustrate how participants' beliefs about each source's expertise (Q) evolved over the course of the learning stage. Each distribution was calculated by averaging the model parameters for each trial across participants. Before observing any evidence pertaining to the sources' competence at answering quiz questions, the prior distribution did not vary by source. Participants' beliefs about each source were updated on each trial in light of the evidence they saw. Over the course of the learning stage, the model suggests that participants became less uncertain in their beliefs – as evidenced by the increasing height of the distributions – and learned which sources were more accurate. Right-hand side: Solid lines show the mean model-predicted probability of betting on each source on every quiz trial of the learning stage. Dotted lines show the proportion of participants that actually bet on each source on each quiz trial.

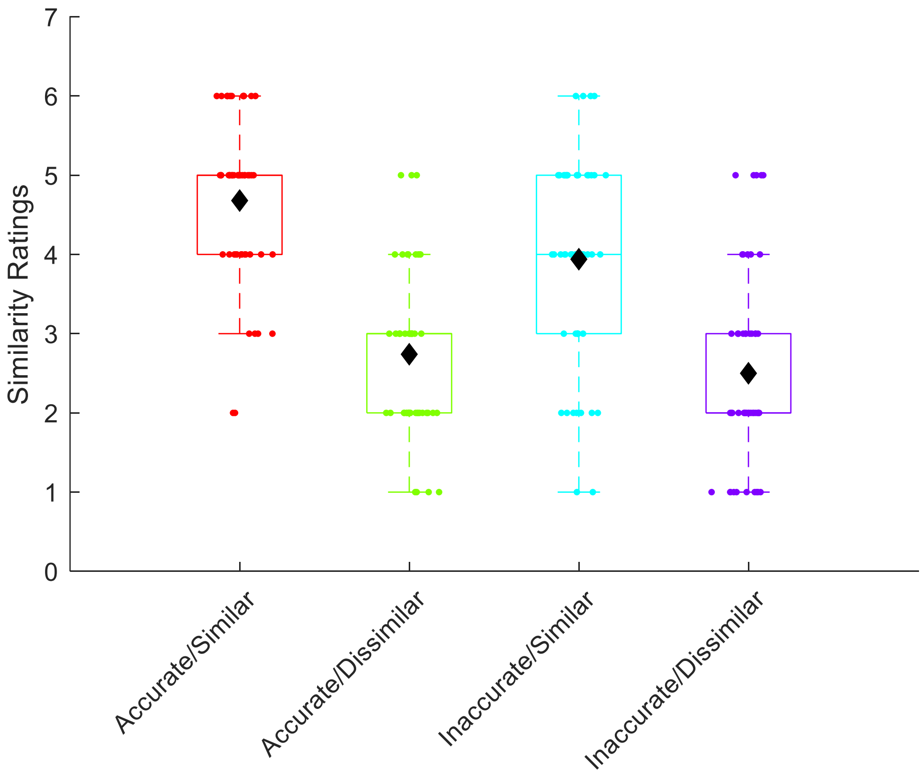
Expertise Learning Interfered with the Ability to Learn About Political Similarity

Similarity Ratings. We next tested whether learning about the sources' expertise on quiz trials interfered with participants' ability to learn about how politically similar the sources were to them. Specifically, we entered participants' similarity ratings into a linear mixed-effects model, with the source's objective similarity (i.e., the percentage of times they agreed with the participant's answer on political questions), objective accuracy (i.e., the percentage of times they answered general knowledge questions correctly), and the interaction between the two entered as fixed and random factors. As expected, this revealed a significant effect of source similarity on participants' perceptions of which sources were politically similar to them (β = 0.58, SE = 0.07, 95% CIs = [0.44, 0.72], t(196) = 8.06, $p < .001$). Surprisingly, there was also a positive effect of source accuracy (β = 0.18, SE = 0.05, 95% CIs = [0.08, 0.28], t(196) = 3.49, $p < .001$), indicating that participants believed that the sources that were more accurate on quiz trials were more similar to them politically. There was no interaction between source accuracy and similarity (β = 0.07, SE = 0.05, 95% CIs = [-0.03, 0.16], t(196) = 1.37, p = .173).

We observed the same pattern of results when performing a 2 (source similarity condition: similar, dissimilar) x 2 (source accuracy condition: accurate, inaccurate) rmANOVA instead of a linear mixed model on these data. The rmANOVA revealed a significant effect of political similarity (F(1,49) = 60.09, $p < .001$, η_p^2 = .55) and a significant effect of source accuracy on participants' similarity ratings (F(1,49) = 11.94, p = .001, η_p^2 = .20). There was no interaction between source similarity and accuracy (F(1,49) = 4.02, p = .051, η_p^2 = .08) (Figure 22).

Figure 22

Participants' Ratings of Each Source's Similarity



Note. Participants rated sources that tended to agree with them on political issues as more politically similar to them than those who tended to disagree with their opinions. They also rated sources that were good at answering general knowledge as more politically similar than those who performed worse on the quiz trials. The coloured dots represent each participant's rating of each source. The black diamonds represent the mean of these ratings. The box plots show the distribution of the generosity ratings: boxes indicate 25–75% interquartile range, whiskers extend from the first and third quartiles to most extreme data point within 1.5 × interquartile range, and the median is shown as a horizontal line within this box.

Betting Behaviour on Politics Trials. A 2 (source similarity condition: similar, dissimilar) x 2 (source accuracy condition: accurate, inaccurate) rmANOVA revealed that participants bet that the similar sources would agree with them on political questions (Accurate/Similar: M = 70.05, SD = 15.90. Inaccurate/Similar: M = 64.25, SD = 17.61) on a greater percentage of trials than dissimilar sources (Accurate/Dissimilar: M = 46.35, SD = 13.23. Inaccurate/Dissimilar: M = 43.15, SD = 15.07; F(1,49) = 49.44, p < .001, η_p^2 = .50). The sources' accuracy on the quiz trials also influenced participants' betting behaviour in the political trials (F(1,49) = 9.14, p = .004, η_p^2 = .16). The interaction between source similarity and source accuracy was not significant (F(1,49) = 0.98, p = .33, η_p^2 = .02).

Computational Modelling. We next fit our computational models to the betting data from the political trials. The results of the four model comparisons are displayed in Table 11. Here, the congruence bias models allowed learning to differ according to the expertise of the source and whether they agreed with the participant's answer on a political trial.

Table 11

Similarity Learning Model Comparison Results

Model number and name	**BIC**	**AIC**	**Mean Pseudo R-Squared**	**% of participants for whom model fit best (BIC)**	**% of participants for whom model fit best (AIC)**
RL models using data from political trials only (Set 1)					
1. RL Unbiased	**10453**	10146	0.10	88%	70%
2. RL Congruence Bias	10567	**10106**	0.12	12%	30%
Bayesian models using data from political trials only (Set 1)					
3. BB Unbiased	**10442**	**9981**	0.13	94%	70%
4. BB with 1 Scaling Parameter	10622	10007	0.13	4%	16%
5. BB	10793	10024	0.14	2%	14%

Congruence Bias					
RL models using data from political and quiz trials (Set 2)					
1. RL Unbiased	**20630**	**20065**	0.11	94%	74%
2. RL Congruence Bias	20839	20085	0.11	6%	26%
Bayesian models using data from political and quiz trials (Set 2)					
3. BB Unbiased	**20765**	19822	0.13	76%	50%
4. BB with 1 Scaling Parameter	20876	**19746**	0.14	22%	38%
5. BB Congruence Bias	21101	19782	0.14	2%	12%

Note: RL = Reinforcement-learning, BB = Beta-binomial.

The model comparisons suggest that a congruence bias did not affect how participants learned about source similarity. The model criterions (BIC and AIC) tended to be lower for unbiased models than congruence bias models, and the unbiased models also fit best to a greater percentage of participants than the congruence bias models. Our modelling procedure cannot therefore help to shed light on why participants rated sources that were more accurate on the quiz task as more politically like-minded.

Wilcoxon signed rank tests indicated that the learning rates [scaling parameters] were not significantly different from each other in three out of the four RL [beta-binomial] congruence bias models. However, in the Set 2 beta-binomial model, which was fit to data from the political and quiz trials, the scaling parameter for congruent evidence was significantly greater than the scaling parameter for incongruent evidence (Table 12). However, as noted above, this model provided a worse fit to the data than a standard unbiased beta-binomial model or an unbiased beta-binomial model with only one scaling parameter, depending on which criterion was used to assess model fit.

Table 12

Wilcoxon Signed Rank Tests Comparing the Magnitude of the Learning Rates [Scaling Parameters] Included in the Congruence Bias Models

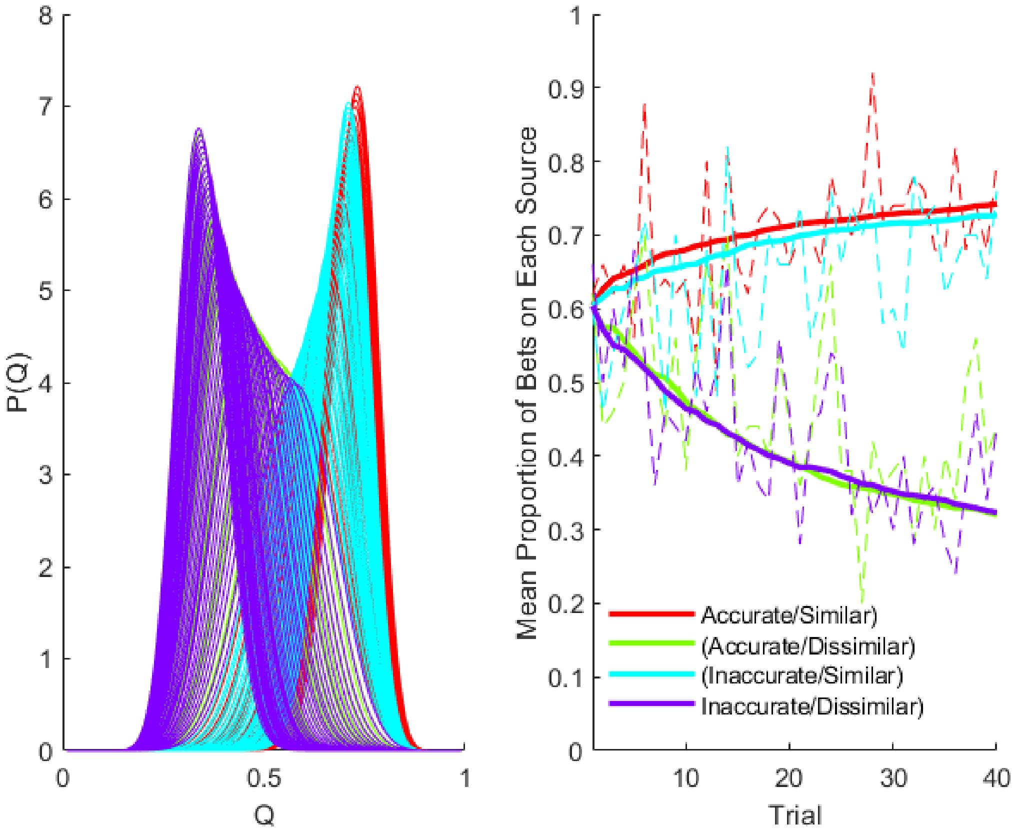
Congruence Bias Model	Median α_1 [γ_1]	Median α_2 [γ_1]	Z	p
RL, using data from political trials only (Set 1)	0.05	0.03	1.70	.088
BB, using data from political trials only (Set 1)	0.35	0.26	-0.19	.850
RL, using data from political and quiz trials (Set 2)	0.01	0.01	-0.81	.420
BB, using data from political and quiz trials (Set 2)	0.63	0.36	3.10	.002

Note: RL = Reinforcement-learning, BB = Beta-binomial. α_1 is the learning rate parameter applied to congruent feedback in the RL models; α_2 is the learning rate parameter applied to incongruent feedback in the RL models; γ_1 is the scaling parameter applied to congruent feedback in the BB models; γ_2 is the scaling parameter applied to incongruent feedback in the BB models.

Bets in the political trials were reasonably well-described by a standard beta-binomial model (Figure 23). As done previously, we simulated participants' bets on the political trials 1000 times using the standard beta-binomial model and computed the mean accuracy (i.e., the percentage of times the model's prediction matched the participant's behaviour) across these simulations. Overall, the model accurately predicted participants' bets on 58% of political trials. In particular, it predicted 62.39% of participant's bets on the Accurate/Similar source, 54.47% of bets on the Accurate/Dissimilar source, 59.75% on the Inaccurate/Similar source, and 55.04% on the Inaccurate/Dissimilar source.

Figure 23

A Standard Beta-Binomial Model Fit to Participants' Bets on Each Source in the Political Trials



Note. Left-hand side: The probability distributions illustrate how participants' beliefs about each source's similarity (Q) evolved over the course of the learning stage. Each distribution was calculated by averaging the model parameters for each trial across participants. Before observing any evidence pertaining to the sources' similarity, the prior distribution did not vary by source. Participants' beliefs about each source were updated on each trial in light of the evidence they saw. Over the course of the learning stage, the model suggests that participants became less uncertain in their beliefs – as evidenced by the increasing height of the distributions – and learned which sources were more and less similar to them. Right-hand side: Solid lines show the mean model-predicted probability of betting on each source on every political trial of the learning stage. Dotted lines show the proportion of participants that actually bet on each source on each political trial.

Discussion

This study was a conceptual replication of those reported in Chapter 2, however unlike the studies reported in Chapter 2, here we did not find an effect of political similarity on participants' expertise learning, information-seeking choices, or advice-utilisation. To help them to answer general knowledge questions, participants tended to seek information from sources that they had previously observed answering such questions with a high level of accuracy, regardless of how similar those sources were to them politically. Politically like-minded sources were not viewed as more competent on quiz trials than those who tended to disagree with the participant on political issues. However, we did find evidence to suggest that participants perceived sources that were more accurate on the quiz task as more similar to them politically. Applying computational models to the data did not help in providing an explanation for this phenomenon.

General Discussion

The two studies presented in this chapter failed to replicate our previous findings. Participants did not believe that politically like-minded (Study 4) or generous (Study 3) sources were better at answering general knowledge (quiz) questions. Rather, expertise learning was driven solely by relevant evidence. Accordingly, participants chose to hear more from sources with greater expertise when answering quiz questions themselves.

Clearly, a lot of changes were made in these studies relative to those conducted in Chapter 2 and it is difficult to disentangle which might have driven the changes in results. One possibility is that motivated reasoning drove the effects that we observed in Study 1 and 2, and participants in Study 3 and 4 lacked sufficient motivation to alter how they learned about others. The theory of motivated cognition suggests that a person's goals and needs bias their thinking towards desirable conclusions (Kunda, 1990; Taylor & Brown, 1988). Accordingly, the desire to maintain a positive self-concept may lead people to enhance their perceptions of others who share a valued group identity (for a review, see Van Bavel & Pereira, 2018). If group-based motivated cognition did indeed drive the effects reported in Chapter 2, it is easy to see why Study 3 would not have produced similar results; observing another acting generously will not activate a sense of shared group membership. The lack of effects in Study 4 are somewhat more puzzling, however it is quite possible that the manipulation used in this study also failed to activate a valued group identity. The US has suffered from particularly high levels of polarisation in recent years (Boxell et al., 2020; Draca & Schwarz, 2020; Iyengar et al., 2012; Iyengar et al., 2019) and our first two studies were run exclusively with American residents. Of course, the UK has also experienced rising polarisation (Boxell et al., 2020), in part as a result of the public's strong views on the issue of Brexit (Curtice, 2018; Hobolt et al., 2018). However, as many of the students who took part in Study 3 (33%) and 4 (40%) were not UK nationals, the average level of connection to British politics in these studies was likely relatively weak. If the political trials in Study 4 did not induce a relevant social identity or a strong sense of 'us' and 'them', then participants may not have been motivated to perceive sources that answered similarly to them on political questions favourably.

Another possibility is that participants paid more attention to the information they received in these studies compared to those in the previous chapter. There are several reasons for why this might have occurred. First, the incentive structure used in the present studies was more transparent; in our previous studies, participants were told that they could earn a bonus payment based on their performance but were not told exactly how performance would be measured. In contrast, here participants gained points for correct answers, lost points for incorrect answers, and received direct feedback about how well they were doing on each trial of the learning stage. Moreover, they were informed that their goal was to learn about others' characteristics, rather than how to categorise shapes, and the points they received were not only explicit and visible but also directly tied to the social aspects of this task.

Feedback interventions, which provide information regarding task performance, have a large effect on motivation, learning, and consequently subsequent performance (e.g., Ammons, 1956). Some have even gone so far as to claim that "the positive effect of feedback interventions on performance has become one of the most accepted principles in psychology" (Pritchard et al., 1988, p. 338). Feedback has a particularly strong effect on behaviour if it is salient (Dolan et al., 2012). It would therefore not be surprising if providing direct feedback on how well participants were learning about the sources' characteristics and how this related to their performance enhanced their ability to attend to relevant information during the learning stage.

There is also good reason to think that the mere act of asking participants to predict how the sources would answer questions might have led to an increased focus on relevant, and by extension a reduced focus on irrelevant, information. Recent evidence suggests that people are intrinsically motivated to see their predictions come true (Scherer et al., 2013). Scherer et al. (2013) found that asking participants to make a prediction about the aesthetic preferences of college students led participants to anticipate enjoyment from being right and selectively seek out new information consistent with that outcome. If the act of making a prediction induces a motivation to see a particular outcome, doing so may cause one to allocate more attention to feedback related to said outcome than would have been otherwise. The implication of this is that asking participants to bet on how sources will answer questions in this task may have led to greater incentivisation and thus enhanced attention and learning.

In a similar vein, not asking participants to answer questions themselves in the quiz trials of the learning stage may have freed up additional attentional resources, thereby allowing them to give greater attention to how accurate the sources were. In the learning stage of Study 1 and 2, participants were asked to guess whether each shape was a blap or not before seeing the source's answer. They were then presented feedback showing them the accuracy of their answer and the source's answer. Given that we told them that their task was to learn how to identify blaps, rather than to learn about how accurate the sources were on these questions, it is likely that attention was allocated away from the sources' outcomes so that they could focus more on the outcomes of their own predictions.

It is also possible that reducing the task length increased the attention paid to the social evidence presented in these studies. It is cognitively demanding to stay focused on a task for long periods of time, especially when the task is repetitive (Langner & Eickhoff, 2013; Pattyn et al., 2008; Warm et al., 2008). Consequently, participants may have had a greater tendency to engage in mind-wandering while completing our previous studies. This may have led participants to encode the information presented to them less efficiently and thus conflate political similarity with accuracy on the shape categorisation task.

Participants in the current studies may have also been more attentive because this study was conducted in person, with undergraduate psychology students, rather than online. Although there is some evidence to suggest that online participants are just as attentive as typical undergraduate subject populations (Hauser & Schwarz, 2016; Paolacci et al., 2010), online participants are also known to multitask and engage in distractions while completing experiments (Chandler et al., 2014; Clifford & Jerit, 2014). Indeed, in Study 1 and 2 we prevented approximately one third of the sample from completing the full experiment, as their answers to a learning test suggested that they had not learned adequately from the information that was presented to them during the learning stage. In contrast, the undergraduate students who took part in Study 3 and 4 were supervised and motivated to engage fully in the task, as they knew that they would subsequently need to write about the study as part of a graded assessment.

Lastly, it is possible that we accidentally sampled participants who selectively paid attention to political similarity in our previous experiments through our learning test. In the learning test, participants were asked to state which sources they thought as more similar to them. Those who answered less than 7 out of 8 of the relevant questions (where similarity varied between the two sources) did not complete the remainder of the task and were not included in the data analysis. Participants were not asked to complete an analogous accuracy learning test. Therefore, our sample may have included participants who learned about how politically similar but not how accurate the sources were. By asking exclusively about the similarity of the sources, participants may also have inferred that this political similarity was of particular importance. That is, demand characteristics may have resulted from the asymmetry of the learning test, leading participants to change their behaviour in the rating and choice stages of the experiment.

Interestingly, we did see one halo effect in Study 4; participants' betting behaviour in the political trials of the learning stage and subjective ratings of similarity revealed that they believed sources that were more accurate on the quiz task were also more politically like-minded. In our previous studies, we found that participants' perceptions of political similarity were determined by an interaction between source accuracy and similarity, whereby sources that were both accurate on blap trials and similar to them on political trials were rated as more similar than sources that were similar to them on political questions but less proficient at answering shape categorisation questions. Here we found a main effect of source accuracy, suggesting that the effect was not driven exclusively by participants' perceptions of the similar sources, however it is possible that the interaction would have arisen with greater experimental power.

The Salient Dimensions model of the halo effect (Fisicaro & Lance, 1990) posits that the direction of influence between two traits will depend on which of them is more salient. According to this model, salient traits draw attention and therefore influence perceptions of other, possibly unrelated traits. That expertise on quiz questions influenced perceptions of political similarity, but not vice-versa, suggests that the former trait was relatively more salient in Study 4. It is conceivable that the reverse was true in Study 1 and 2, due to participants allocating more attention to the shape categorisation questions than the accuracy of the sources. It is possible that quiz questions were more salient than blap questions because people find them inherently more interesting.

Fitting computational models to the data did not help us to delineate the cognitive processes underlying the effect of source accuracy on participants' perceptions of political similarity. We found that unbiased RL and Bayesian models provided a better fit to the data than models that incorporated a congruence bias, such that beliefs about a source on one trait would affect how much people learn about said source on a different trait. This suggests that the positive effects we did observe in Study 4 were not driven by the hypothesised mechanism. Our congruence bias models assumed that people overweight positive [negative] information about sources that possess [un]desirable characteristics and discount evidence to the contrary. It is possible that the generosity learning bias we observed is more specific than this. For example, people may only overweight positive information about sources that possess desirable characteristics or only discount negative information about sources with desirable characteristics. Alternatively, people may update their beliefs about different sources from the evidence they observe equally but apply a constant bonus to sources that possess desirable characteristics.

In conclusion, the present findings suggest that the effects observed in Chapter 2 are less robust than first thought. Not only did we find no evidence to support the notion that beliefs about others' generosity interfere with the ability to learn about expertise in unrelated domains, but we also failed to replicate the key findings from Study 1 and 2 when using a political similarity manipulation. As many changes were made to both the experimental paradigm and setting, it is difficult to disentangle which caused the effects to disappear. However, we suspect that increased attention to the information presented in these studies likely contributes to enhanced learning and diminished interference effects. Investigating these possibilities will be the focus of Chapter 4, in which we try to explain why the results we obtained in this chapter differed from those in the previous chapter, and in so doing increase our understanding of the moderators of epistemic spillovers.

Chapter 4

Chapter Overview

In Chapter 2, we used a novel experimental paradigm to demonstrate that learning about an irrelevant source characteristic – namely, political similarity – interferes with the ability to learn about and utilise others' task-relevant expertise. In Chapter 3, we reported two subsequent studies, in which the experimental design and setting were altered, that failed to corroborate these findings. Specifically, we found that learning about others' generosity (Study 3) and political beliefs (Study 4) did not bias assessments of expertise in a separate task. Consequently, participants chose to hear from sources that were accurate at the task, regardless of how generous or politically like-minded they were. The aim of this chapter was to reconcile these conflicting results. In particular, we ran two studies to test whether the act of predicting how others will answer questions enhances, and thereby de-biases, learning from subsequently observed evidence. In Study 5, we rolled back several of the changes we made in Study 3 and 4 and found that learning about others' similarity again influenced how people learned and utilised their expertise in an unrelated domain. Notably, here, we manipulated how similar the sources were to the participants on questions relating to personal values, rather than political opinions, and used an online sample of UK-based participants, thus demonstrating that epistemic spillover effects are not specific to US politics. In Study 6, we found that making predictions about the accuracy of others' information enhances expertise learning and strengthens the preference for receiving information from competent sources but does not attenuate the biasing effects of source similarity.

Introduction

The aims of this chapter were to (1) reconcile the conflicting results from the previous two chapters and (2) assess whether forming similarity beliefs in a domain outside of US politics can influence expertise learning.

The studies reported in Chapter 2, in which we found that learning about others' political views influenced expertise learning, information-seeking and advice-utilisation in an unrelated domain, were different to those in Chapter 3, in which unrelated source characteristics did not produce these same effects, in several ways. First, the former studies were conducted online with US participants, while the latter were conducted with a UK-based undergraduate student population in a laboratory setting. It is possible that the political similarity manipulation had a greater effect on the US than the UK participants due to the high levels of polarisation in the US (Boxell et al., 2020; Draca & Schwarz, 2020; Iyengar et al., 2012; Iyengar et al., 2019) and the fact that many of the UK participants were not British nationals, reducing the relevance of the anglo-centric political similarity manipulation. It is also possible that online participants were less attentive than those who completed the study in the lab as part of a course exercise, leading them to conflate desirable information on political questions with positive feedback on the blap task.

Second, the way the task was framed differed across these pairs of experiments. In Study 1 and 2, participants were told that “We are interested in how people understand rules” and that their job was “to learn through trial and error how to recognise a certain type of object, called a ‘blap’”, whereas in Study 3 and 4 participants were told that “We are interested in how people learn about other people’s generosity [political views] and general knowledge” and that their job was to “try to guess who gave money to charity and who didn’t [predict how the participants answered the political questions], as well as who answered the general knowledge questions correctly and who answered them incorrectly.” Thus, participants were asked to focus on the blap task in the former experiments and on others’ answers in the latter. Moreover, in the latter experiments we did not ask participants to answer blap questions themselves, thereby allowing them to focus solely on the sources’ responses.

Third, the incentive structure differed between the studies. Participants in the Chapter 2 studies were told that they could earn a bonus payment between $2.50 and $7.50 based on their performance but were not told exactly how performance would be measured. Those in the Chapter 3 studies were told that they would earn points for correct answers throughout the experiment, received feedback indicating how many points they earned through their predictions on how the sources would answer questions on each trial of the learning stage, and were informed that the individual who earned the most points in the study would be rewarded with a £40 cash bonus. The additional feedback in the later studies may have reinforced behaviour and enhanced social learning (Sutton & Barto, 1998), while the competitive nature of the incentive structure may have led to an increased motivation to perform well (Kapp, 2012).

Fourth, participants in the earlier studies learned about sources sequentially, observing answers from one source per block. In contrast, those in the subsequent studies were presented with the responses of all four sources on each trial of the learning stage, thereby allowing them to make relative comparisons between the sources on each trial. Previous research has demonstrated that people's judgements sometimes differ when they make joint evaluations, in which two options are presented and compared together, compared to when they evaluate options separately (e.g., Bazerman et al., 1992). This difference in evaluation mode may also have enhanced learning in our later studies. Moreover, presenting information about all four sources in each trial greatly reduced the task length, which may have reduced fatigue effects.

Fifth, the learning test that was used in Study 1 and 2 to exclude participants who had not learned about similarity was not included in Study 3 and 4. Selectively sampling participants in the former studies based on similarity but not accuracy learning may have biased our results, as too could the effect of focusing participants' attention on similarity immediately prior to the ratings and choice stages of the task.

Sixth, our earlier studies employed a novel shape categorisation task in which participants could not know the correct answer to each question, as there was no ground truth as to whether a shape was a 'blap' or not, while the latter studies used a quiz task in which there were objectively correct and incorrect answers. However, it seems unlikely that the content of the quiz questions evoked stereotypes (e.g., liberal sources tend to know more about art), given that we found no effects of generosity or political opinions on expertise learning in the studies using the quiz task.

In the studies presented in this chapter we investigated whether learning about others' *personal values* would interfere with the ability to learn about and utilise others' expertise in an unrelated domain, using a modified version of the blap task. Values are guiding principles that dictate how people try to live their lives (Aquino & Reed, 2002; Schwartz, 1996). They are central to people's self-identities, are formed early in life (Block et al., 2018; Croft et al., 2014), and have been shown to underlie people’s attitudes and behaviours in different domains, including politics (Boer & Fischer, 2013; Schwartz et al., 2010). For example, Schwartz and colleagues (2010) showed that people’s political attitudes (e.g., towards law and order) mediate the relationship between their more “basic” personal values (e.g., security values) and voting choices, suggesting that political attitudes are derived from higher order personal values.

In the same way that people segregate based on political ideology (Gentzkow & Shapiro, 2011), there is also evidence to suggest that people are drawn to individuals and organisations who possess values that match their own (Hogan et al., 1972; Schneider, 1987). Husbands and wives, too, not only exhibit concordance in their political views but also their personal values (Caspi & Herbener, 1993; Watson et al., 2004). It has been suggested that there are functional benefits to selectively interacting with those with similar values; value congruence is associated with reduced relationship and task-based conflict in working groups (Jehn et al., 1997). This is likely because conflicts over values tend to be harder to resolve than those over interests, as they are concerned with issues that relate to the negotiators’ core self-identities (Harinck & Van Kleef, 2012; Kouzakova et al., 2012; Tetlock et al., 2000; Wade-Benzoni et al., 2002). We thus reasoned that if the influence of political similarity on expertise learning, information-seeking choices and advice utilisation found in Study 1 and 2 was due to an affective halo effect, whereby people’s liking of politically like-minded sources led them to perceive them as more competent on the blap task, we would expect to observe the same pattern of results when people learn about others’ personal values. Replicating these results in a domain outside of US politics would also help to demonstrate the psychological importance and generality of these effects.

We also tested whether asking participants to make predictions (or “bets”) about how sources would answer questions relating to personal values and questions on the blap task would enhance social learning. Given that people find it intrinsically rewarding to make accurate predictions (Scherer et al., 2013) and learn more when actively interacting with a task than passively watching others (e.g., Kardas & O’Brien, 2018), we reasoned that asking participants to guess how sources would answer questions may lead to greater attention to relevant evidence in our task. Therefore, the inclusion of the betting procedure could explain the null results reported in Chapter 3. Here, we hypothesized that the act of predicting how others will respond would increase the amount of attention paid to their responses and thus allow participants to build more accurate (less biased) representations of others' characteristics (i.e., expertise on the shape task and value similarity).

Overview of Experiments

In the first study presented in this chapter (Study 5), we sought to conceptually replicate our initial findings in a context that did not invoke political allegiances, using a shorter task than was used in Study 1 and 2. For the reasons mentioned above, we chose to manipulate similarity in terms of personal values. Participants were not asked to bet on how the sources would answer questions in this experiment. In fact, with the exceptions of the stimuli used to manipulate source similarity, the amount of information presented to participants on each trial of the learning stage (responses from all four sources were observed on each trial), and the number of trials included, this experiment was comparable to those from Chapter 1. That is, participants were initially asked to complete the blap task (although we renamed "blaps" as "blups" – and we will refer to the 'blap task' as the 'blup task' from here – after realising that there was an alternative meaning of the former on Urban Dictionary) and were presented with feedback on their own answers as well as the answers of four sources; they were also asked questions about their personal values and subsequently told whether each of the four sources agreed with their answers on these questions but were not asked to bet on whether the sources agreed with them; they then completed a similarity learning test before going on to rate each source's competence on the blup task and similarity in terms of personal values; lastly, participants completed a choice stage, in which they were presented with a series of novel shapes, asked to indicate whether they thought each shape was a blup, presented with a pair of sources and asked whose answer they wanted to see, and finally given the opportunity to update their own answer.

In the second study of this chapter (Study 6), we directly tested whether betting affected expertise learning on the blup task and, consequently, participants' information seeking choices. A preregistered between-subjects design was employed; in one condition, participants completed the same experiment as those in Study 5, while in the other condition participants completed the same set of tasks except were additionally asked to bet on which sources would answer shape categorisation questions (in)correctly and which would (dis)agree with their answers on questions relating to personal values in the learning stage.

Study 5

Method

Participants

Participants were recruited through Prolific Academic (https://app.prolific.co/) and completed the study online on the survey platform Qualtrics (http://qualtrics.com/). Prolific's custom pre-screening function was used to recruit UK nationals who had an approval rate greater than 97% on previous Prolific studies and had previously stated that they were willing to take part in studies involving an element of deception.

The study description on Prolific stated that participants must complete the experiment on a laptop or PC on Chrome, Firefox, Safari, Microsoft Edge but not Internet Explorer. Participants were informed that those who did not comply with this instruction would be excluded and would not be eligible to receive a bonus payment. The reason for this was that the images on the feedback screens do not show up in Internet Explorer (or legacy browser versions) and are not easily viewable on a phone or tablet screen. No participants were excluded on this basis.

Participants were paid a base rate of £3.75, as the experimental session was expected to last approximately 30 minutes (hourly rate of £7.50 per hour). They were also able to earn a bonus of up to £1, based on their performance. The bonus was determined according to how many attention checks and learning test questions the participant answered correctly, however participants were not told specifically how the bonus would be calculated. Ethical approval was granted from UCL (SHaPS_2015_AH_017).

The sample size was determined according to economic constraints. Funding was provided by the Division of Psychology and Language Sciences, which allowed for a sample size of n = 50. Fortunately, a power analysis using data from Study 1 revealed that a sample size of 50 participants would achieve 95% power to detect the smallest effect size of interest (the effect of source similarity on advice utilisation, $\eta_p^2 = .09$) with an alpha of .05, assuming a correlation among repeated measures of 0.269 (the observed correlation among repeated measures in Study 1). The sample consisted of 12 males and 38 females (mean age = 32.32 years, SD = 12.87).

The data collection was performed in two batches. The first (pilot) batch consisted of five participants, while the second (main) batch consisted of the remaining 45 participants. This was done to allow us to check for technical issues and gauge how well participants were performing on the learning test on a small pilot sample. After reviewing the data from the pilot sample, we made a small change to the instructions; we explicitly told participants that their performance on the similarity learning test would be taken into account when determining their bonus payment. This was because only one out of the five participants included in the pilot would have passed the learning test threshold used in Chapter 2 (i.e., scored > 6). All 50 participants were included in the analyses, however we also tested whether the results held after removing those who failed the learning test.

Study Design

Learning Stage. The learning stage consisted of 4 blocks of 20 trials each (10 shape categorisation ('blup') trials and 10 personal value ('value') trials). The trial types were interleaved, with a value trial always following a blup trial. The order in which trials within each block were presented was randomised using Qualtrics' loop and merge function. In between each block, participants were asked an attention check question (for more details, see below).

Before starting the learning stage, participants completed four practice trials: two blup trials and two value trials. In one practice blup trial all four sources answered the shape categorisation question correctly, while in the other all four sources answered incorrectly. Likewise, in one practice value trial all four sources agreed with the participant, while in the other all of the sources disagreed with them. Participants were then asked three attention check questions (e.g., "How many participants from our previous study will you learn about during this task?") to make sure they had understood the instructions. If a participant answered one of these attention check questions incorrectly they were told that they had given the wrong answer and asked to answer again until they answered correctly. They were also asked to confirm that they would not make notes during this task to aid their memory.

Blup Trials (Figure 24a). On each trial, one of 64 colored shapes was presented on screen along with the question 'Is this a blup?' and the options 'Yes' and 'No' (self-paced). After giving an answer to this question, participants observed a feedback screen for 5s, which displayed whether they answered the question correctly or incorrectly and whether each of the four sources answered the question correctly or incorrectly. The feedback was manipulated so that participants were correct with 50% probability, two ('accurate') sources answered correctly with 80% probability, and two ('inaccurate') sources answered correctly with 50% probability on each trial. The feedback regarding the sources' accuracy was presented on screen in a 2x2 matrix.

Value Trials (Figure 24b). On each value trial, participants indicated whether they agreed or disagreed with one of 40 personal values questions (self-paced), which were adapted from questions used in the Hogan Motives, Values, and Preferences Inventory (Hogan & Hogan, 1996). Participants then observed a feedback screen for 5s, which displayed their answer and whether each of the four sources agreed or disagreed with their answer. The feedback was manipulated so that two ('similar') sources (one that was accurate and one that was inaccurate on blup trials) agreed with the participant's answer with 80% probability and two ('dissimilar') sources (one that was accurate and one that was inaccurate on blup trials) agreed with the participant's answer with 20% probability on each value trial. The feedback regarding the sources' agreement was presented on screen in a 2x2 matrix.

***Sources*.** The same pictures used in the previous studies were used again to represent the sources. The source condition assigned to each animal picture (i.e., Accurate/Similar, Accurate/Dissimilar, Inaccurate/Similar, Inaccurate/Dissimilar) was randomized using the block randomisation feature in Qualtrics. The order that the animal pictures were presented on the feedback screens remained constant throughout the learning stage. Therefore, the bird always appeared in the top left corner, for example, but whether the bird was similar or dissimilar in the value trials, and accurate or inaccurate on blup trials, was randomised across participants.

Attention Checks. At the end of each block, participants were presented with an attention check in which they were asked one of the following questions regarding the previous trial: "Was the question you just answered about a shape or about a personal value?"; "Did the previous participant represented by the fish icon agree or disagree with you on the previous question?"; "Which of these animals is not 1 of the 4 being used to represent a previous participant?" All participants answered at least two out of three of these questions correctly. On each of the three questions, 94%, 70%, and 100% of participants were correct, respectively.

Learning Test. As in Chapter 1, we assessed whether participants successfully learned which sources answered more similarly to them on the value trials using two-alternative forced choice tests. After the learning stage participants were presented with 12 learning test trials. On each trial two sources were presented and the participant was asked to indicate who they thought had more similar personal values to them ("Who do you think has personal values which are more similar to yours?"). Each possible pair of sources (six combinations) was presented twice. The order in which the source pairs were presented was randomised using Qualtrics' loop and merge function. We did not ask participants who they thought was more accurate on blup questions, as we did not test accuracy learning in this way in Study 1 or 2. In contrast to the studies reported in Chapter 2, participants were not excluded from completing the experiment based on their answers in the learning test.

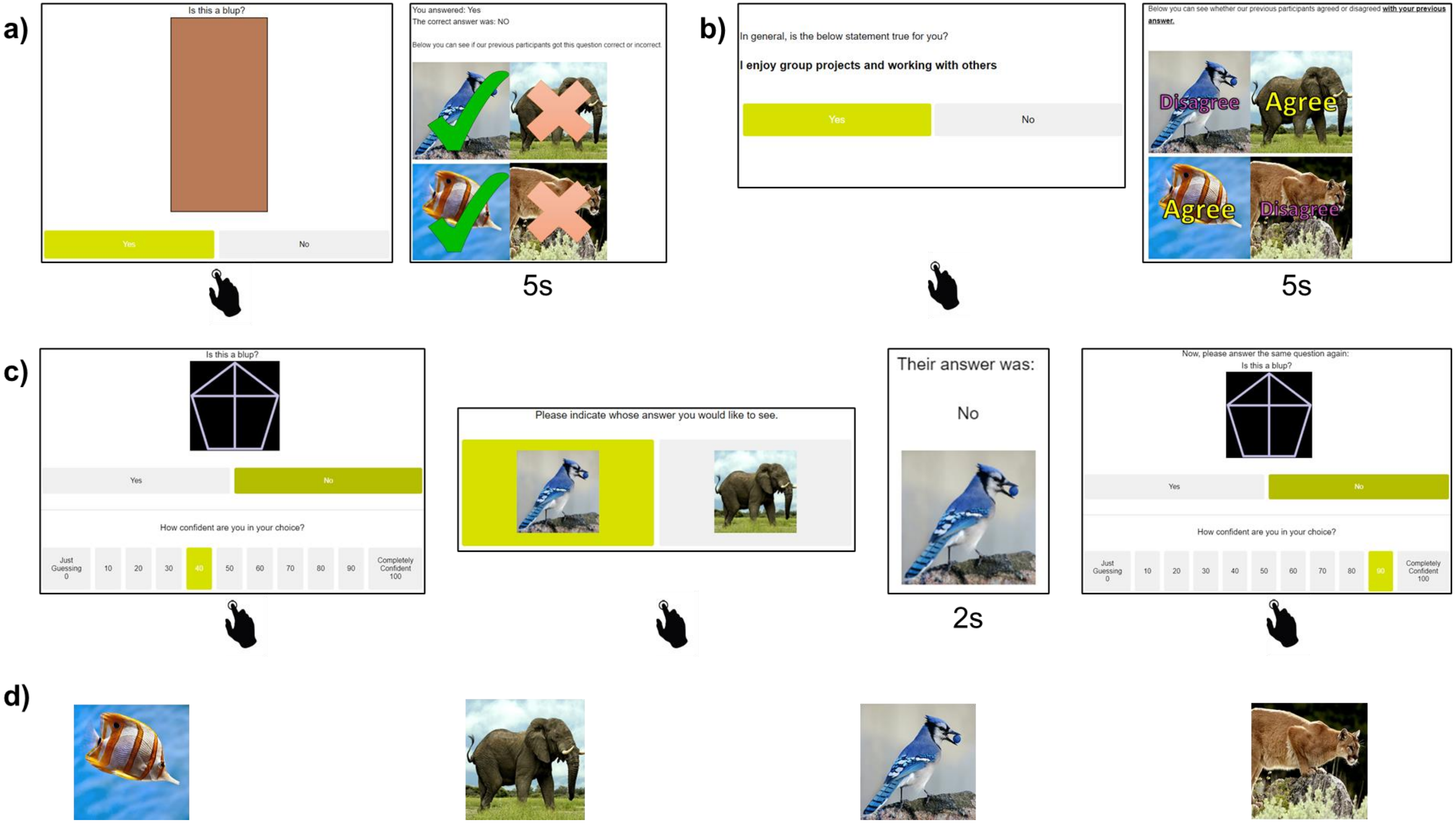
Ratings Stage. Participants then rated each source on: (1) how competent they were at determining if each object was a blup ("What percentage of blup questions did they answer correctly?" from 0 to 100 percent) and (2) how similar the source was to them in terms of personal values ("On what percentage of questions relating to personal values did they agree with you?" from 0 to 100 percent). The order in which each source was presented in the ratings stage was randomised, using the block randomisation feature in Qualtrics.

Choice stage (Figure 24c). On each of 24 trials, participants were presented with a novel shape and asked to indicate whether they thought the shape was a blup ("yes" or "no"; self-paced). They subsequently rated their confidence in this decision (self-paced) on a scale from 0 (Just Guessing) to 100 (Completely Confident). They were then presented with a pair of sources (pseudo-randomized, so that each possible pairing of sources was presented four times) and asked whose response they wanted to see (self-paced). They were then shown the response of the chosen source. Thereafter, the shape was presented again and participants were asked again to indicate whether they thought the shape was a blup ("yes" or "no") (self-paced). Lastly, participants rated their confidence (self-paced) in their final decision.

Within this block, there were an additional three blup questions in which participants did not get to see a source's answer to the question and were not able to update their initial answer. Participants were told this would be the case before starting the choice stage. This was done to motivate participants to provide accurate answers on the initial blup questions.

Figure 24

Experimental Design of the Task Used in Study 5



Is this a blup?

You answered: Yes
The correct answer was: NO

Below you can see if our previous participants got this question correct or incorrect.

In general, is the below statement true for you?

I enjoy group projects and working with others

Below you can see whether our previous participants agreed or disagreed **with your previous answer.**

Is this a blup?

How confident are you in your choice?

Please indicate whose answer you would like to see.

Their answer was:

No

Now, please answer the same question again:
Is this a blup?

How confident are you in your choice?

Note. During the Learning Stage participants learned about four sources' personal values and accuracy on the blup task. (a) Blup trials and (b) value trials were interleaved. (a) On each blup trial a novel shape was presented and the participants had to indicate whether they believed the shape was a blup (yes or no). They then saw whether each of the four sources answered the same question correctly or incorrectly, as well as whether their own answer was correct or incorrect. (b) On value trials a question assessing personal values was presented and the participants had to indicate whether they agreed with it (yes or no). They then saw whether each of the four sources agreed or disagreed with their answer on this question. (c) During the Choice Stage participants completed blup trials only. On each blup trial a novel shape was presented and the participant had to indicate whether they believed the shape was a blup (yes or no) and enter a confidence rating. They were then presented with two sources and asked to choose whose answer they would like to see. They then saw the response of the chosen source. Finally, they were given a chance to update their initial answer and confidence rating. Responses were self-paced. (d) There were four sources represented with animal photos which the participants were led to believe were other participants but were in fact algorithms.

Results

Manipulation Checks

As the percentage of trials on which the sources gave correct answers in the blup trials of the learning stage was not hard-coded, but rather based on the probability of each source answering each question correctly, we first examined whether the algorithm used to manipulate accuracy on the blup trials produced the desired pattern of responses. We did this by entering the percentage of accurate answers given by each source into a 2 (source accuracy condition: accurate, inaccurate) x 2 (source similarity condition: similar, dissimilar) repeated measures (rm) ANOVA. This revealed that the sources that were programmed to be accurate on blup trials answered questions correctly (Accurate/Similar: M = 80.40, SD = 5.74.

Accurate/Dissimilar: M = 80.20, SD = 5.58) more often than those programmed to be inaccurate (Inaccurate/Similar: M = 48.65, SD = 8.30. Inaccurate/Dissimilar: M = 47.85, SD = 6.70; main effect of source accuracy: F(1,49) = 967.06, $p < .001$, η_p^2 = .95). Source accuracy on the blup trials did not vary as a function of source similarity on the value trials (F(1,49) = 0.06, p = .579, η_p^2 = .01), and the interaction effect was not significant (F(1,49) = 0.12, p = .732, $\eta_p^2 < .01$).

We also examined whether the algorithm used to manipulate value similarity achieved its intended aim by entering the percentage of value trials on which each source agreed with each participant's answers into a 2 (source accuracy condition: accurate, inaccurate) x 2 (source similarity condition: similar, dissimilar) rmANOVA. This revealed that sources that were programmed to be similar agreed with participants' answers on value trials (Accurate/Similar: M = 79.60, SD = 6.61. Inaccurate/Similar: M = 79.70, SD = 5.48) more often than those programmed to be dissimilar (Accurate/Dissimilar: M = 20.40, SD = 5.56. Inaccurate/Dissimilar: M = 21.30, SD = 6.23; main effect of source similarity: F(1,49) = 4928.78, $p < .001$, η_p^2 = .99). Source similarity on the value trials did not vary as a function of source accuracy on the blup trials (F(1,49) = 0.28, p = .598, η_p^2 = .01), and the interaction effect was not significant (F(1,49) = 0.30, p = .589, η_p^2 = .01).

Participants Preferred to Receive Information About Shapes from Like-Minded Sources

As in Chapter 3, we used a generalised linear mixed-effects model (GLME) to analyse participants' information-seeking choices. Source choice was entered as the dependent variable and coded as 0 if the participant chose the source presented on the left and as 1 if the participant chose the source presented on the right. The difference in source accuracy and the difference in source similarity between the source presented on the right and left on each trial were included as both fixed and random factors. The interaction between the source accuracy difference and the source similarity difference was included as a fixed and as a random factor. These predictor variables were standardized (z-scored) before being entered into the GLME. Subject ID was entered as a random factor (grouping variable).

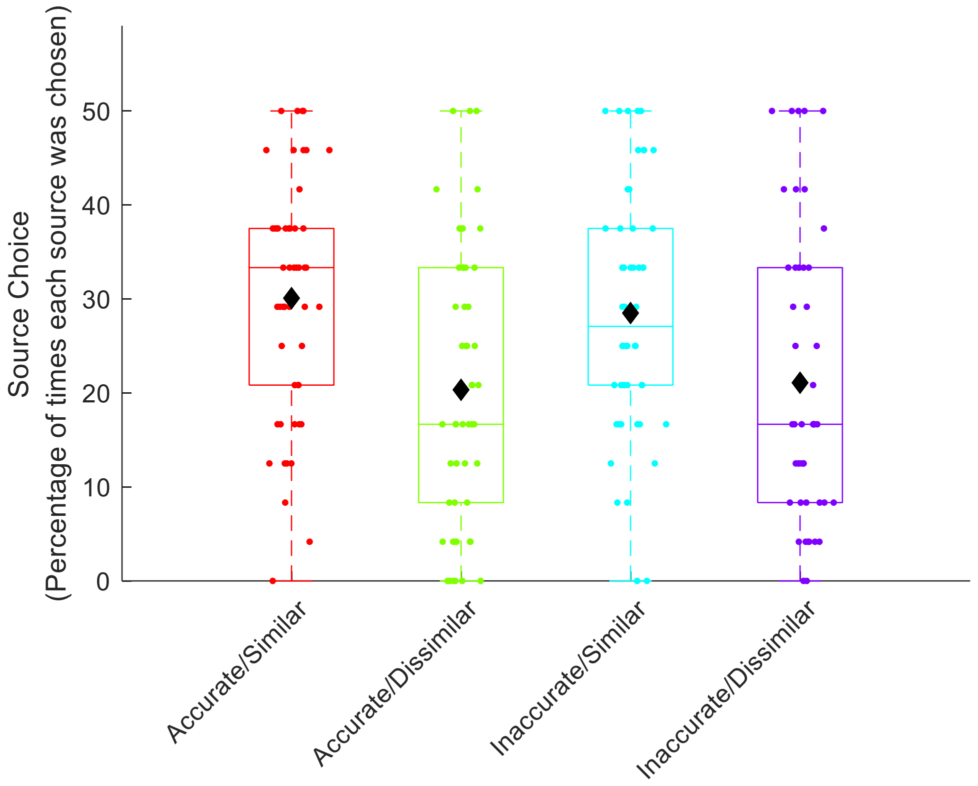
The GLME revealed that the probability of choosing to hear from the source presented on the right-hand side increased with the (standardized) difference in source similarity (β = 0.64, SE = 0.18, 95% CIs = [0.28, 1.00], t(1196) = 3.50, $p < .001$), indicating that participants preferred to seek information on shape categorisation questions in the choice stage from sources that were more similar to them on value trials in the learning stage (Figure 25). Surprisingly, the (standardized) difference in source accuracy did not affect participants' information-seeking decisions (β = -0.05, SE = 0.16, 95% CIs = [-0.38, 0.27], t(1196) = -0.33, p = .740). The interaction between the source accuracy difference and source generosity difference was not significant (β = 0.04, SE = 0.08, 95% CIs = [-0.13, 0.20], t(1196) = 0.43, p = .665).

For ease of interpretation and to facilitate comparison with the studies in Chapter 1, we also calculated the percentage of times each participant chose to hear from each source and entered these values into a 2 (source accuracy condition: accurate, inaccurate) x 2 (source generosity condition: generous, selfish) rmANOVA. This produced the same pattern of results as the GLME: the rmANOVA revealed a main effect of source similarity (F(1,49) = 11.72, p = .001, η_p^2 = .19), no main effect of source accuracy (F(1,49) = 0.04, p = .848, η_p^2 = .01), and no interaction between source similarity and source accuracy (F(1,49) = 0.30, p = .589, η_p^2 = .01).

We also re-ran both of these analyses excluding participants who failed the learning test (n = 22), according to the criterion used in Study 1 and 2. Removing these participants did not change the results (n included = 28). In the GLME, the difference in source similarity had a significant effect on participants' information-seeking choices (β = 1.09, SE = 0.31, 95% CIs = [0.48, 1.71], t(668) = 3.50, $p < .001$), while the source accuracy difference (β = 0.19, SE = 0.18, 95% CIs = [-0.17, 0.55], t(668) = 1.02, p = .308) and the interaction between the two (β = 0.11, SE = 0.13, 95% CIs = [-0.14, 0.36], t(668) = 0.84, p = .399) did not. Likewise, in the rmANOVA, there was a main effect of source similarity (F(1,27) = 14.83, p = .001, η_p^2 = .36), no main effect of source accuracy (F(1,27) = 1.89, p = .180, η_p^2 = .07), and no interaction between source similarity and source accuracy (F(1,27) = 2.54, p = .123, η_p^2 = .09).

Figure 25

Percentage of Trials on Which Participants Chose to Seek Information from Each Source



Note. Participants preferred to receive information on shape categorisation questions from sources that held similar values to them. On each trial of the choice stage, participants were presented with two sources. The (standardised) difference in similarity on value questions between them influenced the probability of the right-hand source being chosen for information, whereas the (standardised) difference in accuracy on the task did not. To facilitate interpretation, in this figure we have plotted the percentage of times each participant selected to hear from each source (coloured dots). The black diamonds represent the mean of these percentages. The box plots show the distribution of these percentages: boxes indicate 25–75% interquartile range, whiskers extend from the first and third quartiles to most extreme data point within 1.5 × interquartile range, and the median is shown as a horizontal line within this box.

Participants' Change of Mind Did Not Vary According to Whom They Received Information From

As in Chapters 2 and 3, we computed participants' Change of Mind (COM) after observing the source's answer on each trial of the choice stage. A linear mixed model was used to assess whether the chosen sources' accuracy on blup trials and similarity on value trials influenced COM. COM was entered as the dependent variable; source accuracy (i.e., the percentage of times the chosen source answered blup questions correctly in the learning stage, z-scored), source similarity (i.e., the percentage of times the chosen source agreed with the participant's answer on value trials, z-scored), a variable indicating whether the source agreed or disagreed with the participant's answer on each trial, and their interactions were all entered as both fixed and random factors; and Subject ID was entered as a random (grouping) factor.

The linear mixed model revealed that COM did not vary depending on the accuracy of the chosen source (β = 0.22, SE = 1.12, 95% CIs = [-1.97, 2.41], t(1192) = 0.20, p = .845) or the similarity of the chosen source (β = 0.56, SE = 1.33, 95% CIs = [-2.05, 3.17], t(1192) = 0.42, p = .676). The intercept of the model was significant (β = 27.02, SE = 2.08, 95% CIs = [22.93, 31.10], t(1192) = 12.98, p < .001), indicating that participants were positively influenced by the sources' answers in the choice stage. COM was greater when the chosen source disagreed with the participant's initial answers than when they agreed (β = 18.50, SE = 1.85, 95% CIs = [14.88, 22.13], t(1192) = 10.01, p < .001). All the interactions included in the model were non-significant (all p-values > .15).

We re-ran this analysis after excluding participants who failed the learning test, according to the criterion used in Chapter 1, and found the results remained unchanged. In particular, the effect of source accuracy was not significant (β = 0.26, SE = 1.47, 95% CIs = [-2.63, 3.16], t(664) = 0.18, p = .858) and nor was the effect of source similarity (β = 2.54, SE = 1.40, 95% CIs = [-0.19, 5.29], t(664) = 1.82, p = .069).

Participants Perceived Sources That Shared Their Values as More Competent on the Blup Task

As in Chapter 3, we entered participants' competence ratings as the dependent variable into a linear mixed-effects model. Source accuracy (i.e., the percentage of times the source answered blup questions correctly) and source similarity (i.e., the percentage of times the source agreed with the participant's answers on value trials) were entered as both fixed and random factors. The interaction between source accuracy and source similarity was included as a fixed and as a random factor. Subject ID was entered as a random factor (grouping variable). This revealed an effect of source accuracy, indicating that participants learned which sources were more competent on the blup trials ($\beta = 0.16$, SE = 0.06, 95% CIs = [0.03, 0.28], t(196) = 2.42, p = .016). There was also an effect of source similarity, indicating that they also perceived sources that tended to agree with their answers on value questions as more competent on the blup task ($\beta = 0.24$, SE = 0.06, 95% CIs = [0.12, 0.37], t(196) = 3.87, p < .001). The interaction between source accuracy and similarity was not significant ($\beta = 0.01$, SE = 0.05, 95% CIs = [-0.08, 0.11], t(196) = 0.29, p = .771).

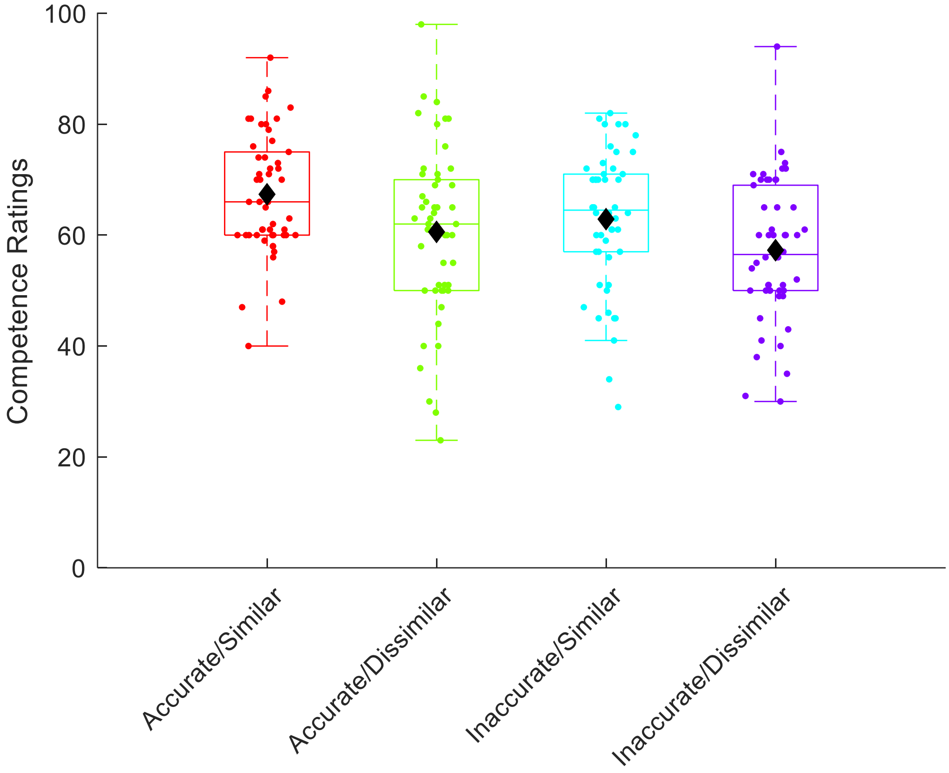
For ease of interpretation and to facilitate comparison with the studies in Chapter 1, we also entered the competence ratings into a 2 (source similarity condition: similar, dissimilar) rmANOVA. The rmANOVA corroborated the results of linear mixed-effects model (Figure 26): we observed a significant main effect of source accuracy (F(1,49) = 5.35, p = .025, η_p^2 = .098), a main effect of source similarity (F(1,49) = 13.65, p = .001, η_p^2 = .22), and no interaction between source accuracy and similarity (F(1,49) = 0.25, p = .619, η_p^2 = .01).

We re-ran the above analyses excluding participants who failed the learning test, according to the criterion used in Study 1 and 2. Here, we found that when we removed participants who failed the learning test, the effect of source accuracy was no longer significant in the linear mixed model ($\beta = 0.16$, SE = 0.10, 95% CIs = [-0.03, 0.35], t(108) = 1.68, p = .095) or the rmANOVA (F(1,27) = 2.89, p = .100, η_p^2 = .097). The effect of source similarity on participants' competence ratings was still significant in both the mixed model ($\beta = 0.29$, SE = 0.09, 95% CIs = [0.12, 0.46], t(108) = 3.35, p < .001) and the rmANOVA (F(1,27) = 9.94, p = .004, η_p^2 = .27). The interaction was not significant in the mixed model ($\beta = 0.03$, SE = 0.06, 95% CIs = [-

0.08, 0.15], t(108) = 0.59, p = .556) nor in the rmANOVA (F(1,27) = 0.85, p = .366, η_p^2 = .03).

Figure 26

Participants' Ratings of Each Source's Competence



Note. Participants rated the more accurate and similar sources as more competent on the blup task. The coloured dots represent each participant's competence rating for each source. The black diamonds represent the mean of these ratings. The box plots show the distribution of the competence ratings: boxes indicate 25–75% interquartile range, whiskers extend from the first and third quartiles to most extreme data point within 1.5 × interquartile range, and the median is shown as a horizontal line within this box.

Expertise Learning Did Not Affect Similarity Learning

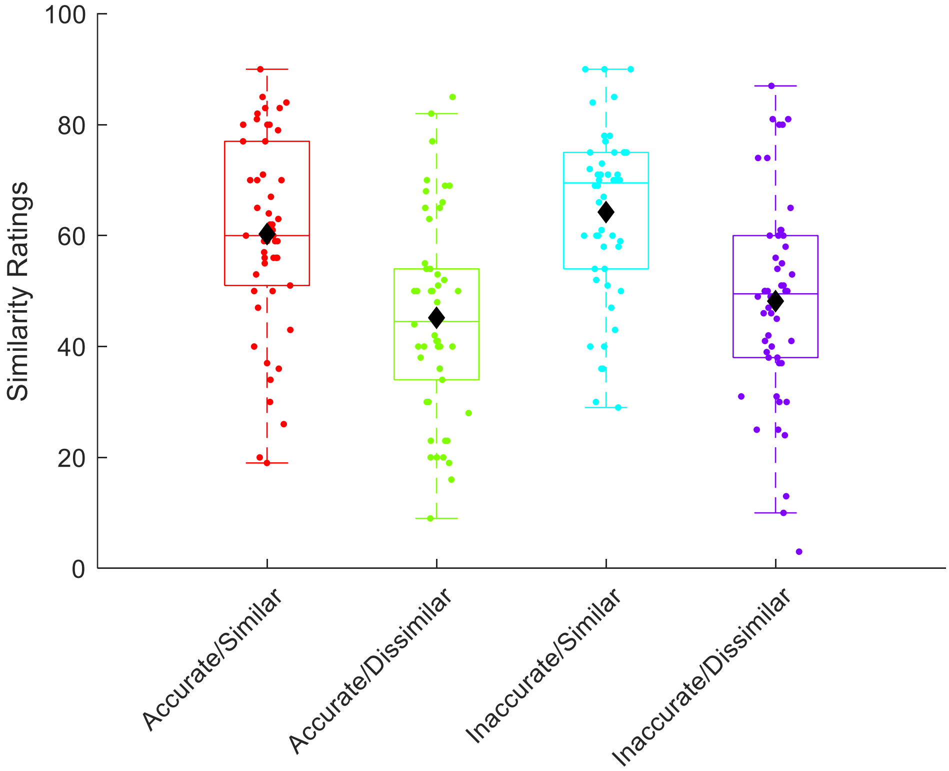
We also entered the similarity ratings as the dependent variable into a linear mixed-effects model. Source accuracy (i.e., the percentage of times the source answered blup questions correctly), source similarity (i.e., the percentage of times the source agreed with the participant's answers on value trials), and their interactions were entered as fixed and random factors, and Subject ID was entered as a random factor (grouping variable). This revealed a significant effect of source similarity on participants' perceptions of which sources agreed with them more on value questions (β = 0.43, SE = 0.08, 95% CIs = [0.27, 0.59], t(196) = 5.37, $p < .001$). The effect of source accuracy was not significant (β = -0.09, SE = 0.05, 95% CIs = [-0.19, 0.02], t(196) = -1.60, p = .11) and there was no interaction between source accuracy and similarity (β = 0.01, SE = 0.05, 95% CIs = [-0.10, 0.11], t(196) = 0.09, p = .926).

As above, we also entered the similarity ratings into a 2 (source similarity condition: similar, dissimilar) rmANOVA. Consistent with the linear mixed model, this revealed a significant main effect of similarity (F(1,49) = 26.06, $p < .001$, $\eta_p^2 = .35$), no main effect of source accuracy on participants' similarity ratings (F(1,49) = 3.43, p = .070, $\eta_p^2 = .07$), and no interaction between source similarity and accuracy (F(1,49) = 0.06, p = .802, $\eta_p^2 < .01$) (Figure 27).

There was also no change in the results when those who failed the learning test were excluded from these analyses. The effect of source similarity was significant in both the GLME (β = 0.80, SE = 0.07, 95% CIs = [0.66, 0.94], t(108) = 11.58, $p < .001$) and the rmANOVA (F(1,27) = 141.18, $p < .001$, $\eta_p^2 = .84$); there was no effect of source accuracy in the GLME (β = -0.03, SE = 0.09, 95% CIs = [-0.14, 0.10], t(108) = -0.29, p = .769) or the rmANOVA (F(1,27) = 0.39, p = .537, $\eta_p^2 = .01$); and there was no interaction effect in either model (GLME: β = 0.05, SE = 0.08, 95% CIs = [-0.10, 0.20], t(108) = 0.64, p = .518. rmANOVA: F(1,27) = 0.33, p = .570, $\eta_p^2 = .01$).

Figure 27

Participants' Ratings of Each Source's Similarity



Note. Participants rated sources that tended to agree with them on value trials as more similar to them, regardless of how accurate the source was on the blup task. The coloured dots represent each participant's rating of each source. The black diamonds represent the mean of these ratings. The box plots show the distribution of the generosity ratings: boxes indicate 25–75% interquartile range, whiskers extend from the first and third quartiles to most extreme data point within 1.5 × interquartile range, and the median is shown as a horizontal line within this box.

Discussion

In Study 5, we removed the betting element from the experimental paradigm and found that participants' perceptions of expertise and their choices of whom to hear from on the shape categorisation task were biased by knowledge of the sources' personal values, thus conceptually replicating the effects documented in Chapter 2.

In contrast to Study 1 and 2, here neither accuracy nor similarity affected how much participants were influenced by a source's response. This is likely because people are choosing to hear from the sources whom they think are most accurate on the task – our measure of source influence does not capture how much participants would update their beliefs in response to information from those they did not select. Surprisingly, we also found that the sources' ability to accurately answer questions on the blup task did not influence participants' information-seeking decisions, even though participants perceived accurate sources as more competent on this task. With regard to this latter finding, it is important to note that the effects of source accuracy were generally weak in this experiment. For example, participants' assessments of how similar the sources were to them were also unaffected by the accuracy of the sources on blup trials, in contrast to our previous studies (i.e., Studies 1, 2 and 4). Furthermore, the tendency to rate accurate sources as more competent on the blup task was not statistically significant when we removed participants who failed the learning test. Indeed, it is possible that the effect of source similarity on people's choices of who to seek information from may only arise when learning about source accuracy is weak. Study 6 helps to shed light on this question by examining whether a manipulation that we hypothesised would enhance learning about source expertise (i.e., betting on how the sources answered questions) reduced the influence of similarity on our key dependent variables.

Study 6

In this study, we tested whether asking participants to bet on how sources will answer questions attenuates, or entirely eliminates, the effects of value similarity on expertise learning and information-seeking decisions on the blup task. We hypothesised that the act of predicting how others will respond increases the amount of attention paid to feedback and thus allows participants to build more accurate, and less biased, representations of others' expertise. The experimental design, statistical analyses and hypotheses for this study were preregistered on the Open Science Framework: https://osf.io/yj3m7.

Method

Participants were randomly assigned to one of two conditions: a betting condition or a non-betting condition. In the betting condition, participants were asked to indicate how they thought each of the four sources would answer questions in the learning stage before seeing the sources' answers. Participants in the non-betting condition completed the same tasks as those in Study 5.

We preregistered the following hypotheses:

H1a: Source similarity will have a greater effect on participants' choices of whom to hear from in the non-betting condition than in the betting condition.

H1b: Source accuracy will have a greater effect on participants' choices of whom to hear from in the betting condition than in the non-betting condition.

H2a: Source similarity will have a greater effect on participants' (post learning stage) competence ratings in the non-betting condition than in the betting condition.

H2b: Source accuracy will have a greater effect on participants' (post learning stage) competence ratings in the betting condition than in the non-betting condition.

Participants

Participants were recruited through the UCL Division of Psychology and Language Sciences' online subject pool. Our recruitment targeted first year undergraduate students, enrolled in a BSc Psychology degree. Participants received course credit in exchange for participation. The two participants (one from each betting condition) with the highest scores in this study were additionally awarded a £40 Amazon eGift voucher. Scores were computed by summing the number of choice stage questions participants would be expected to have answered correctly, based on which source they chose and whether they followed their advice, and the number of correct bets made during the learning stage (if they were assigned to the betting condition) and multiplying this by the number of attention check questions answered correctly and the number of learning test questions answered correctly.

As in the previous study, we stated in the study description that this study must be completed on a laptop or PC on Chrome, Firefox, Safari, Microsoft Edge but not Internet Explorer, and that if participants did not comply with this then their data would be excluded, and they would not be eligible to receive the bonus. We also specified in our preregistration that a participant's data would be excluded from the analyses if they answered two or three attention check questions incorrectly.

To determine the sample size, we calculated a power curve for H1a and H1b, based on data from Study 4 and 5, as one of these included betting in the learning stage (i.e., Study 4) and one did not (i.e., Study 5). We did this using the "powerCurve" function from the R package *simr* (Green & MacLeod, 2015). Specifically, we entered participants' choices of whom to hear from on each trial of these studies into a generalised linear mixed-effects model, with source accuracy, source similarity, the study number, the interaction between source accuracy and source similarity, the interaction between source accuracy and study number, the interaction between study number and source similarity, and the three-way interaction between source accuracy, source similarity and study number specified as fixed effects, and Subject ID specified as a random effect (grouping factor). We then specified a null model, which did not include the main effect of study number or the interaction effects between study number and source accuracy, study number and source similarity, or the three-way interaction, and calculated a power curve for a model comparison between the two models. This revealed that a sample size of 56 participants would achieve 85% power to reject the null model in favor of the alternative model.

Sample size was determined for H2a and H2b using a power analysis (G*Power Version 3.1.9.2; Faul et al., 2007), also based on the results of these previous studies. To estimate the likely effect size of H2a and H2b, we entered the standardized (z-scored) competence ratings from Study 4 and 5 into a 2 (source accuracy: accurate, inaccurate) x 2 (source similarity: similar, dissimilar) x 2 (study number: betting study, non-betting study) mixed ANOVA, with the study number entered as a between-subjects factor. This revealed a significant interaction between source accuracy and study number ($F(1,98) = 5.83$, $p = .018$, $\eta_p^2 = 0.06$)

and a significant interaction between source similarity and study number (F(1,98) = 3.86, p = .052, $\eta_p^2 = 0.04$). We then entered the latter (smaller) effect size ($\eta_p^2 = 0.038$) into a power analysis. This suggested that a sample size of 110 participants would achieve 85% power to detect this effect size with an alpha of .05, assuming a correlation among repeated measures of 0.05 (the observed correlation among repeated measures when combining the data from our two previous studies).

We collected data from a total of 135 participants. The reason we recruited more participants than specified by our power analysis was to allow for exclusions resulting from attention check failures, data recording errors, and participants completing the study on a phone or tablet. Three participants were excluded because they completed the study on phones, one participant was excluded because they used an incompatible browser, one participant was excluded due to a technical error recording the data, and two participants were excluded because they answered more than one attention check incorrectly. This left us with a sample size of n = 128 (113 females, 13 males, and 2 who said “other” when asked about their gender, mean age = 18.57, SD = 0.88).

Study Design

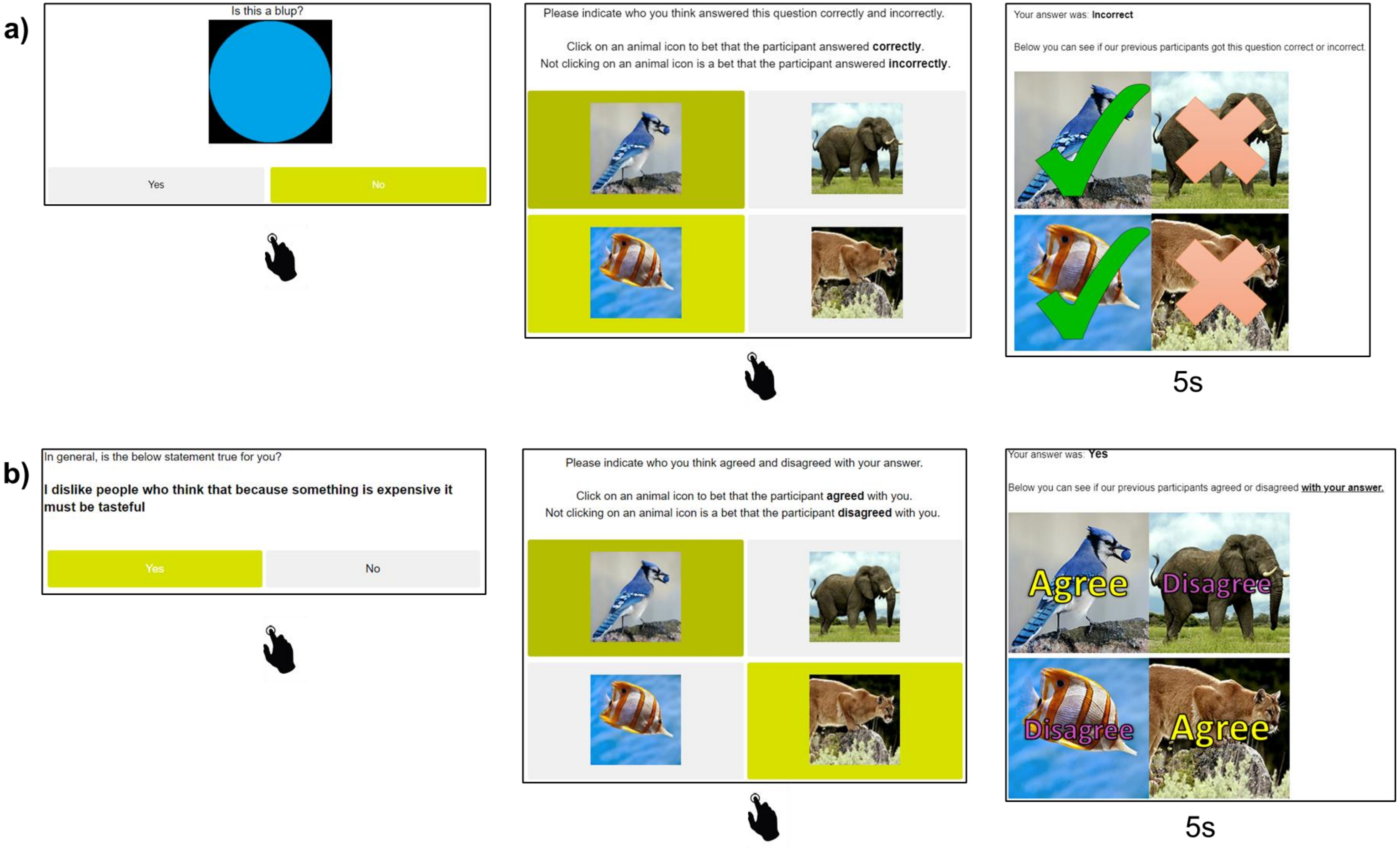
The same design was used as in Study 5, however in this study half of the participants were asked to indicate how they thought each of the four sources would answer questions in the learning stage before seeing the sources’ answers (Figure 28). The instructions were the same for participants in both conditions except those in the betting condition were told that they would be asked to try to predict how the sources answered questions as part of the task (see Appendix 1).

In the blup trials of the learning stage, participants who were randomly assigned to the betting condition were presented with a coloured shape and asked to guess whether it was a ‘blup’ or not (self-paced), like those in the non-betting condition. However, these participants were then shown the pictures representing the four sources (in a 2x2 matrix) and asked to indicate who they thought answered the blup question (in)correctly (self-paced). They were told that clicking on a source would indicate that they thought the source would be correct and not-clicking on a source would indicate that they thought the source would be incorrect.

In the values trials of the learning stage, participants assigned to the betting condition indicated whether they agreed or disagreed with personal values questions (self-paced), just as those in the non-betting condition did. Unlike participants in the non-betting condition, these participants were then shown the pictures representing the four sources (in a 2x2 matrix) and asked to indicate who they thought (dis)agreed with their answer on the same question (self-paced). Whether participants clicked on a source to indicate that they thought the source agreed or disagreed with their answer was counterbalanced in order to account for any influence of habitual (“model-free”) betting behaviour.

Figure 28

Learning Stage in the Betting Condition



Is this a blup?

Please indicate who you think answered this question correctly and incorrectly.

Click on an animal icon to bet that the participant answered **correctly**.
Not clicking on an animal icon is a bet that the participant answered **incorrectly**.

Your answer was: **Incorrect**

Below you can see if our previous participants got this question correct or incorrect.

In general, is the below statement true for you?

I dislike people who think that because something is expensive it must be tasteful

Please indicate who you think agreed and disagreed with your answer.

Click on an animal icon to bet that the participant **agreed** with you.
Not clicking on an animal icon is a bet that the participant **disagreed** with you.

Your answer was: **Yes**

Below you can see if our previous participants agreed or disagreed **with your answer.**

Note. Participants who were assigned to the betting condition completed a modified version of the Learning Stage. As in the non-betting condition (see Figure 24), participants learned about four sources' personal values and accuracy on the blup task. (a) Blup trials and (b) value trials were interleaved. (a) On each blup trial a novel shape was presented and the participants had to indicate whether they believed the shape was a blup (yes or no). They were then asked to bet on who they thought answered the blup question (in)correctly. Afterwards a feedback screen was presented, showing which sources were correct and which were incorrect, as well as whether the participant's answer was correct or incorrect. (b) On each value trial a question assessing personal values was presented and participants had to indicate whether they agreed with it (yes or no). They were then asked to bet on who they thought (dis)agreed with their answer to this question. They then saw whether each of the four sources agreed or disagreed with their answer on this question.

Results

Manipulation Checks

We first checked that the sources responded as we intended. As expected, the accurate sources were correct on approximately 80% of the blup trials (Accurate/Similar: M = 79.63, SD = 6.12. Accurate/Dissimilar: M = 79.49, SD = 6.55) while those programmed to be inaccurate answered approximately 50% of the blup questions correctly (Inaccurate/Similar: M = 51.58, SD = 8.51. Inaccurate/Dissimilar: M = 49.65, SD = 7.90). Entering the percentage of accurate answers given by each source into a 2 (source accuracy condition: accurate, inaccurate) x 2 (source similarity condition: similar, dissimilar) x 2 (betting condition: betting, no-betting) mixed ANOVA, with the betting condition entered as a between-subjects factor and source accuracy and similarity entered as within-subjects factors, revealed a main effect of source accuracy ($F(1,126) = 2141.81$, $p < .001$, $\eta_p^2 = .94$), no main effect of source similarity ($F(1,126) = 2.91$, $p = .091$, $\eta_p^2 = .02$), no interaction between source accuracy and similarity ($F(1,126) = 1.68$, $p = .197$, $\eta_p^2 = .01$), no interaction between source accuracy and the betting condition (F(1,126) = 0.26, p = .611, η_p^2 < .01), no interaction between source similarity and the betting condition (F(1,126) = 0.09, p = .763, η_p^2 = .01), and no three-way interaction (F(1,126) = 0.88, p = .349, η_p^2 = .01).

Likewise, the similar sources agreed with the participants' answers on approximately 80% of the value trials (Accurate/Similar: M = 80.23, SD = 5.21. Inaccurate/Similar: M = 80.45, SD = 6.48), while those programmed to be dissimilar agreed with participants on approximately 20% of these trials (Accurate/Dissimilar: M = 19.67, SD = 6.52. Inaccurate/Dissimilar: M = 19.38, SD = 6.63). A 2 (source accuracy condition: accurate, inaccurate) x 2 (source similarity condition: similar, dissimilar) x 2 (betting condition: betting, no-betting) mixed ANOVA on these percentages revealed a main effect of source similarity (F(1,126) = 13238.57, p < .001, η_p^2 = .99) and no other significant effects (all p-values > .45).

Confirmatory Analyses

Betting Enhanced Participants' Preference for Receiving Information from Accurate Sources but Did Not Attenuate the Effect of Source Similarity. We used a GLME to analyse participants' information-seeking choices. Specifically, we entered participants' choices of whom to seek information from in the choice stage as the dependent variable in a GLME with a binomial response variable distribution. The (standardised) difference in source accuracy and the (standardised) difference in source similarity between the source presented on the right and left were included as fixed factors, as was the interaction between the two. The betting condition was also entered as a fixed factor, as was the two-way interaction between the betting condition and the standardized source accuracy difference, the two-way interaction between the betting condition and the standardized source similarity difference, and the three-way interaction. Subject ID was entered as a random factor (grouping variable). Note, we preregistered that we would use a random-intercept only GLME here, rather than entering the maximal random effects structure, because data from our previous studies suggested that including the interaction between the standardized source accuracy difference and the standardized source similarity difference as a random factor (slope) reduced the model fit, according to both the Bayesian Information Criterion (BIC) and the Akaike Information Criterion (AIC).

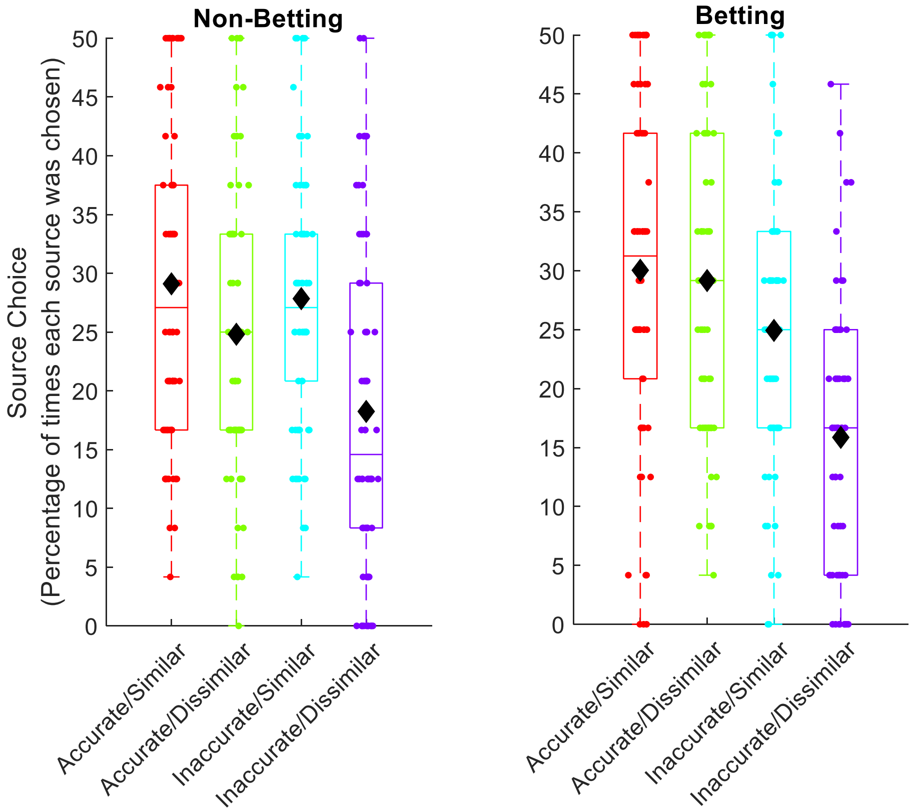
We preregistered that H1a would be supported if we found a significant interaction between source similarity and the betting condition, with a greater effect of source similarity on source choice in the non-betting condition than the betting condition. We preregistered that H1b would be supported if we found a significant interaction between source accuracy and the betting condition, with a greater effect of source accuracy on source choice in the betting condition than in the non-betting condition.

The GLME revealed a main effect of source similarity (Figure 29), indicating that participants preferred to receive information on blup questions from sources that held similar values to them (β = 0.34, SE = 0.05, 95% CIs = [0.24, 0.44], t(3064) = 6.46, p < .001). However, contrary to what we predicted in H1a, there was no interaction between source similarity and the betting condition (β = -0.11, SE = 0.07, 95% CIs = [-0.26, 0.04], t(3064) = -1.46, p = .144), suggesting that betting on how the sources would answer questions in the learning stage did not attenuate the effect of source similarity on participants' information-seeking decisions in the choice stage. The GLME also revealed a main effect of source accuracy, indicating that accurate sources were chosen more often than inaccurate sources (β = 0.22, SE = 0.05, 95% CIs = [0.12, 0.33], t(3064) = 4.25, p < .001). Our second hypothesis (H1b) was supported by a significant interaction between source accuracy and the betting condition; participants in the betting condition were more influenced by source accuracy when choosing who to receive information from than those in the non-betting condition (β = 0.26, SE = 0.08, 95% CIs = [0.12, 0.41], t(3064) = 3.48, p < .001). The interaction between source accuracy and similarity was not significant (β = 0.04, SE = 0.06, 95% CIs = [0.12, 0.41], t(3064) = 0.72, p = .473) and nor was the three-way interaction between source accuracy and source similarity and the betting condition (β = -0.09, SE = 0.09, 95% CIs = [-0.26, 0.07], t(3064) = -1.10, p = .273).

To find out what was driving the interaction between source accuracy and the betting condition, we ran two more GLMEs, as planned in our preregistration – one using participants in the betting condition and one using those in the non-betting condition – with source choice entered as the dependent variable; the (standardised) difference in source accuracy, the (standardised) difference in source similarity, and the interaction between the two entered as fixed factors, and Subject ID entered as a random factor (grouping variable). These analyses revealed that source accuracy and source similarity had significant effects on participants' information-seeking choices in both conditions (all p-values < .001), although the effect of accuracy was greater in the betting condition (β = 0.49, SE = 0.05, 95% CIs = [0.38, 0.59]) than in the non-betting condition (β = 0.22, SE = 0.05, 95% CIs = [0.12, 0.33]) (hence the interaction between source accuracy and the betting condition above). The interaction between source accuracy and source similarity was not significant in either model (both p-values > .40).

Figure 29

Percentage of Trials on Which Participants Chose to Seek Information from Each Source



Note. Participants preferred to receive information on blup questions from the more accurate sources and the more similar sources. However, participants in the betting condition (right-hand side) were more influenced by source accuracy when seeking information than those in the non-betting condition (left-hand side). The percentage of times each participant decided to hear from each source is plotted (coloured dots). The black diamonds represent the mean of these percentages. The box plots show the distribution of these percentages: boxes indicate 25–75% interquartile range, whiskers extend from the first and third quartiles to most extreme data point within 1.5 × interquartile range, and the median is shown as a horizontal line within this box.

Betting Enhanced Expertise Learning but Did Not Attenuate the Effect of Source Similarity. We entered participants' ratings of how competent each source was in the shape categorization trials of the learning stage into a 2 (source accuracy condition: accurate, inaccurate) x 2 (source similarity condition: similar, dissimilar) x 2 (betting condition: betting, no-betting) mixed ANOVA, with the betting condition entered as a between-subjects factor and source accuracy and similarity entered as within-subjects factors, to test whether the source's answers on shape categorisation and personal values questions influenced participants' perception of how many blup questions the sources answered correctly. We preregistered that H2a would be supported if we found a significant interaction between source similarity and the betting condition, with a greater effect of source similarity on competence ratings in the non-betting condition than in the betting condition. Likewise, we preregistered that H2b would be supported if we found a significant interaction between source accuracy and the betting condition, with a greater effect of source accuracy on competence ratings in the betting condition than the non-betting condition. Significant interactions were followed-up with analyses of each betting condition separately. Note, we preregistered that this analysis would be performed using a mixed ANOVA, rather than a mixed effects model, because power analyses revealed that the required sample size was smaller if analysing the competence ratings using within-subjects conditions (i.e., Accurate/Similar, Accurate/Dissimilar, Inaccurate/Similar, Inaccurate/Dissimilar) rather than correlating variability in the sources' accuracy with variability in competence ratings and ignoring the experimental conditions (which is essentially what the mixed model is doing). We therefore also used ANOVAs in the exploratory analyses, where appropriate (e.g., analysing average betting behaviour and similarity ratings, but not COM scores).

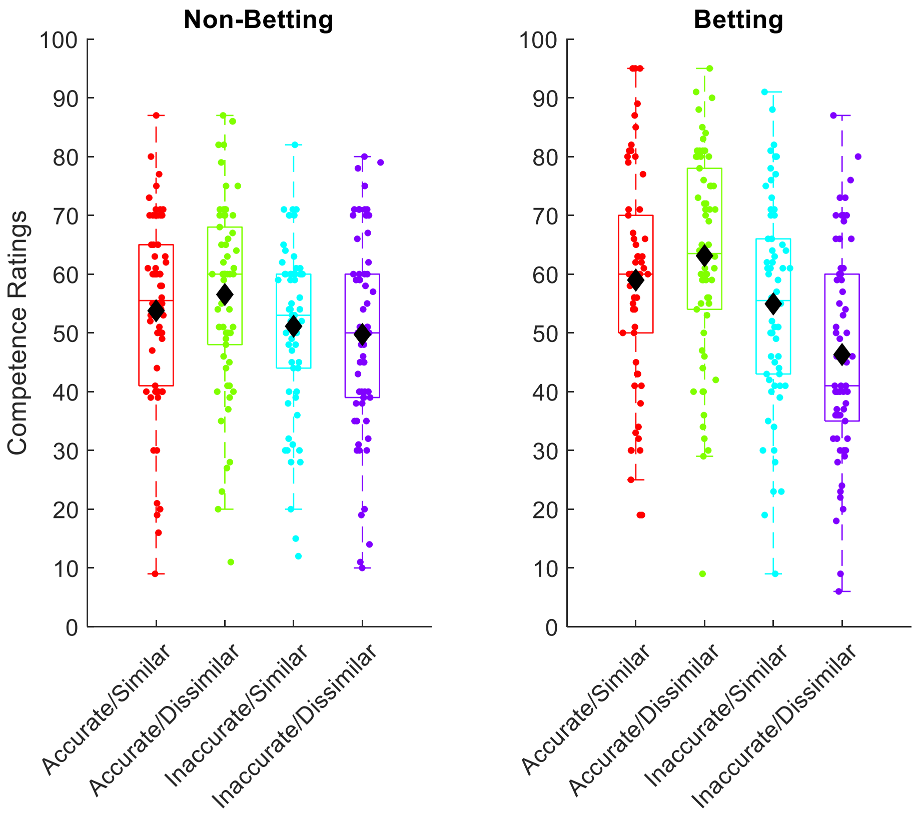
The rmANOVA did not support our hypothesis (H2a) that betting would attenuate the effect of value similarity on participants' competence ratings (Figure 30). In particular, the interaction between source similarity and the betting condition was not significant (F(1,126) = 1.05, p = .308, η_p^2 = .01). There was also no main effect of source similarity (F(1,126) = 0.29, p = .589, η_p^2 < .01), however there was an interaction between source accuracy and source similarity (F(1,126) = 12.12, p = .001, η_p^2 = .09), which was due to participants rating the Inaccurate/Similar source as more competent on the blup task than the Inaccurate/Dissimilar source (t(127) = 2.50, p = .014) while not rating the two accurate sources as significantly different (t(127) = 1.91, p = .059). The three-way interaction between source accuracy, source similarity and the betting condition was not significant (F(1,126) = 3.13, p = .079, η_p^2 = .02), suggesting that the tendency for value similarity to interfere with how participants learned about *inaccurate* sources was not attenuated by the betting manipulation (although, it should be noted that we did not power our experiment to investigate this three-way interaction).

Our final preregistered hypothesis (H2b) – namely, that source accuracy would have a greater effect on participants' competence ratings in the betting condition than in the non-betting condition – was supported. In particular, we found that participants effectively learned that accurate sources answered more blup questions correctly than inaccurate sources (main effect of source accuracy: F(1,126) = 41.49, p < .001, η_p^2 = .25) and that the interaction between source accuracy and the betting condition was significant (F(1,126) = 5.90, p =.017, η_p^2 = .05). To determine what was driving the interaction, we summed the competence ratings for the accurate sources and then subtracted from this the combined competence ratings for the inaccurate sources. We then compared the difference between betting conditions.

This revealed that the difference in competence ratings for accurate versus inaccurate sources was greater in the betting condition (M = 20.87, SD = 26.45) than in the non-betting condition (M = 9.44, SD = 26.75). Thus, as predicted, the interaction was due to participants in the betting condition learning about accuracy more strongly than those in the non-betting condition.

Figure 30

Participants' Ratings of Each Source's Competence in the Non-Betting and Betting Conditions



Note. Participants learned that accurate sources were more competent at answering blup questions than inaccurate sources. Expertise learning was enhanced for participants in the betting condition (right-hand side) compared to those in the non-betting condition (right-hand side). Interestingly, the inaccurate source that tended to agree on questions relating to personal values was rated as more accurate on blups than the inaccurate source that tended to disagree on value questions, suggesting that value similarity partially influences perceived expertise.

The competence rating for each source is plotted (coloured dots). The black diamonds represent the mean of these ratings. The box plots show the distribution of the competence ratings: boxes indicate 25–75% interquartile range, whiskers extend from the first and third quartiles to most extreme data point within 1.5 × interquartile range, and the median is shown as a horizontal line within this box.

Exploratory Analyses

Participants' Betting Behaviour in the Blup Trials Was Unaffected by Source Similarity in the Value Trials. To investigate whether the source's answers on blup trials and value trials influenced who participants thought would answer shape categorisation questions correctly, we entered the percentage of bets on each source in the blup trials (by each participant assigned to the betting condition) into a 2 (source accuracy condition: accurate, inaccurate) x 2 (source similarity condition: similar, dissimilar) rmANOVA. This revealed that participants placed more bets on accurate sources in the blup trials of the learning stage ($F(1,61) = 28.74$, $p < .001$, $\eta_p^2 = .32$). The amount that sources agreed with participants' answers in the value trials did not affect participants betting behaviour in the blup trials ($F(1,61) = 0.79$, $p = .377$, $\eta_p^2 = .01$). The interaction between source accuracy and similarity was not significant ($F(1,61) = 2.70$, $p = .11$, $\eta_p^2 = .04$).

Modelling Participants' Bets in the Blup Trials. As the percentage of bets placed on the sources ignores the temporal dynamics of how participants' beliefs about each source's expertise changes over time, we also fit several computational models to this data. Specifically, we tested whether unbiased reinforcement-learning (RL) and beta-binomial (Bayesian) models provided a better fit to the data than comparable models that included a congruence bias. The congruence bias models assumed that participants' beliefs about the sources personal values influenced how much they update their beliefs about the sources' competence in light of the evidence they received on the blup trials (the models and the fitting and comparison procedures are described in more detail in Chapter 3). The model statistics are presented below (Table 13).

Table 13

Expertise Learning Model Comparison Results

Model No.	BIC	AIC	Mean Pseudo R-Squared	% of participants for whom model fit best (BIC)	% of participants for whom model fit best (AIC)
RL models using data from blup trials only (Set 1)					
1. RL Unbiased	13188	12807	0.09	74%	53%
2. RL Congruence Bias	**13160**	**12588**	0.11	26%	47%
Bayesian models using data from blup trials only (Set 1)					
3. BB Unbiased	**13028**	12456	0.12	65%	45%
4. BB with 1 Scaling Parameter	13116	12354	0.14	15%	11%
5. BB Congruence Bias	13093	**12140**	0.16	21%	44%
RL models using data from blup and value trials (Set 2)					
1. RL Unbiased	**26228**	25527	0.09	82%	58%
2. RL Congruence Bias	26402	**25468**	0.09	18%	42%
Bayesian models using data from blup and value trials (Set 2)					
3. BB	26384	25216	0.11	69%	40%

Unbiased					
4. BB with 1 Scaling Parameter	**26281**	24879	0.12	23%	34%
5. BB Congruence Bias	26478	**24842**	0.13	8%	26%

Note: RL = Reinforcement-learning, BB = Beta-binomial.

The results of the model comparisons are easiest to interpret if one looks at the number of participants to whom each model fit best, as the overall BICs and AICs tend to offer contradictory conclusions. The individual-level fits show that in each model comparison the unbiased model outperformed the congruence bias model. This conclusion is further supported by the fact that Wilcoxon signed rank tests revealed that the two learning rates [scaling parameters] in the RL [beta-binomial] congruence bias models did not differ significantly (Table 14). This is consistent with there being no effect of source similarity in our analysis of participants' average betting behaviour in the blup trials.

Table 14

Wilcoxon Signed Rank Tests Comparing the Magnitude of the Learning Rates [Scaling Parameters] Included in the Congruence Bias Models

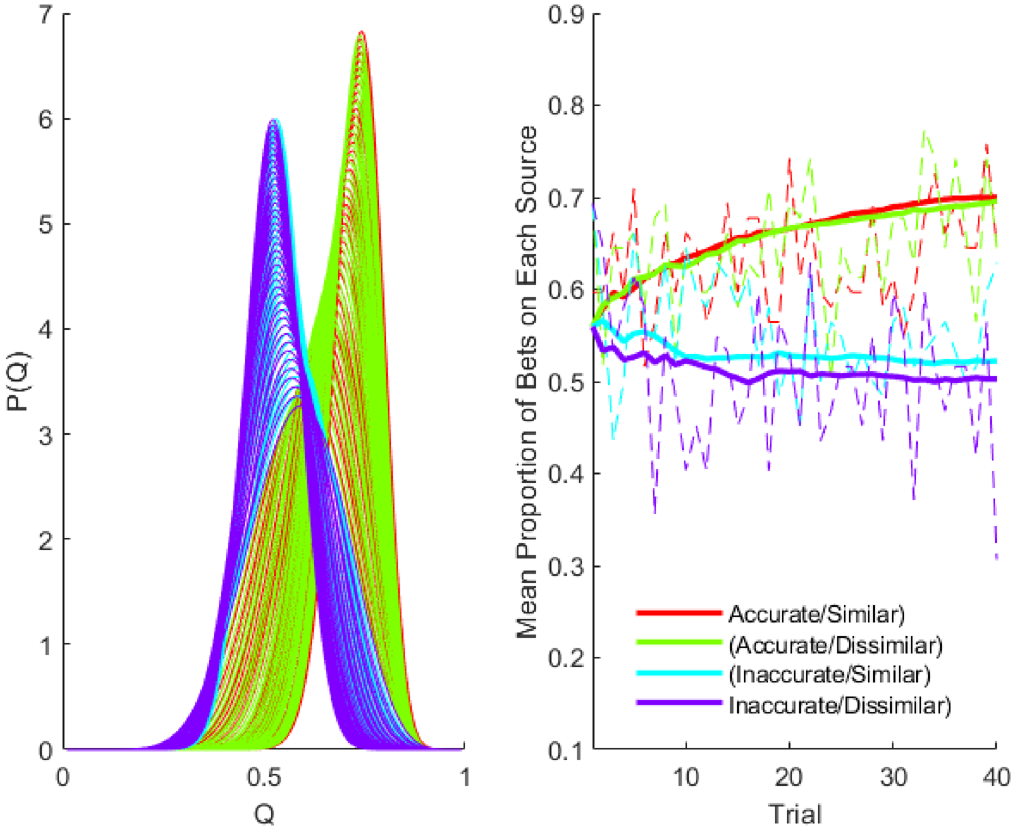
Congruence Bias Model	**Median α_1 [γ_1]**	**Median α_2 [γ_1]**	**Z**	**p**
RL, using data from blup trials only (Set 1)	0.08	0.08	-0.65	.517
BB, using data from blup trials only (Set 1)	0.23	0.60	-0.33	0.739
RL, using data from blup and value trials (Set 2)	0.01	0.03	-1.23	.219
BB, using data from blup and value trials (Set 2)	1.01	1.27	0.35	.729

Note: RL = Reinforcement-learning, BB = Beta-binomial. α_1 is the learning rate parameter applied to congruent feedback in the RL models; α_2 is the learning rate parameter applied to incongruent feedback in the RL models; γ_1 is the scaling parameter applied to congruent feedback in the BB models; γ_2 is the scaling parameter applied to incongruent feedback in the BB models.

Participants' betting behaviour on the blup trials was reasonably well-described by a standard beta-binomial model (Figure 31). Using the best-fit parameters from the standard beta-binomial model (Model 5), we simulated the bets of each participant on each trial 1000 times and computed the mean accuracy of these bets (i.e., the percentage of times the model's prediction matched the participant's behaviour) across these simulations. Overall, the model accurately predicted participants' bets on 57% of blup trials (chance level is 50%). Specifically, it predicted 59.17% of participant's bets on the Accurate/Similar source, 59.90% of bets on the Accurate/Dissimilar source, 54.16% on the Inaccurate/Similar source, and 54.08% on the Inaccurate/Dissimilar source.

Figure 31

A Standard Beta-Binomial Model Fit to Participants' Bets on Each Source in the Blup Trials



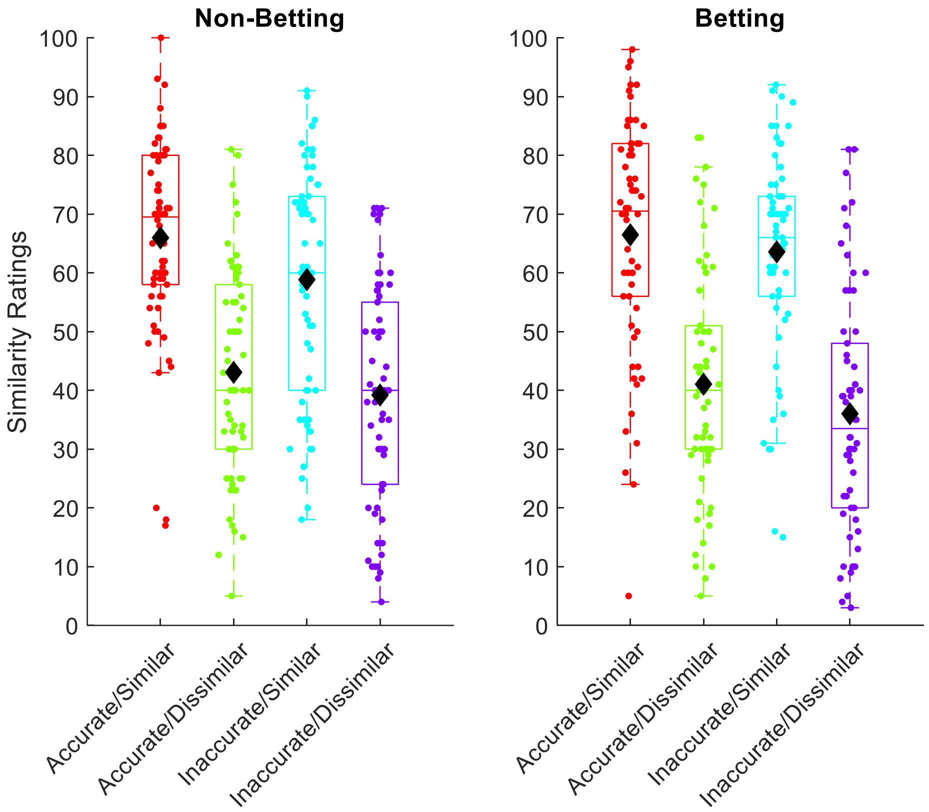
Note. Left-hand side: The probability distributions illustrate how participants' beliefs about each source's expertise (Q) evolved over the course of the learning stage. Each distribution was calculated by averaging the model parameters for each trial across participants. The distributions from each trial are plotted one on top of the other. Before observing any evidence pertaining to the sources' competence at answering blup questions, the prior distribution did not vary by source. Participants' beliefs about each source were updated on each trial. Over the course of the learning stage, the model suggests that participants became less uncertain in their beliefs – as evidenced by the increasing height of the distributions – and learned which sources were more and less accurate – as evidenced by the leftward movement for inaccurate sources and rightward movement for accurate sources. Right-hand side: Solid lines show the mean model-predicted probability of betting on each source on every blup trial of the learning stage. Dotted lines show the proportion of participants that actually bet on each source on each blup trial.

Betting Did Not Affect How Participants Learned About Source Similarity. As we were primarily interested in how biases in expertise learning affect information-seeking decisions, we did not preregister any hypotheses regarding the effects of betting on similarity learning. However, one may wonder whether learning about expertise on the blup task influenced similarity learning on the value trials and if the betting manipulation attenuated this effect. Here, we performed an exploratory analysis on participants' similarity ratings akin to the confirmatory analysis on participants' competence ratings. That is, we entered participants' ratings of how similar each source was to them on value questions into a 2 (source accuracy condition: accurate, inaccurate) x 2 (source similarity condition: similar, dissimilar) x 2 (betting condition: betting, no-betting) mixed ANOVA, with the betting condition entered as a between-subjects factor and source accuracy and similarity entered as within-subjects factors.

This revealed that participants' perceptions of similarity were influenced by the percentage of trials the sources agreed with the participants' answers on questions relating to personal values (i.e., main effect of source similarity: F(1,126) = 143.51, p < .001, η_p^2 = .53) and the percentage of blup questions the sources answered correctly (i.e., main effect of source accuracy: F(1,126) = 13.99, p < .001, η_p^2 = .10). The betting manipulation did not moderate the effect of source similarity (F(1,126) = 1.71, p = .194, η_p^2 = .01) or source accuracy (F(1,126) = 0.40, p = .549, η_p^2 < .01). There was no interaction between source accuracy and source similarity (F(1,126) = 0.04, p = .851, η_p^2 < .01) and no three-way interaction between source accuracy, source similarity, and the betting condition (F(1,126) = 0.83, p = .365, η_p^2 = .01) (Figure 32). Thus, there was an epistemic spillover effect, such that participants perceived sources that performed better on the blup task as having more similar values to them, which was not attenuated by having participants bet on how sources would answer questions during the learning stage.

Figure 32

Participants' Ratings of Each Source's Similarity in the Non-Betting and Betting Conditions



Note. Accuracy on the blup task affected participants' perceptions of how often the sources agreed with their answers on questions relating to personal values. The betting manipulation did not moderate this effect, suggesting that similarity learning was biased by expertise in the blup task in both the betting condition (right-hand side) and the non-betting condition (right-hand side). The similarity rating for each source is plotted (coloured dots). The black diamonds represent the mean of these ratings. The box plots show the distribution of the similarity ratings: boxes indicate 25–75% interquartile range, whiskers extend from the first and third quartiles to most extreme data point within 1.5 × interquartile range, and the median is shown as a horizontal line within this box.

Participants' Average Betting Behaviour in the Value Trials Was Influenced by Source Accuracy on the Blup Task. As with the betting data from the blup trials, we entered the percentage of bets on each source made in the value trials (by each participant assigned to the betting condition) into a 2 (source accuracy condition: accurate, inaccurate) x 2 (source similarity condition: similar, dissimilar) rmANOVA to test whether participants' bets in the value trials were influenced by the sources' answers on the blup trials. The rmANOVA indicated that, on average, participants' betting behaviour in the value trials was influenced by how often the sources agreed with them on these trials ($F(1,61) = 44.87$, $p < .001$, $\eta_p^2 = .42$), as well as how accurate they were on the blup task ($F(1,61) = 4.92$, $p = .030$, $\eta_p^2 = .08$). The interaction between source accuracy and similarity was not significant ($F(1,61) < 0.01$, $p = .956$, $\eta_p^2 < .01$). This is consistent with the finding above indicating that participants believed sources that performed better on the blup task as having more similar values to them.

Modelling Participants' Bets in the Value Trials. We also fit computational models to the betting data from the value trials. We used the same models here as described previously (see Chapter 3 for more details) to test whether participants exhibited a congruence bias when learning about source similarity, such that their beliefs about a source's expertise on the blup task would affect how much they learned from evidence pertaining to that source's similarity to them on value trials. The model statistics are presented below (Table 15).

Table 15

Similarity Learning Model Comparison Results

Model number and name	BIC	AIC	Mean Pseudo R-Squared	% of participants for whom model fit best (BIC)	% of participants for whom model fit best (AIC)
RL models using data from value trials only (Set 1)					
1. RL Unbiased	**13184**	12803	0.09	85%	79%
2. RL	13374	**12802**	0.10	15%	21%

Congruence Bias					
Bayesian models using data from value trials only (Set 1)					
3. BB Unbiased	**13109**	12537	0.12	92%	76%
4. BB with 1 Scaling Parameter	13267	**12504**	0.13	6%	11%
5. BB Congruence Bias	13477	12523	0.13	2%	13%
RL models using data from value and blup trials (Set 2)					
1. RL Unbiased	**26228**	**25527**	0.09	97%	76%
2. RL Congruence Bias	26487	25552	0.09	3%	24%
Bayesian models using data from value and blup trials (Set 2)					
3. BB Unbiased	26384	25216	0.11	73%	50%
4. BB with 1 Scaling Parameter	**26351**	**24950**	0.12	26%	34%
5. BB Congruence Bias	26672	25036	0.12	2%	16%

Note: RL = Reinforcement-learning, BB = Beta-binomial.

The model comparisons indicated that participants did not exhibit a congruence learning bias when betting on how sources would answer in the value trials. The individual-level fits show that in each model comparison the unbiased model outperformed the congruence bias model.

Wilcoxon signed rank tests suggested that the magnitude of the two scaling parameters in Bayesian congruence bias models differed significantly (Table 16), suggesting that participants learned more from congruent than incongruent evidence. The two learning rates in one of the RL congruence bias models (the one in which beliefs about source accuracy were determined using trial-by-trial estimates) were also significantly different, however the effect was in the opposite direction, suggesting that participants learned more from incongruent than congruent evidence. The learning rates in the RL congruence bias model in which accuracy beliefs were coded using predefined categorical variables were not significantly different from each other in the other. However, as these models provided a worse fit to the data than the unbiased versions, and the analyses of the learning rates and scaling parameters is inconsistent, we concluded that they do not indicate that there is a general congruence bias affecting similarity learning.

Table 16

Wilcoxon Signed Rank Tests Comparing the Magnitude of the Learning Rates [Scaling Parameters] Included in the Congruence Bias Models

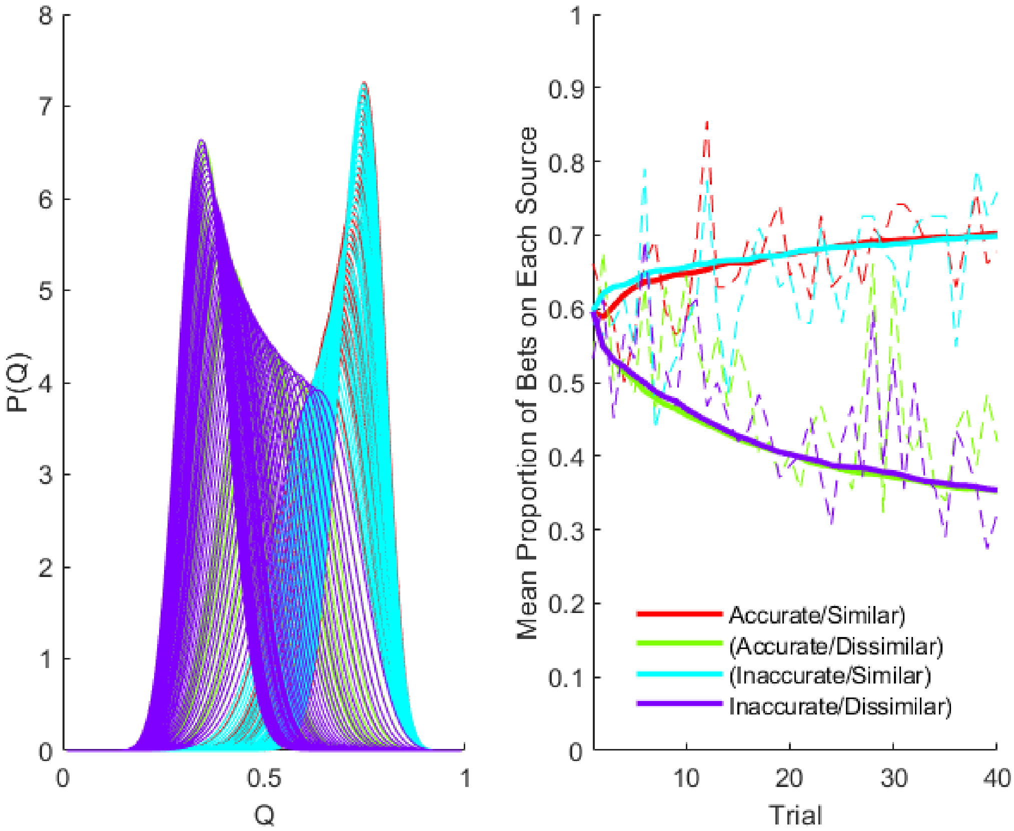
Congruence Bias Model	Median α_1 [γ_1]	Median α_2 [γ_1]	Z	p
RL, using data from political trials only (Set 1)	0.04	0.04	1.21	.23
BB, using data from political trials only (Set 1)	0.49	0.11	3.24	.001
RL, using data from political and quiz trials (Set 2)	0.01	0.02	-2.54	.011
BB, using data from political and quiz trials (Set 2)	0.37	0.23	2.92	.004

Note: RL = Reinforcement-learning, BB = Beta-binomial. α_1 is the learning rate parameter applied to congruent feedback in the RL models; α_2 is the learning rate parameter applied to incongruent feedback in the RL models; γ_1 is the scaling parameter applied to congruent feedback in the BB models; γ_2 is the scaling parameter applied to incongruent feedback in the BB models.

The standard beta-binomial model provided a reasonably good fit to the betting behaviour on the political trials (Figure 33). As done for the blup trials, we simulated the bets of each participant on each value trial 1000 times using the best-fit parameters from the standard beta-binomial model and computed the mean accuracy of these bets (i.e., the percentage of times the model's prediction matched the participant's behaviour) across the simulations. This indicated that on average the model accurately predicted participants' bets on 57% of political trials. Specifically, it predicted 58.60% of participant's bets on the Accurate/Similar source, 55.50% of bets on the Accurate/Dissimilar source, 57.74% on the Inaccurate/Similar source, and 55.56% on the Inaccurate/Dissimilar source.

Figure 33

A Standard Beta-Binomial Model Fit to Participants' Bets on Each Source in the Value Trials



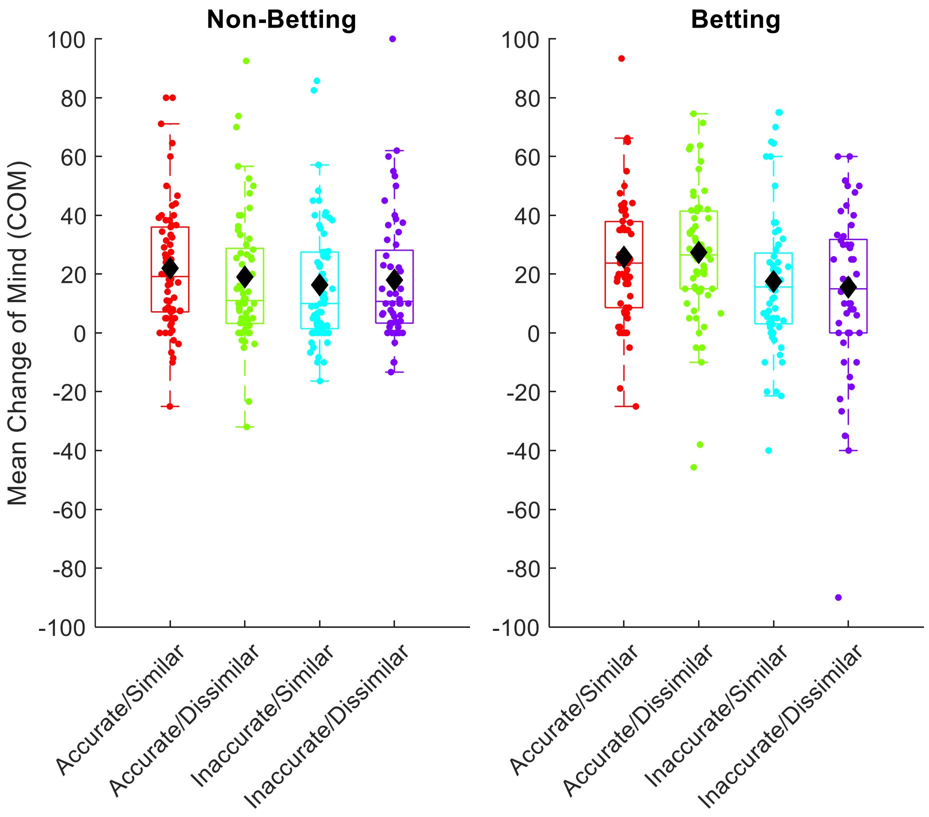
Note. Left-hand side: The probability distributions illustrate how participants' beliefs about each source's expertise (Q) evolved over the course of the learning stage. Each distribution was calculated by averaging the model parameters for each trial across participants. The distributions from each trial are plotted one on top of the other. Before observing any evidence pertaining to the sources' similarity regarding personal values, the prior distribution did not vary by source. Participants' beliefs about each source were updated on each trial. Over the course of the learning stage, the model suggests that participants became less uncertain in their beliefs – as evidenced by the increasing height of the distributions – and learned which sources were more and less similar to them – as evidenced by the leftward movement for dissimilar sources and rightward movement for similar sources. Right-hand side: Solid lines show the mean model-predicted probability of betting on each source on every value trial of the learning stage. Dotted lines show the proportion of participants that actually bet on each source on each value trial.

Betting Enhanced Participants' Receptivity to Information from Accurate Sources. We computed COM scores for each trial of the choice stage (for more details on how COM is calculated, see Chapter 3) and entered it as the dependent variable in a linear mixed model. Source accuracy (i.e., the percentage of times the chosen source answered blup questions correctly in the learning stage, z-scored), source similarity (i.e., the percentage of times the chosen source agreed with the participant's answer on value trials, z-scored), and their interaction were entered as fixed factors. The betting condition was also added as a fixed factor, along with its interactions with source accuracy and source similarity. We also included a variable indicating whether the source agreed or disagreed with the participant's answer on each trial, and its interactions with source accuracy and similarity, as fixed factors. Subject ID was entered as a random (grouping) factor.

This analysis revealed a main effect of source accuracy on COM (Figure 34), suggesting that participants' belief updating in response to socially acquired information was sensitive to the accuracy of the chosen source (β = 2.09, SE = 0.78, 95% CIs = [0.57, 3.61], t(3060) = 2.69, p = .007). There was also an interaction between source accuracy and the betting condition (β = 1.94, SE = 0.77, 95% CIs = [0.43, 3.46], t(3060) = 2.51, p = .012), indicating that participants were more influenced by accurate sources in the betting condition than the non-betting condition. There was no main effect of source similarity on COM (β = 0.37, SE = 0.76, 95% CIs = [-1.12, 1.86], t(3060) = 0.49, p = .622) and the interaction between source similarity and the betting condition was not significant (β = -0.11, SE = 0.76, 95% CIs = [-1.60, 1.37], t(3060) = -0.15, p = .881). The intercept of the model was significant (β = 24.45, SE = 1.37, 95% CIs = [21.76, 27.15], t(3060) = 17.79, p < .001), indicating that overall participants were positively influenced by the sources' answers on blup questions, and COM was greater when the chosen source disagreed with the participant's initial answers than when they agreed (β = 19.50, SE = 0.74, 95% CIs = [18.05, 20.96], t(3060) = 26.29, p < .001). All other factors in the model were non-significant (all p-values > .10).

Figure 34

Participants' Mean Change of Mind After Receiving Information from Each Source in the Non-Betting and Betting Conditions



Note. Participants' blup judgments were more influenced by accurate sources than inaccurate sources, with those in the betting condition (right-hand side) displaying increased sensitivity to source accuracy compared to those in the non-betting condition (left-hand side). Change of Mind (COM) quantifies the average change in participants' beliefs about blups after receiving information from a source by taking into account both the participant's change of decision and confidence (see methods in Chapter 3 for more details). The mean COM for each source per participant is plotted (coloured dots). The black diamonds represent the average of these means. The box plots show the distribution of the mean COM values: boxes indicate 25–

75% interquartile range, whiskers extend from the first and third quartiles to most extreme data point within 1.5 × interquartile range, and the median is shown as a horizontal line within this box.

Discussion

The results of Study 6 indicate that betting on how accurate others will be in their predictions enhances expertise learning and thereby the ability to make wise information-seeking choices. However, it does not, in and of itself, attenuate the effects of similarity on expertise learning and therefore suggests that the null results reported in Study 3 and 4 are not solely due to inattentiveness.

Our confirmatory analyses revealed that participants chose to hear from accurate sources more often than inaccurate sources, especially if they were asked to bet on how sources would answer questions during the learning stage. Thus, H1b was supported. Participants' subjective ratings of the sources indicated that those who were asked to make bets learned about the sources' expertise on the blup task more efficiently than those who were not, thus supporting H2b. However, the degree to which source similarity influenced information-seeking decisions was unaffected by our betting manipulation. Therefore, H1a was not supported. Likewise, betting did not attenuate participants' misperception that the source that was inaccurate on the blup task but shared their values was better at the task than an equally incompetent source that disagreed with them on questions relating to personal values. Overall, these results suggest that while making predictions about others before seeing (dis)confirming evidence does enhance learning, the inclusion of betting in the experimental paradigm did not alone explain why source generosity and similarity did not bias expertise learning or information-seeking in Study 3 and 4.

Our exploratory analyses produced a more mixed pattern of results. First, in contrast to the results of the confirmatory analysis, two analyses of the betting data from the Blup trials – one using the percentage of bets each participant made on each source and the other using a computational modelling approach – indicated that participants were no more likely to expect sources that shared their values to be correct when answering shape categorisation questions than those who had different values. This suggests that participants' belief updating in response to feedback about the sources' expertise on the blup task was unbiased by value similarity, while their recall of which sources were accurate and which were less so was affected by source similarity ex post.

Perceptions of value similarity were influenced by how accurate the sources were on the blup task. Consistent with the results of our confirmatory analyses, this bias was not attenuated by the betting manipulation. However, here, the betting manipulation did not enhance learning about similarity either. That is, those who were asked to try to predict how the sources would respond to questions in the learning stage were no more likely to rate (dis)similar sources as (dis)similar to them than those who were not asked to make bets. It is possible that participants in the *non-betting* condition were able to fully learn about value similarity (i.e., they were at ceiling levels), due to the large discrepancy between similar sources (80% agreement) and dissimilar sources (20% agreement). Therefore, including betting in the procedure could not boost similarity learning.

Participants' average betting behaviour on the value trials also showed evidence of an epistemic spillover effect; they bet that the accurate sources would agree with them on value questions more often than the inaccurate sources. However, computational modelling did not indicate that this difference in betting behaviour at the average level was characterised by a congruence bias. That is, models assuming that participants overweighted evidence suggesting that accurate sources were similar to them and inaccurate sources were dissimilar to them and underweighted evidence to the contrary provided a worse fit to the data than comparable unbiased models. It is possible that the modelling procedure provided a less sensitive test than the analysis of the mean behaviour across all trials. Alternatively, the bias that we observed at the average level may have been driven by a different mechanism from the one we tested.

Finally, we found that participants were more influenced by the information they received from accurate than inaccurate sources. Moreover, betting on how the sources would respond when learning about source accuracy on the blup task and source similarity on personal values questions accentuated the degree to which participants were influenced by the accuracy of the source when receiving information. These results align well with our other findings suggesting that the betting manipulation enhanced expertise learning.

General Discussion

Reverting a number of the changes that were made in Study 3 and 4 led to us once again finding the epistemic spillover effects that were observed in Study 1 and 2. Specifically, in the two studies presented here, we found that learning about others' similarity in a domain outside of politics biased expertise learning and whom people chose to seek information from, thus revealing that the effects observed in Study 1 and 2 are not specific to US politics. The fact that a sample of first-year undergraduate Psychology students were recruited for Study 6 also clarifies that the effects reported in Study 1 and 2 are not only observed with online (e.g., MTurk, Prolific) participants.

Here, we manipulated value, rather than political, similarity because people's values are core to their identity and provide standards for what they find most desirable when evaluating behaviours and situations (Schwartz, 1996). Therefore, learning that others share one's values should activate an important social identity. Our results are consistent with findings from previous research indicating that people cluster and segregate on the basis of personal values (Lee et al., 2009; Lönnqvist & Itkonen, 2016). As values are theorised to underlie political attitudes (Schwartz et al., 2010), it is also possible that an underlying sense of shared values explains why people formed overly positive views of others in Study 1 and 2.

Based on the findings from Study 3, 4, and 5, we hypothesised that asking participants to bet on how the sources would respond focused their attention on the sources' answers, thereby enhancing learning and reducing halo effects. We tested this hypothesis in Study 6 and found that while including betting in the learning stage did enhance expertise learning, it did not attenuate the biasing effects of source similarity on perceptions of expertise or information-seeking decisions. This suggest that the null effects we observed in Chapter 3 were not fully explained by the fact that we asked participants to bet on how sources would respond to questions before seeing their actual answers.

There are several features, other than the inclusion of betting, that could explain why we found similarity influenced expertise learning and information-seeking in the studies in chapter two and four but not three. First, the two studies presented in this chapter, like the Chapter 2 studies, were conducted online rather than in the lab (even though participants recruited in Study 6 were undergraduate students completing the study for course credit as opposed to online participants completing the study for monetary payment). Participants in the Chapter 3 studies may have thus been more attentive than the others, as they were supervised while completing the task. However, research showing that online participants are just as attentive as typical undergraduate subject populations (Hauser & Schwarz, 2016; Paolacci et al., 2010), casts some doubt on this explanation. Second, participants in the Chapter 3 studies were informed that they could earn points based on their ability to predict how the sources would answer questions, shown feedback after each trial indicating how many points they had earned from betting on the sources' responses, and were not asked to answer shape categorisation while learning about the sources. In contrast, those who took part in our other studies were not awarded points for correct answers in the experiment and were asked to focus on learning which shapes were blaps (or blups) and which were not whilst also learning about the characteristics of the sources. Participants in the Chapter 3 studies may therefore have been more incentivised to perform well on the task and have had greater attentional capacity than those in our other studies. Third, participants in the Chapter 3 studies, in which we did not find effects of generosity or political similarity on expertise learning or information-seeking on the blup task, did not complete the learning test, whereas those in the Chapter 2 and 4 studies did. In the learning test, participants were asked to assess which of the sources were more similar to them. This may have primed them to focus on similarity, thus affecting their subsequent ratings and choices. Consistent with this hypothesis, the results of Study 6 indicated that source similarity influenced participants' post-learning-test competence ratings and information-seeking choices, but not their pre-learning-test bets on which sources would be accurate on the blup task. Still, further research would be needed to disentangle whether each of these features affect the degree to which irrelevant source characteristics influence expertise learning and information-seeking in this experimental task. Fourth, it is possible that betting did not attenuate the effects of value similarity in Study 6 because personal values are important to people's identities. We previously hypothesised that generosity and political similarity manipulations used in Chapter 3 may not have induced a relevant social identity, as many of our participants were not UK-nationals. If the value stimuli used in Chapter 4 tapped into a more relevant underlying social dimension than the charity donation or political statement stimuli used in Chapter 3, then we might conclude that epistemic spillovers only influence expertise learning, and thus information-seeking and advice utilisation decisions, when people learn about messenger characteristics that are important to their social identity.

As in Chapter 3, there were some discrepancies here between the results of the behavioural analysis and computational modelling. In particular, an rmANOVA showed that participants' bets on which sources would agree with their answers on personal value questions were influenced by how accurate the sources were on the blup task, whereas a model comparison indicated that unbiased RL and Bayesian models provided a better fit to this betting data than comparable models that assumed beliefs about source accuracy on the blup trials affected how much participants learned about similarity on the value trials. As noted in the previous chapter, it is possible that our models are not correctly parameterised to capture the specific learning bias that participants are exhibiting.

Overall, the studies presented in this chapter offer two additional findings. The first is that learning about others' personal values can interfere with the ability to assess and use others' expertise in unrelated domains, suggesting that the epistemic spillover effects we observed in Chapter 2 do not only occur when political allegiances are invoked. Note, even though we did not find that participants were more influenced by similar than dissimilar sources in terms of COM scores in this chapter, it is still valid to conclude that value similarity influenced the *use* of others' expertise, since participants chose to hear more often from similar than dissimilar sources and were therefore influenced by similar sources on a greater number of occasions than by dissimilar sources. The second is that while making predictions about others' accuracy enhances expertise learning, doing so does not attenuate the influence of homophily on perceptions of expertise or whom people choose to seek information from. These findings, along with those from chapter two and three, will be discussed within the context of existing psychological theories in the final chapter of this thesis.

Chapter 5. Discussion

Much attention has recently been paid to the potential for selective attention, information seeking, and belief updating to drive polarisation between social groups (Bail et al., 2018; Leong et al., 2020; Prior, 2007), produce ‘echo chambers’ (Colleoni et al., 2014; Sunstein, 2017; Kleinberg & Lau, 2016), and exacerbate the spread of misinformation online (Del Vicario et al., 2016; Faris et al., 2017; Kahan, 2017; Lewandowsky et al., 2012). The surge of interest in this area is, in part, due to the rise in political polarisation that has occurred in recent years, particularly in the United States (Boxell et al., 2020; Campbell, 2016; Iyengar et al., 2019; McCoy et al., 2018; Doherty et al., 2019), but has likely also been spurred by the surprising (and, to many academics, unsettling) results in the 2016 US Presidential election and the UK Brexit referendum. This body of work suggests that immensely consequential societal events and social dynamics may be negatively affected by individual-level biases in social learning.

Normative models of information-seeking assume that agents act to obtain information that helps them to make better decisions and therefore has ‘instrumental utility’ (Edwards, 1965; Kobayashi & Hsu, 2019; Stigler, 1961). Knowledge of others’ expertise can aid learners in this pursuit (Harvey & Fischer, 1997; Soll & Larrick, 2009). By selectively learning from sources who possess relevant expertise, people can improve the accuracy of their beliefs (Coady, 1992; Hahn et al., 2009; Henrich & Gil-White, 2001; Madsen, 2019a), increase the rewards they receive from their actions (Biele et al., 2011; Li et al., 2011), and avoid costly losses (Olsson & Phelps, 2007). The overarching aim of this thesis was to provide a novel account for why people might fail to seek information and utilise advice from others in a manner that is consistent with normative models. Our approach can be broken down into two discrete goals: First, we sought to test whether people would seek information and listen to sources with demonstrably low task-relevant expertise (relative to others), who displayed desirable characteristics in unrelated domains. Second, we sought to contribute a mechanistic account for why people might do so.

Theoretical Implications

In Chapter 1 we found evidence to support the hypothesis that people choose to hear from sources that share their political views on non-political topics, even when they could receive information from sources with greater expertise but different political opinions. Our data also suggested that the tendency to learn from the politically like-minded is mediated by an illusory perception that politically like-minded sources are more competent on non-political tasks than those with opposing political views.

These findings are consistent with previous research examining how halo effects influence beliefs about others' characteristics. Past works on the halo effect have demonstrated that people who are perceived to possess one desirable characteristic, such as attractiveness, are expected to possess a host of other desirable characteristics, such as intelligence, trustworthiness, and happiness (Dion et al., 1972; Eagly et al., 1991; Griffin & Langlois, 2006). Our results suggest that knowledge of others' (un)desirable characteristics not only influences expectations about unrelated characteristics but also interferes with the ability to learn about unrelated characteristics from observed evidence. Consequently, irrelevant messenger characteristics can bias how people learn about and utilise others' expertise, even in the presence of diagnostic information.

We speculated that people judged politically (dis)similar messengers as (in)competent on non-political tasks, even after observing evidence suggesting that this is not the case, because they overweight information indicating that messengers with (un)desirable, yet irrelevant characteristics are (in)competent and underweight evidence to the contrary. In Chapter 3, we attempted to formalise this learning bias using computational models. Here, however, we found that desirable yet irrelevant messenger characteristics – namely, generosity and political similarity – did not influence how people learned about others' task-relevant expertise or how they choose whom to hear from. Rather, the data indicated that people seek information from sources that are most likely to possess accurate knowledge, regardless of how generous or politically aligned they are. We hypothesised from these results that differences in the experimental task and setting between the studies in Chapter 2 and those in Chapter 3 may moderate the extent to which desirable, yet irrelevant, messenger characteristics interfere with expertise learning and information-seeking decisions.

Interestingly, although knowledge of others' political similarity did not influence expertise learning, in Study 4 (Chapter 3) we did find significant effects in the opposite causal direction: competence on general knowledge quiz questions influenced how likely participants were to bet that others would share their political views and affected their ratings of how similar others were to them on political questions. One explanation for this is that political similarity was more salient than general knowledge competence. The Salient Dimensions model of the halo effect (Fisicaro & Lance, 1990) suggests that the direction of a halo effect will depend on which of two observed traits is more salient. It is possible that the quiz task was more salient than the political task in Study 4 and therefore participants' estimates of others' general knowledge expertise influenced perceptions of political similarity, but not vice-versa. It is also worth noting that that there is more subjectivity in similarity judgments than accuracy judgments. Therefore, the influence of expertise beliefs on similarity learning provides weaker evidence that people violate normative principles of learning than effects in the reverse causal direction.

The final empirical chapter of this thesis (i.e., Chapter 4) sought to test whether the findings from Chapter 2 were driven by features of the current political climate in the US, as opposed to a more general cognitive bias, and whether they could be explained by an alternative explanation: a lack of attention. The results demonstrated that epistemic spillover effects are not specific to US politics – learning about others' personal values can also interfere with the ability to assess and utilise expertise – and may occur even when people make active predictions about others' accuracy before observing outcomes (and are, therefore, more attentive to the evidence presented to them). Betting on others' accuracy in a shape categorization task did enhance expertise learning, indicating that doing so leads people to pay more attention to the social evidence presented to them than they would otherwise. However, contrary to our preregistered hypothesis, betting did not attenuate the effect of value similarity on expertise learning or the effect of expertise on similarity learning, suggesting that the tendency to perceive like-minded others as more competent is not due to a lack of attention. Nonetheless, the findings from Chapter 4 convincingly show that contextual factors can affect the degree to which people learn from diagnostic information relating to others' traits.

Whether epistemic spillovers are costly or not will depend on how misperceptions of expertise affect individuals' choices. If people decide to seek information and take advice from like-minded sources instead of those with more task-relevant expertise, as when a participant chooses to hear from an inaccurate/similar source rather than an accurate/dissimilar source in our studies, then they will suffer a cost in terms of the expected utility of their decisions. On the other hand, if people only choose to listen to those who share their views when choosing between equally competent sources, as when a participant chooses to hear from an [in]accurate/similar source rather than an [in]accurate/dissimilar source in our studies, then a preference to hear from like-minded sources will not reduce the quality of their decisions. In Study 1 and Study 5, we observed clear evidence that participants chose to hear from less accurate sources who shared their beliefs rather than more accurate sources who did not, suggesting that at least in some circumstances epistemic spillovers can lead to suboptimal social learning decisions.

Yet, even in cases where people make suboptimal decisions, it is hard to say that these decisions are not rational, as they may reflect the optimal use of the brains' limited computational resources and time (Gershman et al., 2015; Gigerenzer, 2008; Griffiths et al., 2015; Lewis et al., 2014; Lieder & Griffiths, 2020). For example, if the cognitive costs of tracking others' expertise are greater than the expected benefits that could be accrued from doing so appropriately, then individuals may rely on heuristic mechanisms to make judgements and decisions. It is notable, then, that when the experimental conditions were designed so as to facilitate expertise learning, we observed little to no evidence that participants made *costly* social learning decisions. Even in the betting condition of Study 6, where our results showed that participants did prefer to hear from like-minded sources, we did not find evidence to suggest that they systematically chose to learn from inaccurate yet similar sources over accurate yet dissimilar sources (see Figure 29, right-hand panel). Thus, it is possible that the degree to which irrelevant messenger characteristics bias expertise learning, and thus information-seeking and advice-utilisation decisions, may reflect optimal trade-offs between the benefits of increased accuracy and the costs of performing resource intensive cognitive operations, in accordance with a resource-rational account of cognition (Gershman et al., 2015; Lieder & Griffiths, 2020).

Applied Implications

Previous research has demonstrated that expertise judgements do not only affect whom people choose to go to for information and advice but also influence a host of other consequential decisions. For example, a large body of research on first impressions has demonstrated that intuitive judgments of competence are predictive of which candidates receive more votes in elections (Antonakis & Dalgas, 2009; Ballew & Todorov, 2007; Lawson et al., 2010; Olivola & Todorov, 2010; Sussman et al., 2013; Todorov et al., 2005) and, in business, which job applicants get hired and negotiate better salaries (Pfann et al., 2000; Rule and Ambady, 2008; Rule & Ambady, 2009). Our findings suggest that evaluators hold illusory perceptions of competence based on epistemic-based factors too and it is possible that these affect many social decisions in domains such as politics and business.

While it is alarming that beliefs about irrelevant messenger characteristics can influence how people learn about and utilise others' task-relevant expertise, our finding that it is possible to influence the degree to which people attend to diagnostic information may provide some comfort to those alarmed by the tendency for selective attention, information seeking, and belief updating to drive negative societal outcomes. This latter finding indicates that informed interventions can help to stymie messenger biases, not by reducing halo effects per se but rather by facilitating learning of relevant messenger characteristics.

Many studies focus on reducing judgement and decision-making biases (e.g., Axt et al., 2019; Stone & Moskowitz, 2011). Yet increasing individuals' ability to learn from diagnostic information is also a viable route to improving decision-making (Axt & Lai, 2019). For example, in situations where a person can choose to seek information from either an accurate or an inaccurate source, both of whom share their political views, reducing bias in favour of similar others will not help them to make a better decision, whereas interventions that improve expertise learning will. Halo effects and epistemic spillovers will only lead to inaccurate judgements and decisions when individuals chose to hear from sources with relatively low expertise rather those with greater expertise. This can be remedied by either reducing bias or increasing expertise learning. The results of Study 6 suggest that it may be easier to do the latter than the former.

Of course, in some instances, such as when trying to create a diverse workforce, it may be necessary to reduce messenger biases, even if accuracy assessments are so well refined that those with relatively low expertise are never consulted or chosen. We know from previous research that job applicants are more likely to receive a call-back if their CV suggests that they share the employer's political affiliation than if it signals they support an opposing political party (Gift & Gift, 2015). There is also existing evidence to suggest that teams that are cognitively diverse – that is, have large intra-team differences in perspective or information processing styles – are better at complex problem-solving tasks (Reynolds & Lewis, 2017; Syed, 2019). Now, consider a case where two equally qualified and competent candidates are applying for a job. One shares the hiring manager's political views, while the other does not. A bias that leads the similar candidate to be chosen will reduce the belief diversity within the organisation and may consequently cause the team to be less effective than if the dissimilar candidate were hired. Improving expertise learning would not help to better the hiring manager's decision-making in this scenario but reducing bias would.

Efforts to improve social decisions need to therefore consider how interventions impact judgemental bias, accuracy, or both. Our findings suggest that facilitating expertise learning, for example by asking people to predict whether messengers will possess accurate knowledge before observing diagnostic evidence, will help people to make better social learning decisions without reducing bias.

Directions For Future Research

In Chapter 4, we demonstrated that betting on how others will answer questions, before observing those answers, improved learning from outcomes even though it did not reduce bias. Future research is needed to test whether making predictions about the accuracy of others' assertions is an effective method for improving expertise learning in real-world settings. For example, it would be of interest to test whether betting on whether a political candidate will answer a fact-based question accurately reduces partisan viewers' propensity to process subsequent evidence in a biased manner (Cohen, 2003).

In our studies (Studies 3, 4, 6), participants bet on how sources would answer questions before seeing the sources' answers and before observing whether those answers were correct. It remains untested whether betting would still increase expertise learning if it occurred after the source's answer had been observed but before the outcome was revealed. Indeed, it is possible that the congruence between a source's answer and a learner's prior beliefs may nullify the impact of betting on expertise learning, because the learner must not only update their beliefs about another's expertise but also their own. This is an important question because prediction-based interventions may be ineffective if they are implemented after a learner has received information (e.g., after a viewer has heard a politician's answer but before finding out whether that answer was factually accurate).

Future research could also explore the moderators of epistemic spillovers. In Chapter 3 we found no effect of source generosity (Study 3) or political similarity (Study 4) on expertise learning, information seeking, or advice utilisation. We hypothesised that these null effects were driven by the inclusion of pre-evidence betting in these studies, but the results of Study 6 do not support this notion. It is possible that differences between Study 4 and Study 6, such as the inclusion of a similarity learning test, the stimuli, the setting, the sample, or the existence of a points-based reward system, moderate the impact of epistemic similarities, but at present it is unclear which, if any, of these features do so or why.

Concluding Remarks

Evolutionary and decision theorists have puzzled over why people selectively seek out and believe information from sources that they find congenial rather than those with the most expertise (e.g., Henrich & Broesch, 2011; Sunstein, 2017). This thesis provides a novel account of why people might choose to seek information from like-minded sources. In four out of the six studies conducted, we found that people judge those who share their political beliefs or personal values as more competent at unrelated tasks than those with differing views, even when they are presented with diagnostic information pertaining to those others' task-relevant expertise. This suggests that knowledge of others' beliefs does not only influence perceptions of expertise but also how people learn about others' expertise from observable social evidence. Consequently, inaccurate beliefs about others' expertise can be maintained in the face of reality and people may choose to learn from relatively inaccurate sources when they should be able to make wiser social learning decisions. There is a growing concern that this behaviour is driving undesirable real-world behaviours, including the spread of fake news (Faris et al., 2017; Friggeri, 2014; Kahne & Bowyer, 2017; Traberg & van der Linden, 2022), conspiracy theories (Del Vicario et al., 2016), and polarisation (Bail et al., 2018; Druckman, 2013; Prior, 2007).

This thesis also highlights that, under certain conditions, people will choose to hear from accurate sources regardless of how similar their political beliefs are. Moreover, it demonstrates that contextual factors can affect the degree to which people learn from diagnostic information relating to others' traits. In particular, when people are asked to make predictions about others' accuracy, they show enhanced expertise learning after observing outcomes. The fact that it is possible to increase attention to social evidence suggests that informed interventions can improve people's ability to judge who will provide them with useful information. This latter finding provides some hope to those who are alarmed by the tendency for selective attention, information seeking, and belief updating to drive negative societal outcomes. For those wanting to reduce harmful effects of messenger biases, it may be comforting to know that change is possible.

References

Aaker, J., Vohs, K. D., & Mogilner, C. (2010). Nonprofits are seen as warm and for-profits as competent: Firm stereotypes matter. *Journal of Consumer Research*, *37*, 224-237.

Abele, A. E., & Wojciszke, B. (2007). Agency and communion from the perspective of self versus others. *Journal of Personality and Social Psychology, 93*, 751-763.

Abele, A. E., & Wojciszke, B. (2014). Communal and agentic content in social cognition: A Dual Perspective Model. In J. M. Olson & M.P. Zanna (Eds.), *Advances in Experimental Social Psychology* (pp. 195- 255). San Diego: Academic.

Abikoff, H., Courtney, M., Pelham, W. E., & Koplewicz, H. S. (1993). Teachers' ratings of disruptive behaviors: The influence of halo effects. *Journal of Abnormal Child Psychology*, 21, 519-533.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716-723.

Alister, M., Vickers-Jones, R., Sewell, D. K., & Ballard, T. (2021). How Do We Choose Our Giants? Perceptions of Replicability in Psychological Science. *Advances in Methods and Practices in Psychological Science*, *4*, 1-21.

Ambady, N., LaPlante, D., Nguyen, T., Rosenthal, R., Chaumeton, N., & Levinson, W. (2002). Surgeons' tone of voice: a clue to malpractice history. *Surgery*, *132*, 5-9.

Ames, D. L., & Fiske, S. T. (2013). Outcome dependency alters the neural substrates of impression formation. *NeuroImage*, *83*, 599-608.

Ammons, R. B. (1956). Effects of knowledge of performance: A survey and tentative theoretical formulation. *Journal of General Psychology, 54*, 279-299.

Anderson, J. R. (1990). The adaptive character of thought. Erlbaum.

Antonakis, J., & Dalgas, O. (2009). Predicting elections: Child's play! *Science, 323*, 1183-1183.

Anvari, F., & Lakens, D. (2018). The replicability crisis and public trust in psychological science. *Comprehensive Results in Social Psychology*, *3*, 266-286.

Apesteguia, J., Huck, S., & Oechssler, J. (2007). Imitation—theory and experimental evidence. *Journal of Economic Theory*, *136*, 217-235.

Aquino, K., & Reed, A. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology, 83*, 1423-1440.

Asch, S. E. (1946). Forming impressions of personality. *Journal of Abnormal and Social Psychology, 41*, 258-290.

Axt, J. R., & Lai, C. K. (2019). Reducing discrimination: A bias versus noise perspective. *Journal of Personality and Social Psychology*, *117*, 26-49.

Axt, J. R., Casola, G., & Nosek, B. A. (2019). Reducing social judgment biases may require identifying the potential source of bias. *Personality and Social Psychology Bulletin*, *45*, 1232-1251.

Baier, A. (1986). Trust and antitrust. E*thics*, *96*, 231-260.

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. F., ... & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, *115*, 9216-9221.

Bakan, D. (1966). *The duality of human existence: Isolation and communion in Western man*. Beacon Press.

Ballew, C. C., & Todorov, A. (2007). Predicting political elections from rapid and unreflective face judgments. *Proceedings of the National Academy of Sciences*, *104*, 17948-17953.

Bandura, A. (1977). *Social learning theory*. Prentice Hall.

Baron, J. (1996). Norm-endorsement utilitarianism and the nature of utility. *Economics and Philosophy, 12*, 165–82.

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, *117*, 497-529.

Bazerman, M. H., Loewenstein, G. F., & White, S. B. (1992). Reversals of preference in allocation decisions: Judging an alternative versus choosing among alternatives. *Administrative Science Quarterly, 37*, 220-240.

Behrens, T. E. J., Hunt, L. T., Woolrich, M. W., & Rushworth, M. F. S. (2008). Associative learning of social value. *Nature, 456*, 245–249.

Behrens, T. E., Hunt, L. T., & Rushworth, M. F. (2009). The computation of social behavior. *Science*, *324*, 1160-1164.

Behrens, T. E., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*, 1214-1221.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, *10*, 122-142.

Berger, J., Cohen, B. P., & Zelditch Jr, M. (1972). Status characteristics and social interaction. *American Sociological Review*, *37*, 241-255.

Berger, J., Rosenholtz, S. J., & Zelditch Jr, M. (1980). Status organizing processes. *Annual Review of Sociology, 6*, 479-508.

Berman, J. S., & Kenny, D. A. (1976). Correlational bias in observer ratings. *Journal of Personality and Social Psychology, 34*, 263–273.

Biele, G., Rieskamp, J., Krugel, L. K., & Heekeren, H. R. (2011). The neural basis of following advice. *PLoS biology*, *9*, e1001089.

Birch, S. A., Akmal, N., & Frampton, K. L. (2010). Two-year-olds are vigilant of others' non-verbal cues to credibility. *Developmental Science*, *13*, 363-369.

Birch, S. A., Vauthier, S. A., & Bloom, P. (2008). Three-and four-year-olds spontaneously use others' past performance to guide their learning. *Cognition, 107*, 1018-1034.

Block, K., Gonzalez, A. M., Schmader, T., & Baron, A. S. (2018). Early Gender Differences in Core Values Predict Anticipated Family Versus Career Orientation. *Psychological Science*, *29*, 1540–1547.

Boer, D., & Fischer, R. (2013). How and when do personal values guide our attitudes and sociality? Explaining cross-cultural variability in attitude-value linkages. *Psychological Bulletin*, *139*, 1113–1147.

Bonaccio, S., & Dalal, R. S. (2006). Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational Behavior and Human Decision Processes, 101*, 127–151.

Boorman, E. D., Behrens, T. E., & Rushworth, M. F. (2011). Counterfactual choice and learning in a neural network centered on human lateral frontopolar cortex. *PLoS biology*, *9*, e1001093.

Boorman, E. D., O'Doherty, J. P., Adolphs, R., & Rangel, A. (2013). The behavioral and neural mechanisms underlying the tracking of expertise. *Neuron*, *80*, 1558-1571.

Boseovski, J. J. (2010). Evidence for "rose-colored glasses": An examination of the positivity bias in young children's personality judgments. *Child Development Perspectives, 4,* 212–218.

Bossaerts, P., & Murawski, C. (2017). Computational complexity and human decision-making. *Trends in Cognitive Sciences*, *21*, 917-929.

Bovens, L., & Hartmann, S. (2003). *Bayesian Epistemology.* Oxford University Press.

Boxell, L., Gentzkow, M., & Shapiro, J. M. (2020). *Cross-country trends in affective polarization*. National Bureau of Economic Research.

Boyd, R., & Richerson, P. J. (1985). *Culture and the evolutionary process*. University of Chicago Press.

Boyd, R., Richerson, P. J., & Henrich, J. (2011). The cultural niche: Why social learning is essential for human adaptation. *Proceedings of the National Academy of Sciences*, *108*, 10918-10925.

Brainard, D. H., & Freeman, W. T. (1997). Bayesian color constancy. *JOSA A*, *14*, 1393-1411.

Brainard, D. H., Longère, P., Delahunt, P. B., Freeman, W. T., Kraft, J. M., & Xiao, B. (2006). Bayesian model of human color constancy. *Journal of Vision*, *6*, 10-10.

Brewer, M. B. (1999). The psychology of prejudice: Ingroup love or outgroup hate? *Journal of Social Issues*, *55*, 429-444.

Brewer, M. B., & Caporael, L. R. (2006). An evolutionary perspective on social identity: Revisiting groups. In M. Schaller, J. A. Simpson, & D. T. Kenrick (Eds.), *Evolution and Social Psychology* (pp. 143–161). Psychosocial Press.

Brinol, P., & Petty, R. E. (2009). Source factors in persuasion: A self-validation approach. *European Review of Social Psychology*, *20*, 49-96.

Brock, T. C. (1965). Communicator-recipient similarity and decision change. *Journal of Personality and Social Psychology, 1*, 650–654.

Brooks, A. W., Gino, F., & Schweitzer, M. E. (2015). Smart people ask for (my) advice: Seeking advice boosts perceptions of competence. *Management Science*, *61*, 1421-1435.

Burger, J. M., Messian, N., Patel, S., Del Prado, A., & Anderson, C. (2004). What a coincidence! The effects of incidental similarity on compliance. *Personality and Social Psychology Bulletin*, *30*, 35-43.

Burke, C. J., Tobler, P. N., Baddeley, M., & Schultz, W. (2010). Neural mechanisms of observational learning. *Proceedings of the National Academy of Sciences*, *107*, 14431-14436.

Byrne, D. (1969). Attitudes and attraction. In *Advances in experimental social psychology* (Vol. 4, pp. 35-89). Academic Press.

Byrne, R., & Whiten, A. (1988). *Machiavellian intelligence: Social expertise and the evolution of intellect in monkeys, apes, and humans.* Oxford University Press.

Camerer, C., & Weigelt, K. (1998). Experimental tests of a sequential equilibrium reputation model. *Econometrica*, *56*, 1-36.

Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin*, *54*, 297-312.

Campbell, J. E. (2016). *Polarized: Making sense of a divided America*. Princeton University Press.

Caspi, A., & Herbener, E. S. (1993). Marital assortment and phenotypic convergence: Longitudinal evidence. *Social Biology, 40*, 48–60.

Chaiken, S., & Maheswaran, D. (1994). Heuristic processing can bias systematic processing: Effects of source credibility, argument ambiguity, and task importance on attitude judgment. *Journal of Personality and Social Psychology, 66*, 460–473.

Chandler, J., Mueller, P., & Paolacci, G. (2014). Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behavior Research Methods*, *46*, 112-130.

Chang, L. J., Doll, B. B., van't Wout, M., Frank, M. J., & Sanfey, A. G. (2010). Seeing is believing: Trustworthiness as a dynamic belief. *Cognitive Psychology*, *61*, 87-105.

Chater, N., & Manning, C. D. (2006). Probabilistic models of language processing and acquisition. *Trends in Cognitive Sciences*, *10*, 335-344.

Chater, N., & Oaksford, M. (1999). The probability heuristics model of syllogistic reasoning. *Cognitive Psychology*, *38*, 191-258.

Chater, N., Tenenbaum, J. B., & Yuille, A. (2006). Probabilistic models of cognition: Conceptual foundations. *Trends in Cognitive Sciences*, *10*, 287-291.

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological review, 104*, 367–405.

Cialdini, R. B. (1984). *Influence: The new psychology of modern persuasion*. Morrow.

Cialdini, R. B. (2001). The science of persuasion. *Scientific American*, *284*, 76-81.

Cialdini, R. B., & Trost, M. R. (1998). Social influence: Social norms, conformity and compliance. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (pp. 151–192). McGraw-Hill.

Clifford, S., & Jerit, J. (2014). Is there a cost to convenience? An experimental comparison of data quality in laboratory and online studies. *Journal of Experimental Political Science*, *1*, 120-131.

Coady, C. A. J. (1992). *Testimony: A philosophical study*. Oxford University Press.

Cohen, G. L. (2003). Party over policy: The dominating impact of group influence on political beliefs. *Journal of Personality and Social Psychology*, *85*, 808-822.

Colleoni, E., Rozza, A., & Arvidsson, A. (2014). Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data. *Journal of Communication*, *64*, 317-332.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *8*, 240-247.

Corriveau, K. H., Harris, P. L., Meins, E., Fernyhough, C., Arnott, B., Elliott, L., ... & De Rosnay, M. (2009). Young children's trust in their mother's claims: Longitudinal links with attachment security in infancy. *Child Development*, *80*, 750-761.

Corriveau, K., & Harris, P. L. (2009). Choosing your informant: Weighing familiarity and recent accuracy. *Developmental Science*, *12*, 426-437.

Coussi-Korbel, S., & Fragaszy, D. M. (1995). On the relation between social dynamics and social learning. *Animal Behaviour*, *50*, 1441-1453.

Cowan, M. L., & Little, A. C. (2013). The effects of relationship context and modality on ratings of funniness. *Personality and Individual Differences*, *54*, 496-500.

Cox, R. (1946). Probability, frequency, and reasonable expectation. *American Journal of Physics, 14,* 1-13.

Croft, A., Schmader, T., Block, K., & Baron, A. S. (2014). The second shift reflected in the second generation: Do parents' gender roles at home predict children's aspirations? *Psychological Science, 25*, 1418–1428.

Csibra, G., & Gergely, G. (2011). Natural pedagogy as evolutionary adaptation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*, 1149-1157.

Cuddy, A. J., Fiske, S. T., & Glick, P. (2008). Warmth and competence as universal dimensions of social perception: The stereotype content model and the BIAS map. *Advances in Experimental Social Psychology*, *40*, 61-149.

Curtice, J. (2018). The emotional legacy of Brexit: How Britain has become a country of 'Remainers' and 'Leavers'. London: NatCen Social Research. Retrieved from: https://ukandeu.ac.uk/research-papers/the-emotional-legacy-of-brexit-how-britainhas-become-a-country-of-remainers-and-leavers/

Deaner, R. O., Khera, A. V., & Platt, M. L. (2005). Monkeys pay per view: adaptive valuation of social images by rhesus macaques. *Current Biology*, *15*, 543-548.

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., & Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, *113*, 554-559.

Delgado, M. R., Frank, R. H., & Phelps, E. A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nature Neuroscience*, *8*, 1611-1618.

Dennis, I. (2007). Halo effects in grading student projects. *Journal of Applied Psychology, 92*, 1169–1176.

Deska, J. C., Lloyd, E. P., & Hugenberg, K. (2018). Facing humanness: Facial width-to-height ratio predicts ascriptions of humanity. *Journal of Personality and Social Psychology, 114*, 75–94.

Diaconescu, A. O., Mathys, C., Weber, L. A., Daunizeau, J., Kasper, L., Lomakina, E. I., ... & Stephan, K. E. (2014). Inferring on the intentions of others by hierarchical Bayesian learning. *PLoS Computational Biology*, *10*, e1003810.

Diaconescu, A. O., Mathys, C., Weber, L. A., Kasper, L., Mauer, J., & Stephan, K. E. (2017). Hierarchical prediction errors in midbrain and septum during social learning. *Social Cognitive and Affective Neuroscience*, *12*, 618-634.

Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology*, *24*, 285-290.

Doherty, C., Kiley, J., & Asheer, N. (2019). *Partisan Antipathy: More Intense, More Personal*. Pew Research Center. Retrieved from https://www.pewresearch.org/politics/2019/10/10/partisan-antipathy-more-intense-more-personal/

Dolan, P., Hallsworth, M., Halpern, D., King, D., Metcalfe, R., & Vlaev, I. (2012). Influencing behaviour: The MINDSPACE way. *Journal of Economic Psychology*, *33*, 264-277.

Dovidio, J. F., & Gaertner, S. L. (2010). Intergroup bias. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (pp. 1084–1121). John Wiley & Sons, Inc.

Doya, K., Ishii, S., Pouget, A., & Rao, R. P. (2007). *Bayesian brain: Probabilistic approaches to neural coding*. MIT press.

Draca, M., & Schwarz, C. (2020). How polarized are citizens? Measuring ideology from the ground-up. Available at SSRN: https://ssrn.com/abstract=3154431

Druckman, J. N., Peterson, E., & Slothuus, R. (2013). How elite partisan polarization affects public opinion formation. *American Political Science Review*, *107*, 57-79.

Dunbar, R. I. M. (2004). Gossip in evolutionary perspective. *Review of General Psychology, 8,* 100–110.

Dunbar, R. I., & Shultz, S. (2007). Evolution in the social brain. *Science*, *317*, 1344-1347.

Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). What is beautiful is good, but…: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin*, *110*, 109-128.

Edelson, M. G., Dudai, Y., Dolan, R. J., & Sharot, T. (2014). Brain Substrates of Recovery from Misleading Influence. *Journal of Neuroscience, 34*, 7744-7753.

Edwards, W. (1954). The theory of decision making. *Psychological Bulletin*, *51*, 380-417.

Edwards, W. (1965). Optimal strategies for seeking information: Models for statistics, choice reaction times, and human information processing. *Journal of Mathematical Psychology*, *2*, 312-329.

Eriksson, K., & Strimling, P. (2009). Biases for acquiring information individually rather than socially. *Journal of Evolutionary Psychology*, *7*, 309-329.

Faraji-Rad, A., Samuelsen, B. M., & Warlop, L. (2015). On the persuasiveness of similar others: The role of mentalizing and the feeling of certainty. *Journal of Consumer Research*, *42*, 458-471.

Faraji-Rad, A., Warlop, L., & Samuelsen, B. (2012). When the Message "Feels Right": When and How Does Source Similarity Enhance Message Persuasiveness? *Advances in Consumer Research, 40,* 682-683.

Faris, R., Roberts, H., Etling, B., Bourassa, N., Zuckerman, E., & Benkler, Y. (2017). Partisanship, propaganda, and disinformation: Online media and the 2016 US presidential election. *Berkman Klein Center Research Publication*, *6*. Available at SSRN: https://ssrn.com/abstract=3019414

Faul, F., Erdfelder, E., Lang, A., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39,* 175–191.

Feinberg, M., Willer, R., & Schultz, M. (2014). Gossip and ostracism promote cooperation in groups. *Psychological Science, 25*, 656–664.

Feinberg, M., Willer, R., Stellar, J., & Keltner, D. (2012). The virtues of gossip: Reputational information sharing as prosocial behavior. *Journal of Personality and Social Psychology, 102*, 1015–1030.

Fiser, J., Berkes, P., Orbán, G., & Lengyel, M. (2010). Statistically optimal perception and learning: from behavior to neural representations. *Trends in Cognitive Sciences*, *14*, 119-130.

Fisicaro, S. A., & Lance, C. E. (1990). Implications of three causal models for the measurement of halo error. *Applied Psychological Measurement*, 14, 419-429.

Forgas, J. P. (2011). She just doesn't look like a philosopher…? Affective influences on the halo effect in impression formation. *European Journal of Social Psychology*, *41*, 812-817.

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*, 998-998.

Freeman, J. B., Schiller, D., Rule, N. O., & Ambady, N. (2010). The neural origins of superficial and individuated judgments about ingroup and outgroup members. *Human Brain Mapping*, *31*, 150-159.

Friggeri, A., Adamic, L. A., Eckles, D., & Cheng, J. (2014). Rumor Cascades. In *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media.* Retrieved from https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/viewFile/8122/8110

Frith, C. D., & Frith, U. (2012). Mechanisms of social cognition. *Annual Review of Psychology*, *63*, 287-313.

Frone, M. R., Adams, J., Rice, R. W., & Instone-Noonan, D. (1986). Halo error: A field study comparison of self- and subordinate evaluations of leadership process and leader effectiveness. *Personality and Social Psychology Bulletin, 12*, 454-461.

Galef Jr, B. G., Dudley, K. E., & Whiskin, E. E. (2008). Social learning of food preferences in 'dissatisfied' and 'uncertain' Norway rats. *Animal Behaviour*, *75*, 631-637.

Galizzi, M. M., & Navarro-Martinez, D. (2019). On the external validity of social preference games: a systematic lab-field study. *Management Science*, *65*, 976-1002.

Garimella, K., Morales, G. D. F., Gionis, A., & Mathioudakis, M. (2018). Political Discourse on Social Media: Echo Chambers, Gatekeepers, and the Price of Bipartisanship. *arXiv preprint,* arXiv:1801.01665.

Gelman, S. A., (1988). The development of induction within natural kind and artifact categories. *Cognitive Psychology, 20*, 65–95.

Gentzkow, M., & Shapiro, J. M. (2011). Ideological segregation online and offline. *The Quarterly Journal of Economics*, *126*, 1799-1839.

Gershman, S. J., Horvitz, E. J., & Tenenbaum, J. B. (2015). Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. *Science*, *349*, 273-278.

Gift, K., & Gift, T. (2015). Does politics influence hiring? Evidence from a randomized experiment. *Political Behavior*, *37*, 653-675.

Gigerenzer, G. (2008). Why heuristics work. *Perspectives on Psychological Science*, *3*, 20-29.

Gilovich, T. (1987). Secondhand information and social judgment. *Journal of Experimental Social Psychology, 23*, 59–74.

Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge University Press.

Gino, F., Brooks, A. W., & Schweitzer, M. E. (2012). Anxiety, advice, and the ability to discern: Feeling anxious motivates individuals to seek and use advice. *Journal of Personality and Social Psychology, 102*, 497-512.

Gino, F., Shang, J., & Croson, R. (2009). The impact of information from similar or different advisors on judgment. *Organizational Behavior and Human Decision Processes*, *108*, 287-302.

Giraldeau, L. A., Valone, T. J., & Templeton, J. J. (2002). Potential disadvantages of using socially acquired information. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *357*, 1559-1566.

Good, I. J. (1950). *Probability and the weighing of evidence.* Hafners.

Goodman, N. D., Tenenbaum, J. B., Feldman, J., & Griffiths, T. L. (2008). A rational analysis of rule-based concept learning. *Cognitive Science*, *32*, 108-154.

Gräf, M., & Unkelbach, C. (2016). Halo effects in trait assessment depend on information valence: Why being honest makes you industrious, but lying does not make you lazy. *Personality and Social Psychology Bulletin*, *42*, 290-310.

Gräf, M., & Unkelbach, C. (2018). Halo effects from agency behaviors and communion behaviors depend on social context: Why technicians benefit more from showing tidiness than nurses do. *European Journal of Social Psychology*, *48*, 701-717.

Green, P., & MacLeod, C. J. (2016). SIMR: an R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*, 493-498.

Griffin, A. M., & Langlois, J. H. (2006). Stereotype directionality and attractiveness stereotyping: Is beauty good or is ugly bad? *Social Cognition*, *24*, 187-206.

Griffiths, T. L., & Tenenbaum, J. B. (2006). Optimal predictions in everyday cognition. *Psychological Science*, *17*, 767-773.

Griffiths, T. L., & Tenenbaum, J. B. (2009). Theory-based causal induction. *Psychological Review, 116*, 661-716.

Griffiths, T. L., Lieder, F., & Goodman, N. D. (2015). Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Topics in cognitive science, 7*, 217-229.

Hackel, L. M., & Amodio, D. M. (2018). Computational neuroscience approaches to social cognition. *Current Opinion in Psychology*, *24*, 92-97.

Hackel, L. M., Doll, B. B., & Amodio, D. M. (2015). Instrumental learning of traits versus rewards: dissociable neural correlates and effects on choice. *Nature Neuroscience*, *18*, 1233-1235.

Hackel, L. M., Mende-Siedlecki, P., & Amodio, D. M. (2020). Reinforcement learning in social interaction: The distinguishing role of trait inference. *Journal of Experimental Social Psychology*, *88*, 103948.

Hahn, U., & Harris, A. J. (2014). What does it mean to be biased: Motivated reasoning and rationality. In *Psychology of Learning and Motivation* (Vol. 61, pp. 41-102). Academic Press.

Hahn, U., & Oaksford, M. (2007). The rationality of informal argumentation: a Bayesian approach to reasoning fallacies. *Psychological Review*, *114*, 704–732.

Hahn, U., Harris, A. J., & Corner, A. (2009). Argument content and argument source: An exploration. *Informal Logic*, *29*, 337-367.

Hahn, U., Merdes, C., & von Sydow, M. (2018). How good is your evidence and how would you know? *Topics in Cognitive Science*, *10*, 660-678.

Hahn, U., Oaksford, M., & Harris, A. J. (2013). Testimony and argument: A Bayesian perspective. In *Bayesian argumentation* (pp. 15-38). Springer, Dordrecht.

Hampton, A. N., Bossaerts, P., & O'Doherty, J. P. (2008). Neural correlates of mentalizing-related computations during strategic interactions in humans. *Proceedings of the National Academy of Sciences*, *105*, 6741-6746.

Hannak, A., Sapiezynski, P., Molavi Kakhki, A., Krishnamurthy, B., Lazer, D., Mislove, A., & Wilson, C. (2013, May). Measuring personalization of web search. In *Proceedings of the 22nd international conference on World Wide Web* (pp. 527-538).

Harari, H., & McDavid, J. W. (1973). Name stereotypes and teachers' expectations. *Journal of Educational Psychology, 65*, 222–225.

Hardin, R. (1993). The street-level epistemology of trust. *Politics & Society, 21,* 505–529.

Harinck, F., & Van Kleef, G. A. (2012). Be hard on the interests and soft on the values: Conflict issue moderates the effects of anger in negotiations. *British Journal of Social Psychology*, *51*, 741-752.

Harris, A. J., Hahn, U., Madsen, J. K., & Hsu, A. S. (2016). The appeal to expert opinion: Quantitative support for a Bayesian network approach. *Cognitive Science*, *40*, 1496-1533.

Harris, P. L., & Corriveau, K. H. (2011). Young children's selective trust in informants. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*, 1179-1187.

Harvey, N., & Fischer, I. (1997). Taking advice: Accepting help, improving judgment, and sharing responsibility. *Organizational Behavior and Human Decision Processes*, *70*, 117-133.

Hastie, R., & Kumar, P. A. (1979). Person memory: Personality traits as organizing principles in memory for behaviors. *Journal of Personality and Social Psychology, 37*, 25–38.

Hauser, D. J., & Schwarz, N. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods*, *48*, 400-407.

Heider, F. (1958). *The psychology of interpersonal relations*. Wiley.

Hendrick, C., & Costantini, A. F. (1970). Effects of varying trait inconsistency and response requirements on the primacy effect in impression formation. *Journal of Personality and Social Psychology, 15*, 158–164.

Henrich, J., & Broesch, J. (2011). On the nature of cultural transmission networks: evidence from Fijian villages for adaptive learning biases. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*, 1139-1148.

Henrich, J., & Gil-White, F. J. (2001). The evolution of prestige: Freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. *Evolution and Human Behavior, 22*, 165-196.

Henrich, J., & Henrich, N. (2010). The evolution of cultural adaptations: Fijian food taboos protect against dangerous marine toxins. *Proceedings of the Royal Society B: Biological Sciences*, *277*, 3715-3724.

Henrich, J., & McElreath, R. (2003). The evolution of cultural evolution. *Evolutionary Anthropology: Issues, News, and Reviews: Issues, News, and Reviews*, *12*, 123-135.

Heyes, C. (2012). What's social about social learning? *Journal of Comparative Psychology*, *126*, 193-202.

Heyes, C. (2016). Who knows? Metacognitive social learning strategies. *Trends in Cognitive Sciences*, *20*, 204-213.

Hobolt, S. B., Leeper, T. J., & Tilley, J. (2020). Divided by the vote: Affective polarization in the wake of the Brexit referendum. *British Journal of Political Science*, 1-18.

Hofmann, D. A., Lei, Z., & Grant, A. M. (2009). Seeking help in the shadow of doubt: the sensemaking processes underlying how nurses decide whom to ask for advice. *Journal of Applied Psychology*, *94*, 1261- 1274.

Hogan, R., & Hogan, J., (1996). *Motives, Values, Preferences Inventory Manual*. Hogan Assessment Systems.

Hogan, R., Hall, R., & Blank, E. (1972). An extension of the similarity-attraction hypothesis to the study of vocational behavior. *Journal of Counseling Psychology*, *19*, 238-240.

Holleman, G. A., Hooge, I. T., Kemner, C., & Hessels, R. S. (2020). The 'real-world approach' and its problems: A critique of the term ecological validity. *Frontiers in Psychology*, *11*, 1-12.

Holzbach, R. L. (1978). Rater bias in performance ratings: Superior, self-, and peer ratings. *Journal of Applied Psychology, 63*, 579-588.

Howe, L. C., Goyer, J. P., & Crum, A. J. (2017). Harnessing the placebo effect: Exploring the influence of physician characteristics on placebo response. *Health Psychology, 36*, 1074–1082.

Howson, C., & Urbach, P. (1996). *Scientific Reasoning: The Bayesian Approach* (2nd ed.). Open Court.

Hughes, B. L., Zaki, J., & Ambady, N. (2017). Motivation alters impression formation and related neural systems. *Social Cognitive and Affective Neuroscience*, *12*, 49-60.

Iyengar, S., & Westwood, S. J. (2015). Fear and loathing across party lines: New evidence on group polarization. *American Journal of Political Science*, *59*, 690-707.

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The origins and consequences of affective polarization in the United States. *Annual Review of Political Science*, *22*, 129-146.

Iyengar, S., Sood, G., & Lelkes, Y. (2012). Affect, not ideology: A social identity perspective on polarization. *Public Opinion Quarterly, 76*, 405–431.

Jaswal, V. K., & Malone, L. S. (2007). Turning believers into skeptics: 3-year-olds' sensitivity to cues to speaker credibility. *Journal of Cognition and Development, 8*, 263-283.

Jehn, K. A., Chadwick, C., & Thatcher, S. M. (1997). To agree or not to agree: The effects of value congruence, individual demographic dissimilarity, and conflict on workgroup outcomes. *International Journal of Conflict Management*, 8, 287-305.

Jones, E. E. (1990). *Interpersonal perception.* Freeman.

Judd, C. M., James-Hawkins, L., Yzerbyt, V. Y., & Kashima, Y. (2005). Fundamental dimensions of social judgment: Understanding the relations between judgments of competence and warmth. *Journal of Personality and Social Psychology, 89*, 899-913.

Kahan, D. M. (2016). The politically motivated reasoning paradigm, part 1: What politically motivated reasoning is and how to measure it. In R. A. Scott, S. M. Kosslyn, & M. C. Buchmann (Eds.), *Emerging trends in the social and behavioral sciences: an interdisciplinary, searchable, and linkable resource* (pp. 1–16). Wiley.

Kahan, D. M. (2017). Misconceptions, misinformation, and the logic of identity-protective cognition. *Cultural Cognition Project Working Paper Series No. 164*. Available at SSRN: https://ssrn.com/abstract=2973067

Kahne, J., & Bowyer, B. (2017). Educating for democracy in a partisan age: Confronting the challenges of motivated reasoning and misinformation. *American Educational Research Journal*, *54*, 3-34.

Kahneman, D. & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology 3,* 430–454.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49-81). Cambridge University Press.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47,* 263-291.

Kapp, K. M. (2012). *The gamification of learning and instruction: game-based methods and strategies for training and education*. John Wiley & Sons.

Kardas, M., & O'Brien, E. (2018). Easier seen than done: Merely watching others perform can foster an illusion of skill acquisition*. Psychological Science, 29*, 521–536.

Kavaliers, M., Colwell, D. D., & Choleris, E. (2005). Kinship, familiarity and social status modulate social learning about "micropredators" (biting flies) in deer mice. *Behavioral Ecology and Sociobiology*, *58*, 60-71.

Kendal, J. R., Rendell, L., Pike, T. W., & Laland, K. N. (2009). Nine-spined sticklebacks deploy a hill-climbing social learning strategy. *Behavioral Ecology*, *20*, 238-244.

Kendal, R. L., Boogert, N. J., Rendell, L., Laland, K. N., Webster, M., & Jones, P. L. (2018). Social learning strategies: Bridge-building between fields. *Trends in Cognitive Sciences*, *22*, 651-665.

Kendal, R., Hopper, L. M., Whiten, A., Brosnan, S. F., Lambeth, S. P., Schapiro, S. J., & Hoppitt, W. (2015). Chimpanzees copy dominant and knowledgeable individuals: implications for cultural diversity. *Evolution and Human Behavior*, *36*, 65-72.

King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., & Montague, P. R. (2005). Getting to know you: reputation and trust in a two-person economic exchange. *Science*, *308*, 78-83.

Klein, N., & Epley, N. (2014). The topography of generosity: Asymmetric evaluations of prosocial actions. *Journal of Experimental Psychology: General, 143*, 2366–2379.

Kleinberg, M. S., & Lau, R. R. (2016). Candidate extremity, information environments, and affective polarization: Three experiments using dynamic process tracing. In A. Blais, J. F. Laslier, & K. Van der Straeten (Eds.), *Voting experiments* (pp. 67-87). Springer International.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences*, *27*, 712-719.

Knill, D. C., & Richards, W. (Eds.). (1996). *Perception as Bayesian inference*. Cambridge University Press.

Kobayashi, K., & Hsu, M. (2019). Common neural code for reward and information value. *Proceedings of the National Academy of Sciences*, *116*, 13061-13066.

Koch, A., Yzerbyt, V., Abele, A., Ellemers, N., & Fiske, S. T. (2021). Social evaluation: Comparing models across interpersonal, intragroup, intergroup, several-group, and many-group contexts. *Advances in Experimental Social Psychology*, *63*, 1-68.

Koenig, M. A., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development*, *76*, 1261-1277.

Koenig, M. A., Clément, F., & Harris, P. L. (2004). Trust in testimony: Children's use of true and false statements. *Psychological Science*, *15*, 694-698.

Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, *427*, 244-247.

Kouzakova, M., Ellemers, N., Harinck, F., & Scheepers, D. (2012). The implications of value conflict: How disagreement on values affects self-involvement and perceived common ground. *Personality and Social Psychology Bulletin*, *38*, 798-807.

Kuklinski, J. H., Quirk, P. J., Jerit, J., Schweider, D., & Rich, R. F. (2000). Misinformation and the currency of democratic citizenship. *The Journal of Politics, 62*, 790–816.

Kull, S., Ramsay, C., & Lewis, E. (2003). Misperceptions, the media, and the Iraq war. *Political Science Quarterly, 118*, 569–598.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*, 480-498.

Kuzmanovic, B., Rigoux, L., & Tittgemeyer, M. (2018). Influence of vmPFC on dmPFC predicts valence-guided belief formation. *Journal of Neuroscience*, *38*, 7996-8010.

Laland, K. N. (2004). Social learning strategies. *Animal Learning & Behavior*, *32*, 4-14.

Lance, C. E., LaPointe, J. A., & Fisicaro, S. A. (1994). Tests of three causal models of halo rater error. *Organizational Behavior and Human Decision Processes*, *57*, 83-96.

Landy, D., & Sigall, H. (1974). When beauty is talent: Task evaluation as a function of the performer's physical attractiveness. *Journal of Personality and Social Psychology, 29*, 299–304.

Langner, R., & Eickhoff, S. B. (2013). Sustaining attention to simple tasks: a meta-analytic review of the neural mechanisms of vigilant attention. *Psychological Bulletin*, *139*, 870-900.

Lawson, C., Lenz, G. S., Baker, A., & Myers, M. (2010). Looking like a winner: Candidate appearance and electoral success in new democracies. *World Politics*, *62*, 561-593.

Lee, K., Ashton, M. C., Pozzebon, J. A., Visser, B. A., Bourdage, J. S., & Ogunfowora, B. (2009). Similarity and assumed similarity in personality reports of well-acquainted persons. *Journal of Personality and Social Psychology, 96*, 460–472.

Lefebvre, G., Lebreton, M., Meyniel, F., Bourgeois-Gironde, S., & Palminteri, S. (2017). Behavioural and neural characterization of optimistic reinforcement learning. *Nature Human Behaviour*, *1*, 1-9.

Lefkowitz, M., Blake, R. R., & Mouton, J. S. (1955). Status factors in pedestrian violation of traffic signals. *The Journal of Abnormal and Social Psychology*, *51*, 704-706.

Leong, Y. C., & Zaki, J. (2018). Unrealistic optimism in advice taking: A computational account. *Journal of Experimental Psychology: General, 147*, 170–189.

Leong, Y. C., Chen, J., Willer, R., & Zaki, J. (2020). Conservative and liberal attitudes drive polarized neural responses to political content. *Proceedings of the National Academy of Sciences*, *117*, 27731-27739.

Levine, M., Prosser, A., Evans, D., & Reicher, S. (2005). Identity and emergency intervention: How social group membership and inclusiveness of group boundaries shape helping behavior. *Personality and Social Psychology Bulletin*, *31*, 443-453.

Levine, T. R. (2014). Truth-default theory (TDT) a theory of human deception and deception detection. *Journal of Language and Social Psychology*, *33*, 378-392.

Lewandowsky, S., Ecker, U. K., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, *13*, 106-131.

Lewis, R. L., Howes, A., & Singh, S. (2014). Computational rationality: Linking mechanism and behavior through bounded utility maximization. *Topics in Cognitive Science*, *6*, 279-311.

Li, J., Delgado, M. R., & Phelps, E. A. (2011). How instructed knowledge modulates the neural systems of reward learning. *Proceedings of the National Academy of Sciences of the United States of America, 108,* 55–60.

Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, *43*.

Lindley, D. (1994). Foundations. In G. Wright & P. Ayton (Eds.), *Subjective probability* (pp. 3-15). John Wiley & Sons.

Lönnqvist, J. E., & Itkonen, J. V. (2016). Homogeneity of personal values and personality traits in Facebook social networks. *Journal of Research in Personality*, *60*, 24-35.

Madsen, J. K. (2016). Trump supported it?! A Bayesian source credibility model applied to appeals to specific American presidential candidates’ opinions. In

Proceedings of the 38th Annual Meeting of the Cognitive Science Society (pp. 165–170).

Madsen, J. K. (2019a). Voter reasoning bias when evaluating statements from female and male political candidates. *Politics & Gender*, *15*, 310-335.

Madsen, J. K. (2019b). Source Credibility. In *The Psychology of Micro-Targeted Election Campaigns* (pp. 103-133). Palgrave Macmillan.

Maestripieri, D., Henry, A., & Nickels, N. (2017). Explaining financial and prosocial biases in favor of attractive people: Interdisciplinary perspectives from economics, social psychology, and evolutionary psychology. *Behavioral and Brain Sciences*, *40,* Article e19.

Maner, J. K., DeWall, C. N., & Gailliot, M. T. (2008). Selective attention to signs of success: Social dominance and early stage interpersonal perception. *Personality and Social Psychology Bulletin*, *34*, 488-501.

Markman, A. B. (2018). Combining the strengths of naturalistic and laboratory decision-making research to create integrative theories of choice. *Journal of Applied Research in Memory and Cognition*, *7*, 1-10.

Martin, S. & Marks, J. (2019). *Messengers: Who We Listen To, Who We Don't, and Why*. Penguin Random House.

Masip, J., Garrido, E., & Herrero, C. (2009). Heuristic versus systematic processing of information in detecting deception: Questioning the truth bias. *Psychological Reports*, *105*, 11-36.

Mathys, C., Daunizeau, J., Friston, K. J., & Stephan, K. E. (2011). A Bayesian foundation for individual learning under uncertainty. *Frontiers in Human Neuroscience*, *5*, 39.

McCornack, S. A., & Parks, M. R. (1986). Deception detection and relationship development: The other side of trust. *Annals of the International Communication Association*, *9*, 377-389.

McCoy, J., Rahman, T., & Somer, M. (2018). Polarization and the global crisis of democracy: Common patterns, dynamics, and pernicious consequences for democratic polities. *American Behavioral Scientist*, *62*, 16-42.

Mende-Siedlecki, P. (2018). Changing our minds: the neural bases of dynamic impression updating. *Current Opinion in Psychology*, *24*, 72-76.

Michelson, L., Mannarino, A., Marchione, K., Kazdin, A. E., & Costello, A. (1985). Expectancy bias in behavioral observations of therapeutic outcome: An experimental analysis of treatment and halo effects. *Behaviour Research and Therapy, 23*, 407-414.

Miller, D. T., Downs, J. S., & Prentice, D. A. (1998). Minimal conditions for the creation of a unit relationship: The social bond between birthdaymates. *European Journal of Social Psychology*, *28*, 475-481.

Morgan, T. J., Rendell, L. E., Ehn, M., Hoppitt, W., & Laland, K. N. (2012). The evolutionary basis of human social learning. *Proceedings of the Royal Society B: Biological Sciences*, *279*, 653-662.

Morrison, E. W., & Vancouver, J. B. (2000). Within-person analysis of information seeking: The effects of perceived costs and benefits. *Journal of Management*, *26*, 119-137.

Moutoussis, M., Bentall, R. P., El-Deredy, W., & Dayan, P. (2011). Bayesian modelling of Jumping-to-Conclusions bias in delusional patients. *Cognitive Neuropsychiatry*, *16*, 422-447.

Mumma, G. H. (2002). Effects of three types of potentially biasing information on symptom severity judgments for major depressive episode. *Journal of Clinical Psychology, 58*, 1327-1345.

Newell, A., Shaw, J. C., & Simon, H. A. (1958). Elements of a theory of human problem solving. *Psychological Review, 65*, 151–166.

Nisbett, R. E., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*, *35*, 250-256.

Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, *32*(2), 303-330.

Oaksford, M., & Chater, N. (2001). The probabilistic approach to human reasoning. *Trends in Cognitive Sciences, 5*, 349-357.

Olivola, C. Y., & Todorov, A. (2010). Fooled by first impressions? Reexamining the diagnostic value of appearance-based inferences. *Journal of Experimental Social Psychology*, *46*, 315-324.

Olsson, A., & Phelps, E. A. (2007). Social learning of fear. *Nature neuroscience, 10*, 1095-1102.

Olsson, A., Knapska, E., & Lindström, B. (2020). The neural and computational systems of social learning. *Nature Reviews Neuroscience*, *21*, 197-212.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, *105*, 11087-11092.

Orehek, E., Dechesne, M., Fishbach, A., Kruglanski, A. W., & Chun, W. Y. (2010). On the inferential epistemics of trait centrality in impression formation. *European Journal of Social Psychology*, *40*, 1120-1135.

Osherson, D. N., Smith, E. E., Wilkie, O., Lopez, A., & Shafir, E. (1990). Category-based induction. *Psychological Review*, *97*, 185-200.

Palmer, C. L., & Peterson, R. D. (2016). Halo effects and the attractiveness premium in perceptions of political expertise. *American Politics Research*, *44*, 353-382.

Palminteri, S., Lefebvre, G., Kilford, E. J., & Blakemore, S. J. (2017). Confirmation bias in human reinforcement learning: Evidence from counterfactual feedback processing. *PLoS Computational Biology*, *13*, e1005684.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on amazon mechanical turk. *Judgment and Decision Making*, *5*, 411-419.

Pasquini, E. S., Corriveau, K. H., Koenig, M., & Harris, P. L. (2007). Preschoolers monitor the relative accuracy of informants. *Developmental Psychology*, *43*, 1216-1226.

Pattyn, N., Neyt, X., Henderickx, D., & Soetens, E. (2008). Psychophysiological investigation of vigilance decrement: boredom or cognitive fatigue? *Physiology & Behavior*, *93*, 369-378.

Paulhus, D. L., & Morgan, K. L. (1997). Perceptions of intelligence in leaderless groups: The dynamic effects of shyness and acquaintance. *Journal of Personality and Social Psychology, 72*, 581–591.

Peterson, C. R., & Beach, L. R. (1967). Man as an intuitive statistician. *Psychological Bulletin*, *68*, 29-46.

Petty, R. E., & Cacioppo, J. T. (1984). Source Factors and the Elaboration Likelihood Model of Persuasion. *Advances in Consumer Research*, *11*, 668–672.

Pfann, G. A., Biddle, J. E., Hamermesh, D. S., & Bosman, C. M. (2000). Business success and businesses' beauty capital. *Economics Letters*, *67*, 201-207.

Pike, T. W., Kendal, J. R., Rendell, L. E., & Laland, K. N. (2010). Learning by proportional observation in a species of fish. *Behavioral Ecology*, *21*, 570-575.

Pillemer, J., Graham, E. R., & Burke, D. M. (2014). The face says it all: CEOs, gender, and predicting corporate performance. *The Leadership Quarterly*, *25*, 855-864.

Preacher, K. J. (2015). Advances in mediation analysis: A survey and synthesis of new developments. *Annual Review of Psychology*, *66*, 825–852.

Prior, M. (2007). *Post-broadcast democracy: How media choice increases inequality in political involvement and polarizes elections*. Cambridge University Press.

Pritchard, R. D., Jones, S. D., Roth, P. L., Stuebing, K. K., & Ekeberg, S. E. (1988). Effects of group feedback, goal setting, and incentives on organizational productivity. *Journal of Applied Psychology, 73*, 337-358.

Reynolds, A., & Lewis, D. (2017, March 30). Teams solve problems faster when they're more cognitively diverse. Harvard Business Review. Retrieved from https://hbr.org/2017/03/teams-solve-problems-faster-when-theyremore-cognitively-diverse

Rieucau, G., & Giraldeau, L. A. (2011). Exploring the costs and benefits of social information use: an appraisal of current experimental evidence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*, 949-957.

Rilling, J. K., & Sanfey, A. G. (2011). The neuroscience of social decision-making. *Annual Review of Psychology*, *62*, 23-48.

Rips, L. J. (1975). Inductive judgments about natural categories. *Journal of Verbal Learning and Verbal Behavior, 14,* 665–681.

Rogers, A. R. (1988). Does biology constrain culture? *American Anthropologist, 90*, 819-831.

Rogers, C. R. (1957). The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology, 21*, 95–103.

Rosenberg, S., Nelson, C., & Vivekananthan, P. S. (1968). A multidimensional approach to the structure of personality impressions. *Journal of Personality and Social Psychology, 9*, 283–294.

Rosenkrantz, R. D. (1992). The justification of induction. *Philosophy of Science, 59*, 527-539.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, *23*, 393-404.

Rule, N. O., & Ambady, N. (2008). The face of success: Inferences from chief executive officers' appearance predict company profits. *Psychological Science*, *19*, 109-111.

Rule, N. O., & Ambady, N. (2009). She’s got the look: Inferences from female chief executive officers’ faces predict their success. *Sex Roles*, *61*, 644-652.

Russell, S. J., & Subramanian, D. (1994). Provably bounded-optimal agents. *Journal of Artificial Intelligence Research*, *2*, 575-609.

Sabbagh, M. A., & Baldwin, D. A. (2001). Learning words from knowledgeable versus ignorant speakers: Links between preschoolers' theory of mind and semantic development. *Child Development*, *72*, 1054-1070.

Sanborn, A. N., & Chater, N. (2016). Bayesian brains without probabilities. *Trends in Cognitive Sciences*, *20*, 883-893.

Savage, L. J. (1954). *The Foundations of Statistics.* John Wiley & Sons.

Saxe, R. (2006). Uniquely human social cognition. *Current Opinion in Neurobiology, 16*, 235-239.

Scherer, A. M., Windschitl, P. D., & Smith, A. R. (2013). Hope to be right: Biased information seeking following arbitrary and informed predictions. *Journal of Experimental Social Psychology*, *49*, 106-112.

Schilbach, L., Eickhoff, S. B., Schultze, T., Mojzisch, A., & Vogeley, K. (2013). To you I am listening: perceived competence of advisors influences judgment and decision-making via recruitment of the amygdala. *Social Neuroscience*, *8*, 189-202.

Schlag, K. H. (1998). Why imitate, and if so, how?: A boundedly rational approach to multi-armed bandits. *Journal of Economic Theory*, *78*, 130-156.

Schneider, B. (1987). The people make the place. *Personnel psychology*, *40*, 437-453.

Schonberg, T., Fox, C. R., & Poldrack, R. A. (2011). Mind the gap: bridging economic and naturalistic risk-taking with cognitive neuroscience. *Trends in Cognitive Sciences*, *15*, 11-19.

Schrah, G. E., Dalal, R. S., & Sniezek, J. A. (2006). No decision-maker is an Island: integrating expert advice with information acquisition. *Journal of Behavioral Decision Making*, *19*, 43-60.

Schum, D. A. (1981). Sorting out the effects of witness sensitivity and response-criterion placement upon the inferential value of testimonial evidence. *Organizational Behavior and Human Performance*, *27*, 153-196.

Schwartz, J., Luce, M. F., & Ariely, D. (2011). Are consumers too trusting? The effects of relationships with expert advisers. *Journal of Marketing Research*, *48*, S163-S174.

Schwartz, S. H. (1996). Value priorities and behavior: Applying a theory of integrated value systems. In C. Seligman, J. M. Olson, & M. P. Zanna (Eds.), *The psychology of values: The Ontario symposium* (Vol. 8, pp. 1–24). Erlbaum.

Schwartz, S. H., Caprara, G. V., & Vecchione, M. (2010). Basic personal values, core political values, and voting: A longitudinal analysis. *Political Psychology, 31*, 421–452.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6,* 461-464.

Segal-Caspi, L., Roccas, S., & Sagiv, L. (2012). Don't judge a book by its cover, revisited: Perceived and reported traits and values of attractive women. *Psychological Science*, *23*, 1112-1116.

Shafto, P., Eaves, B., Navarro, D. J., & Perfors, A. (2012). Epistemic trust: Modeling children's reasoning about others' knowledge and intent. *Developmental Science*, *15*, 436-447.

Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The expected value of control: an integrative theory of anterior cingulate cortex function. *Neuron*, *79*, 217-240.

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., & Botvinick, M. M. (2017). Toward a rational and mechanistic account of mental effort. *Annual Review of Neuroscience*, *40*, 99-124.

Shepherd, S. V., Deaner, R. O., & Platt, M. L. (2006). Social status gates social attention in monkeys. *Current Biology*, *4*, 138-147.

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, *69*, 99-118.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review, 63*, 129–138.

Simon, H. A. (1979). Rational decision making in business organizations. *The American Economic Review*, *69*, 493-513.

Sîrbu, A., Pedreschi, D., Giannotti, F., & Kertész, J. (2018). Algorithmic bias amplifies opinion polarization: A bounded confidence model. *arXiv preprint,* arXiv:1803.02111.

Sloman, S.A. (1993). Feature-based induction. *Cognitive Psychology, 25*, 231–280.

Soll, J. B., & Larrick, R. P. (2009). Strategies for revising judgment: How (and how well) people use others' opinions. *Journal of experimental psychology: Learning, Memory, and Cognition*, *35*, 780–805.

Stigler, G. J. (1961). The economics of information. *Journal of Political Economy, 69*, 213-225.

Stone, J., & Moskowitz, G. B. (2011). Non-conscious bias in medical decision making: what can be done to reduce it?. *Medical Education*, *45*, 768-776.

Strevens, M. (2005). The Baysian approach in the philosophy of science. In D. M. Borchet (Ed.), *Encyclopedia of philosophy* (2nd ed.). Macmillan Reference.

Suchow, J. W., Bourgin, D. D., & Griffiths, T. L. (2017). Evolution in mind: Evolutionary dynamics, cognitive processes, and bayesian inference. *Trends in Cognitive Sciences*, *21*, 522-530.

Sunstein, C. R. (2017). *# Republic: Divided democracy in the age of social media*. Princeton University Press.

Sussman, A. B., Petkova, K., & Todorov, A. (2013). Competence ratings in US predict presidential election outcomes in Bulgaria. *Journal of Experimental Social Psychology*, *49*, 771-775.

Sutton, R. S., & Barto, A. G. (1998). *Introduction to reinforcement learning*. MIT press.

Suzuki, S., Harasawa, N., Ueno, K., Gardner, J. L., Ichinohe, N., Haruno, M., ... & Nakahara, H. (2012). Learning to simulate others' decisions. *Neuron*, *74*, 1125-1137.

Syed, M. (2019). *Rebel ideas: The power of diverse thinking*. Hachette UK.

Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behavior. In S. Worchel, & W. G. Austin (Eds.), *Psychology of intergroup relations* (pp. 7–24). Nelson-Hall.

Tajima, S., Drugowitsch, J., & Pouget, A. (2016). Optimal policy for value-based decision-making. *Nature Communications*, *7*, 1-12.

Tappin, B. M., van der Leer, L., & McKay, R. T. (2017). The heart trumps the head: Desirability bias in political belief revision. *Journal of Experimental Psychology: General*, *146*, 1143-1149.

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: a social psychological perspective on mental health. *Psychological Bulletin*, *103*, 193-210.

Tenenbaum, J. B. (1999). *A Bayesian framework for concept learning* (Doctoral dissertation, Massachusetts Institute of Technology).

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences*, *24*, 629-640.

Tenenbaum, J. B., & Griffiths, T. L. (2001). Structure learning in human causal induction. *Advances in Neural Information Processing Systems*, 59-65.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*, 1279-1285.

Tennie, C., Call, J., & Tomasello, M. (2009). Ratcheting up the ratchet: on the evolution of cumulative culture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*, 2405-2415.

Tetlock, P. E., Kristel, O. V., Elson, S. B., Green, M. C., & Lerner, J. S. (2000). The psychology of the unthinkable: taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology*, *78*, 853–870.

The Pew Research Center. (2009, October 30). Partisanship and Cable News Audiences. Retrieved from http://www.pewresearch.org/2009/10/30/partisanship-and-cable-news-audiences/

Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology*, *4*, 25-29.

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science*, *308*, 1623-1626.

Toelch, U., Bach, D. R., & Dolan, R. J. (2014). The neural underpinnings of an optimal exploitation of social information under uncertainty. *Social Cognitive and Affective Neuroscience*, *9*, 1746-1753.

Tomasello, M. (1999). The human adaptation for culture. *Annual Review of Anthropology*, *28*, 509-529.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*, 675-691.

Traberg, C. S., & van der Linden, S. (2022). Birds of a feather are persuaded together: Perceived source credibility mediates the effect of political bias on misinformation susceptibility. *Personality and Individual Differences*, *185*, 111269.

Trivers, R. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology*, *46*, 35–57.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124-1131.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the rationality of choice. *Science, 211*, 453-458.

Uleman, J. S., Rim, S., Saribay, S. A., & Kressel, L. M. (2012). Controversies, questions, and prospects for spontaneous social inferences. *Social and Personality Psychology Compass, 6*, 657–673.

Van Bavel, J. J., & Pereira, A. (2018). The partisan brain: An identity-based model of political belief. *Trends in Cognitive Sciences*, *22*, 213-224.

Van Swol, L. M., & Sniezek, J. A. (2005). Factors affecting the acceptance of expert advice. *British Journal of Social Psychology*, *44*, 443-461.

Von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior.* Princeton University Press.

Vrieze, S. I. (2012). Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological Methods*, *17*, 228-243.

Vul, E., Goodman, N., Griffiths, T. L., & Tenenbaum, J. B. (2014). One and done? Optimal decisions from very few samples. *Cognitive Science*, *38*, 599-637.

Vuorre, M., & Bolger, N. (2017). Within-subject mediation analysis for experimental data in cognitive psychology and neuroscience. *Behavior Research Methods*, 1-19.

Wade-Benzoni, K. A., Hoffman, A. J., Thompson, L. L., Moore, D. A., Gillespie, J. J., & Bazerman, M. H. (2002). Barriers to resolution in ideologically based negotiations: The role of values and institutions. *Academy of Management Review*, *27*, 41-57.

Walton, D. (1997). *Appeal to expert opinion: Arguments from authority.* Pennsylvania State University Press

Warm, J. S., Parasuraman, R., & Matthews, G. (2008). Vigilance requires hard mental work and is stressful. *Human Factors*, *50*, 433-441.

Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, *20*, 273-281.

Watson, D., Klohnen, E. C., Casillas, A., Nus Simms, E., Haig, J., & Berry, D. S. (2004). Match makers and deal breakers: Analyses of assortative mating in newlywed couples. *Journal of Personality*, *72*, 1029-1068.

Wilson, R. C., & Collins, A. G. (2019). Ten simple rules for the computational modeling of behavioral data. *Elife*, *8*, e49547.

Wingen, T., Berkessel, J. B., & Englich, B. (2020). No replication, no trust? How low replicability influences trust in psychology. *Social Psychological and Personality Science*, *11*, 454-463.

Wojciszke, B., Bazinska, R., & Jaworski, M. (1998). On the dominance of moral categories in impression formation. *Personality and Social Psychology Bulletin*, *24*, 1251-1263.

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review, 114*, 245-272.

Yaniv, I., & Kleinberger, E. (2000). Advice taking in decision making: Egocentric discounting and reputation formation. *Organizational Behavior and Human Decision Processes*, *83*, 260-281.

Yaniv, I., & Milyavsky, M. (2007). Using advice from multiple sources to revise and improve judgments. *Organizational Behavior and Human Decision Processes*, *103*, 104-120.

Yaniv, I., Choshen-Hillel, S., & Milyavsky, M. (2011). Receiving advice on matters of taste: Similarity, majority influence, and taste discrimination. *Organizational Behavior and Human Decision Processes*, *115*, 111-120.

Yoshida, W., Dolan, R. J., & Friston, K. J. (2008). Game theory of mind. *PLoS Computational Biology*, *4*, e1000254.

Yu, M., Saleem, M., & Gonzalez, C. (2014). Developing trust: First impressions and experience. *Journal of Economic Psychology*, *43*, 16-29.

Yzerbyt, V. Y. (2018). The dimensional compensation model: Reality and strategic constraints on Warmth and Competence in intergroup perceptions. In A. E. Abele, & B. Wojciszke (Eds.), *Agency and communion in social psychology* (pp. 126–141). Routledge.

Yzerbyt, V. Y., Kervyn, N., & Judd, C. M. (2008). Compensation versus halo: The unique relations between the fundamental dimensions of social judgment. *Personality and Social Psychology Bulletin*, *34*, 1110-1123.

Yzerbyt, V. Y., Provost, V., & Corneille, O. (2005). Not competent but warm . . . Really? Compensatory stereotypes in the French-speaking world. *Group Processes and Intergroup Relations, 8*, 291-308.

Zaki, J., Kallman, S., Wimmer, G. E., Ochsner, K., & Shohamy, D. (2016). Social cognition as reinforcement learning: feedback modulates emotion inference. *Journal of Cognitive Neuroscience, 28*, 1270-1282.

Zysberg, L., & Nevo, B. (2004). “The smarts that counts?” Psychologists’ decision-making in personnel selection. *Journal of Business and Psychology, 19,* 117-124.

Appendix 1

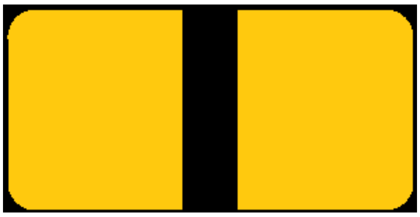
Instructions for Study 1

(Differences between Study 1 and Study 2 are highlighted in the main text)

Please read the following carefully to understand how to complete the task and earn the most money.

Welcome to our experiment! We are interested in how people understand rules.

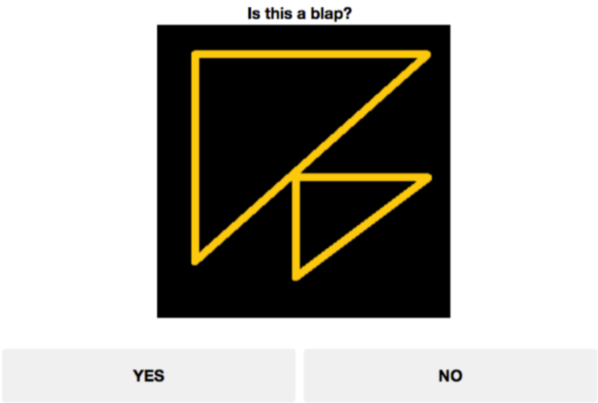
In some questions you will see pictures of objects, with different shapes and colors. Here are two examples:



Your job is to **learn through trial and error** how to recognize a certain type of object, called a 'blap'.

There are certain **rules that determine whether the object is likely to be a blap or not.** For example, the rule could be '80% of the time a shiny shape is a blap' (this is just an example).

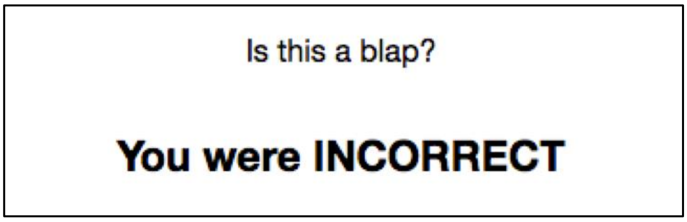
For each of these questions, the computer will show you a picture of an object, and will ask you whether you think it is a blap or not.



Press the **A** key on your keyboard for **“yes”** (if you think the object is a blap)

Press the **S** key on your keyboard for **“no”** (if you think the object is NOT a blap)

After you respond, **you will receive feedback**. You can use that feedback to get better at the task.

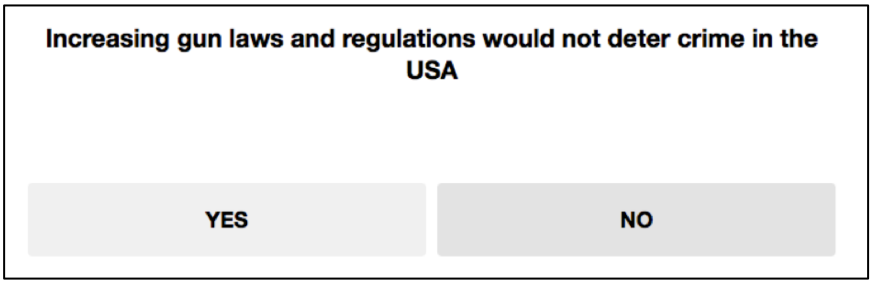


Is this a blap?

You were INCORRECT

Try your best to learn what the rules are, so that you can get better at classifying the objects as time goes on. This task is difficult, but you will **win more money in this experiment if you are good at recognizing when something is a blap or not!**

Other questions will involve your understanding of societal rules.
On each trial, the computer will show you a statement, such as

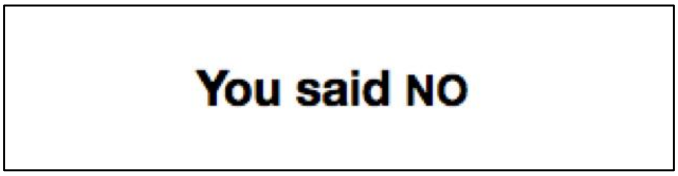


Increasing gun laws and regulations would not deter crime in the USA

Press the **A** key on your keyboard for **“yes”** (if you agree that increasing gun laws and regulations will not deter crime)

Press the **S** key on your keyboard for **“no”** (if you disagree that increasing gun laws and regulations will not deter crime)

You will then see your answer.



You said NO

Let’s practice. You will do 8 trials as practice.

Press next when you are ready to begin.

Remember to respond using the **A** and **S** keyboard keys

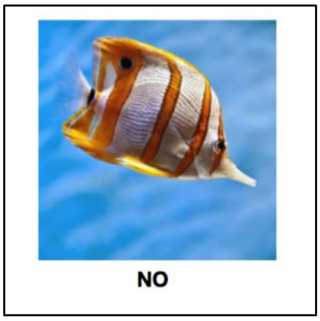
You have now finished the practice session.

Click next to proceed

Well done for completing the practice! During this next session, after you give your response you will see how one of four previous participants responded to the same question. These are four participants, who also completed the same task online like you, we will show you the answers that they put for the exact same questions.

To differentiate the four participants and keep them anonymous, they have been given different arbitrary animal icons, these are:

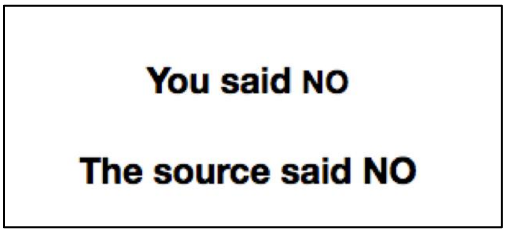
This is what the screen will look like:



NO

This means the fish responded 'no'.

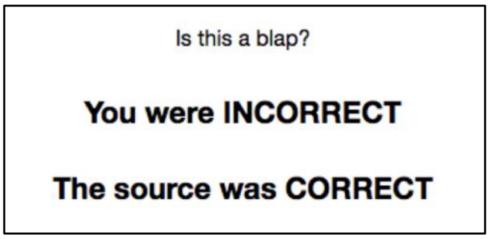
You will then see a feedback screen:



You said NO

The source said NO

OR



Is this a blap?

You were INCORRECT

The source was CORRECT

<u>Bonus Payment</u>

Over the course of the experiment, you should try to learn the characteristics of the 4 different sources. Knowing about how the other sources respond will help you later in the task, so make sure to pay attention to how they respond.

Your bonus payment will be based on the answers you give to the questions in the various stages of the task.

<u>If you are not performing the task properly you will lose your bonus and may be kicked out of the study prematurely.</u>

Before you can continue to the learning task, please answer the following questions to confirm you understand the task.

Congratulations, you answered all responses correctly. Press next to continue to the task.

REMEMBER:

- Blaps are defined by probabilistic rules, such as '80% of the time a shiny shape is a blap' (just an example)
- You must learn about the performance of 4 sources in order to do well in the later stages of this study.
- Your bonus is based on the accuracy of your answers
- Use the A and S keyboard keys to respond

Finally, throughout the task there will be a series of quiz trials about previously presented information. Too many incorrect responses on these questions will terminate the study so be sure to pay attention!

Instructions for Study 3

(Differences between Study 3 and Study 4 are highlighted in the main text)

Welcome to our study! We are interested in how people learn about other people's knowledge and generosity. This study will consist of two sessions.

In the first session, you will be shown the responses of four participants who completed one of our previous studies.

On half of the rounds, these participants were asked a multiple-choice general knowledge question. They earned £0.50 if they answered the question correctly and lost £0.50 if they answered incorrectly.

On the other half, they were given £1.00 and asked whether they would like to give half of this amount (i.e. £0.50) to a specific charity or keep the full amount (i.e. £1.00) for themselves.

Your task is to try to guess who answered the general knowledge questions correctly and who answered them incorrectly, as well as who gave money to each charity and who didn't.

On each round of this study, you will either be shown the general knowledge question the participants were asked or the name of the charity they could give money to.

For each general knowledge question, you will be asked to indicate who you think answered it correctly and who answered it incorrectly. You will then see who actually answered correctly and who didn't.

For each charity, you will be asked to indicate who you think gave £0.50 and who kept the £0.50 that could have been donated. You will then see who actually gave money and who kept it.

You will gain or lose points for each of your predictions.

If you accurately predict how a participant responded you will **gain 10 points**.

If you fail to predict how a participant responded you will **lose 10 points**.

The person with the highest score at the end of the experiment will win £40 (ties will be decided by a random draw). The winner will be announced next week.

To enable you to differentiate the four participants and to keep them anonymous, they have been given different arbitrary animal icons, as shown below:

On each **General Knowledge** question, you will be shown the general knowledge question the participants were asked and the four animal icons.

You will then be asked to bet on whether you think each participant answered the question correctly or incorrectly.

You should **click** on an animal icon to bet that the participant answered **correctly**.

You should **not click** on an animal icon to bet that the participant answered **incorrectly**.

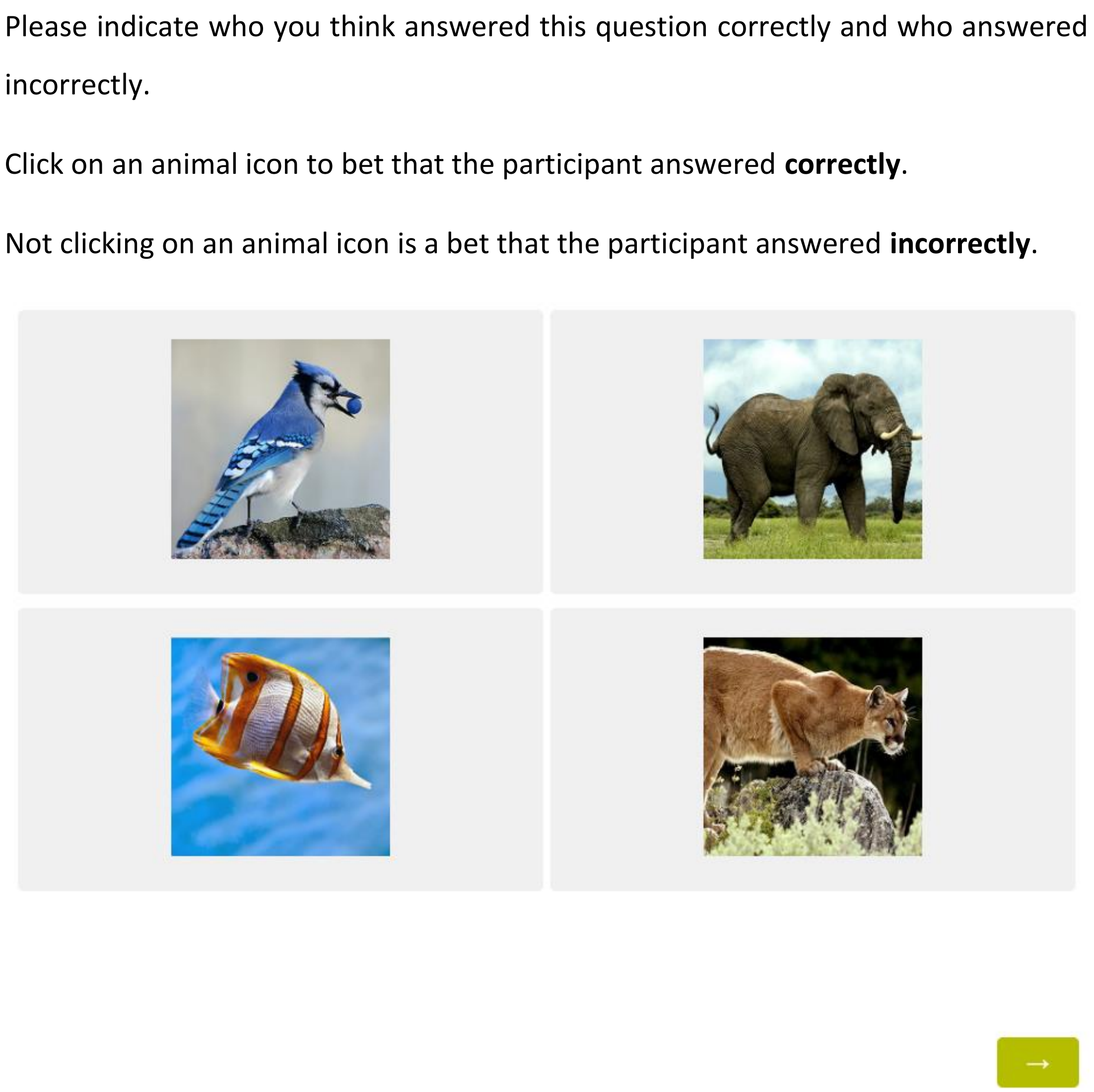
*Here is an example of what a **General Knowledge** question will look like:*

Question: Which is the fifth planet from the sun?

Options:

a) Jupiter

b) Saturn



Please indicate who you think answered this question correctly and who answered incorrectly.

Click on an animal icon to bet that the participant answered **correctly**.

Not clicking on an animal icon is a bet that the participant answered **incorrectly**.

After placing your bets, you will be shown the answer to the general knowledge question and whether the participants answered it correctly or incorrectly. If the participant was correct, a green tick will appear overlaid on the participant's animal icon.

Here is an example of what it would look like if all of the participants answered the question ***correctly*** *(in reality, some participants may answer correctly while others answer incorrectly):*

The correct answer was: Jupiter

Below you can see who answered this question correctly.

Below the animal icons, you will see how many points you earned through your betting.

---

*Here is another example of what a **General Knowledge** question will look like:*

Question: What is a ‘falchion’?

Options:

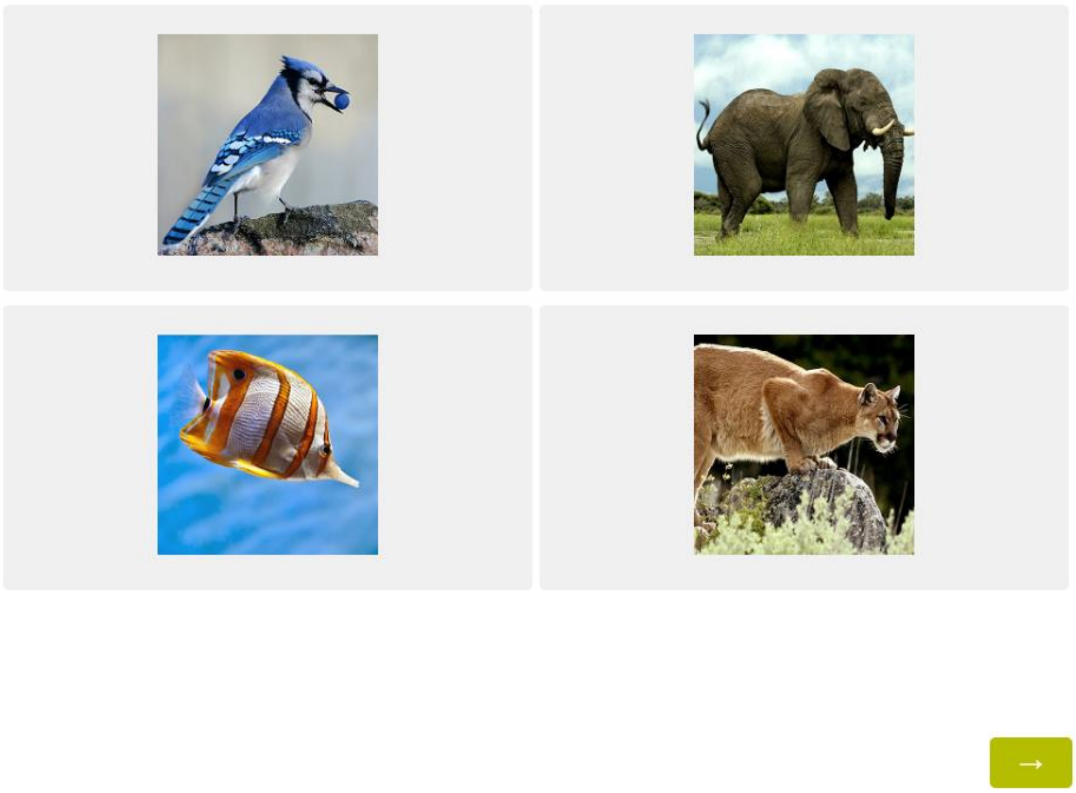
a) A type of bird

b) A type of sword

Please indicate who you think answered this question correctly and who answered incorrectly.

Click on an animal icon to bet that the participant answered **correctly**.

Not clicking on an animal icon is a bet that the participant answered **incorrectly**.

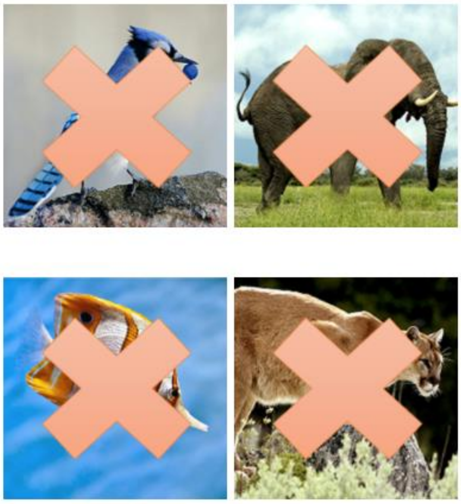


After placing your bets, you will be shown the answer to the general knowledge question and whether the participants answered it correctly or incorrectly. If the participant was incorrect, a red cross will appear overlaid on the participant's animal icon.

Here is an example of what it would look like if all of the participants answered the question ***incorrectly*** *(in reality, some participants may answer correctly while others answer incorrectly):*

The correct answer was: A type of sword

Below you can see who answered this question correctly.



Below the animal icons, you will see how many points you earned through your betting.

On each **Charity** question, you will be shown the name of the charity the participants could donate to and the four animal icons.

You will then be asked to bet on whether you think each participant gave £0.50 to the charity or kept the £1.00 for themselves.

You should **click** on an animal icon to bet that the participant **gave** £0.50 to the charity.

You should **not click** on an animal icon to bet that the participant **kept** the £1.00 for themselves.

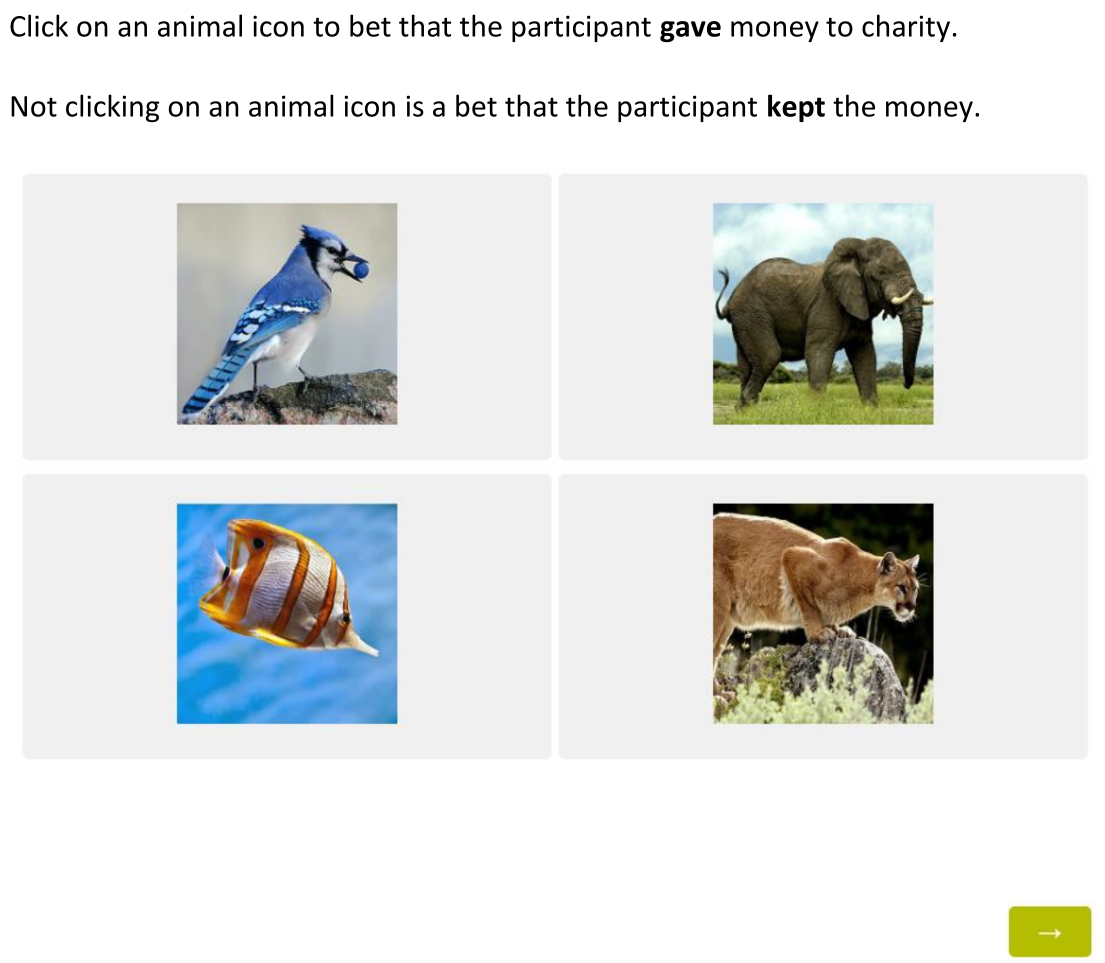
---

*Here is an example of what a **Charity** round will look like:*

Charity: THE PROSTATE CANCER CHARITY

Options: Give £0.50 (and Keep £0.50) **OR** Keep £1.00 (and Give £0.00)

Please indicate who you think gave away £0.50 and who kept the £1.00.



Click on an animal icon to bet that the participant **gave** money to charity.

Not clicking on an animal icon is a bet that the participant **kept** the money.

After placing your bets, you will be shown whether the participants gave or kept the money. If the participant gave money, the word 'Give' will appear in yellow overlaid on the participant's animal icon.

Here is an example of what it would look like if all of the participants chose to ***Give*** *money to the charity (in reality, some participants may choose to give while others choose to keep the money):*

Below you can see who gave £0.50 to the charity and who kept the full amount for themselves.

Below the animal icons, you will see how many points you earned from your bets.

---

*Here is another example of what a **Charity** round will look like:*

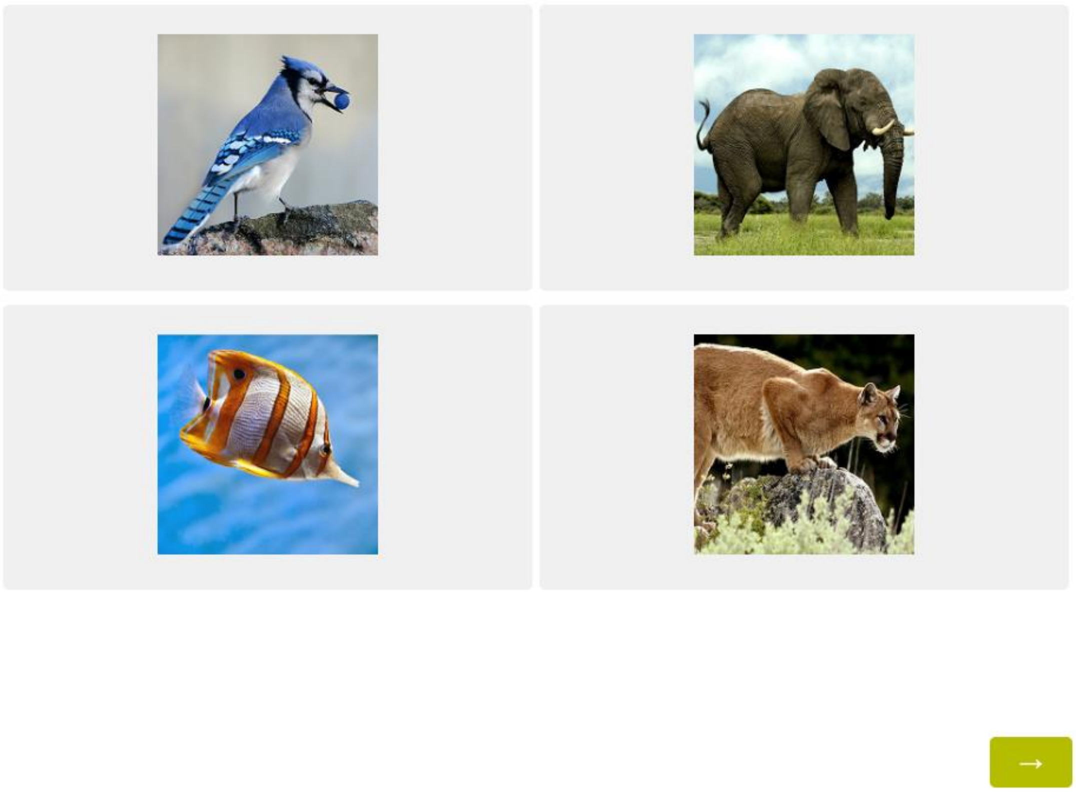
Charity: AMNESTY INTERNATIONAL

Options: Give £0.50 (and Keep £0.50) **OR** Keep £1.00 (and Give £0.00)

Please indicate who you think gave away £0.50 and who kept the £1.00.

Click on an animal icon to bet that the participant gave money to charity.

Not clicking on an animal icon is a bet that the participant **kept** the money.



After placing your bets, you will be shown whether the participants gave or kept the money. If the participant kept the money, the word 'Keep' will appear in purple overlaid on the participant's animal icon.

Here is an example of what it would look like if all of the participants chose to ***Keep*** *the money (in reality, some participants may choose to give while others choose to keep the money):*

Below you can see who gave £0.50 to the charity and who kept the full amount for themselves.

Below the animal icons, you will see how many points you earned from your bets.

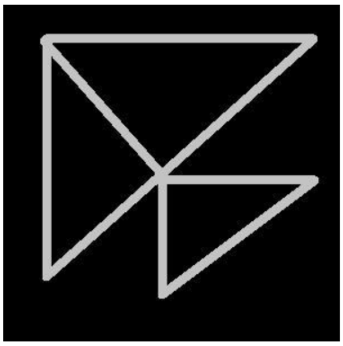
Instructions for Study 5

(Differences between Study 5 and Study 6 are highlighted in the main text)

Welcome to our study! We are interested in how information people learn about other people relates to their ability to learn patterns.

In some questions you will be asked questions related to your personal values.

In other questions you will see pictures of objects of different shapes and colours. Here are two examples:



Your job is to **learn through trial and error** how to recognize a certain type of object, called a ‘blup’.

There are **certain rules that determine whether the object is likely to be a blup or not.** For example, the rule could be '80% of the time a shiny shape is a blup' (this is just an example).

For each of these questions, you will be shown a picture of an object, and will asked whether you think it is a blup or not.

After you make each guess, you will then be told whether you were correct or incorrect.

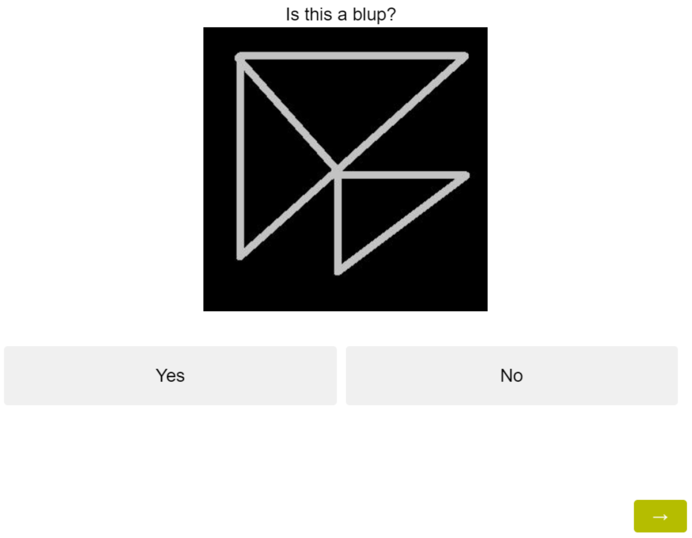
Try your best to learn what the rules are, so that you can get better at classifying the objects as time goes on. This task is difficult, but you will **receive a bonus payment of up to £1 if you do well in the tasks!**

After you answer each question, you will see **how four previous participants responded to the same question.** These four participants completed this task online for one of our previous studies. We will show you the answers that they put for the exact same questions.

Pay close attention to how these previous participants answered the questions as you will be tested on this during the study!

To enable you to differentiate the four participants and to keep them anonymous, they have been given different arbitrary animal icons, as shown below:

Here is an example of what a blup question will look like.



Is this a blup?

After you give your answer you will be shown the correct answer and whether the four previous participants ***answered the same question correctly or incorrectly****. Here is an example of what this will look like. If a previous participant was correct, a green tick will appear overlaid on the participant's animal icon. In this example you and all the participants answered the question* ***correctly*** *(in reality, some participants may answer correctly while others answer incorrectly).*

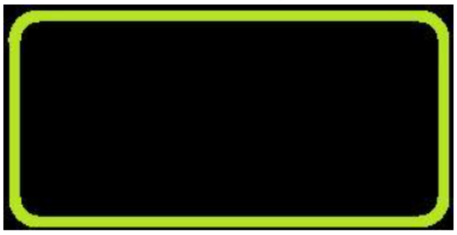
You answered: No

The correct answer was: No

Below you can see if our previous participants got this question correct or incorrect.

Here is an another example of a blup question.

Is this a blup?



Yes

No

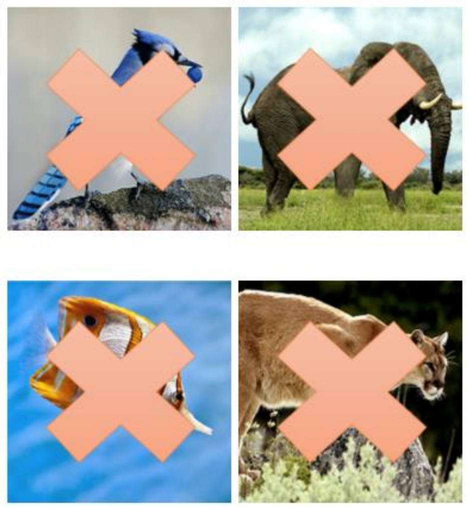
→

Again, after you give your answer you will be shown the correct answer and whether the four previous participants ***answered the same question correctly or incorrectly****. Here is another example of what this will look like. If a previous participant was incorrect, a red cross will appear overlaid on the participant's* *animal icon. In this example you and all the participants answered the question* ***incorrectly*** *(in reality, some participants may answer incorrectly while others answer correctly).*

You answered: No

The correct answer was: Yes

Below you can see if our previous participants got this question correct or incorrect.

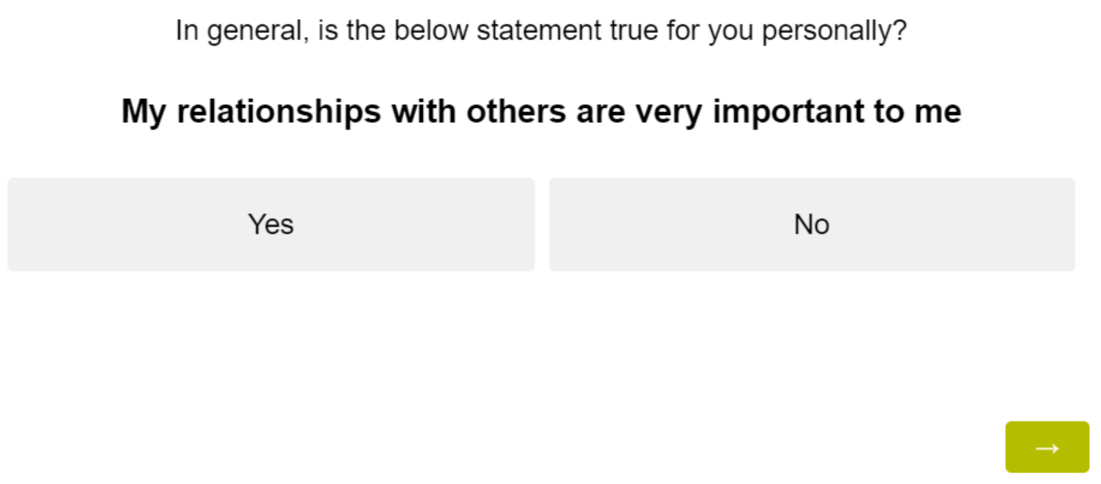


On other questions, you will be shown a statement designed to tell us something about your personal values and asked to tell us if it is, in general, true for you.

After you answer a question about your personal values you will be shown how the four previous participants answered the same question.

Pay close attention to how these previous participants answered the questions, as you will be tested on this during the study!

Here is an example of what a personal values question will look like.



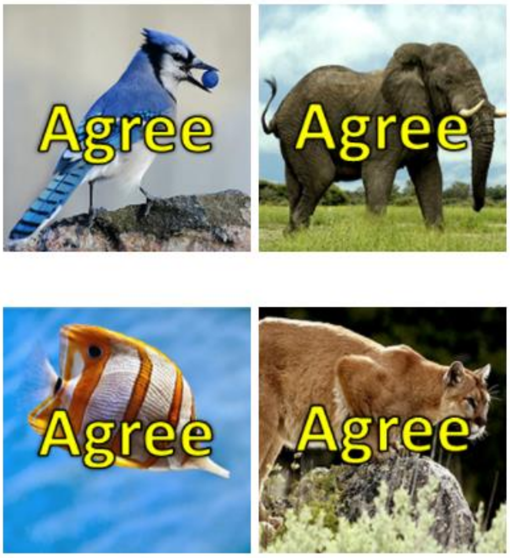
In general, is the below statement true for you personally?

My relationships with others are very important to me

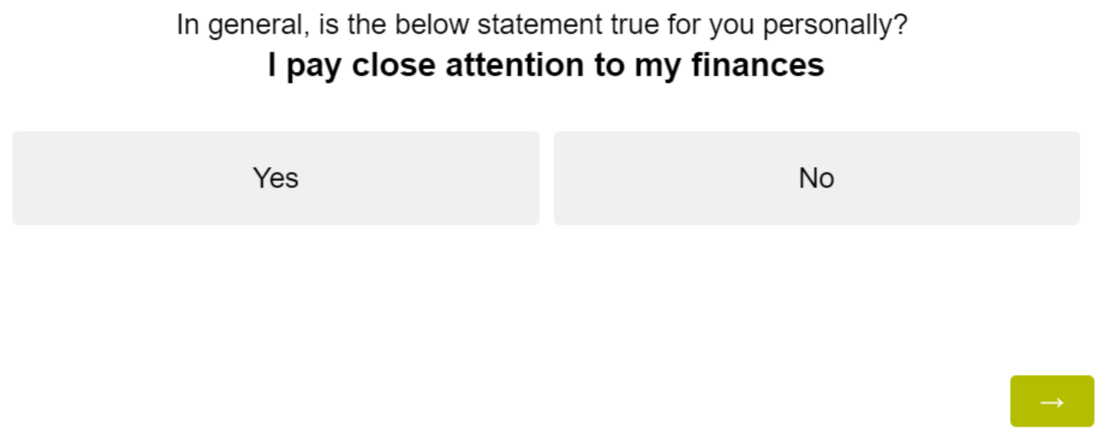
*After you give your answer you will be shown whether the four previous participants agreed or disagreed **with your answer**.*

*Here is an example of what this will look like. **If a participant agreed with you, the word 'Agree' will appear in yellow overlaid on the participant's animal icon**. In this example all of the participants agreed with you (in reality, some participants may agree and some may disagree).*

Below you can see whether our previous participants agreed or disagreed **with your previous answer.**



Here is another example of a personal values question.



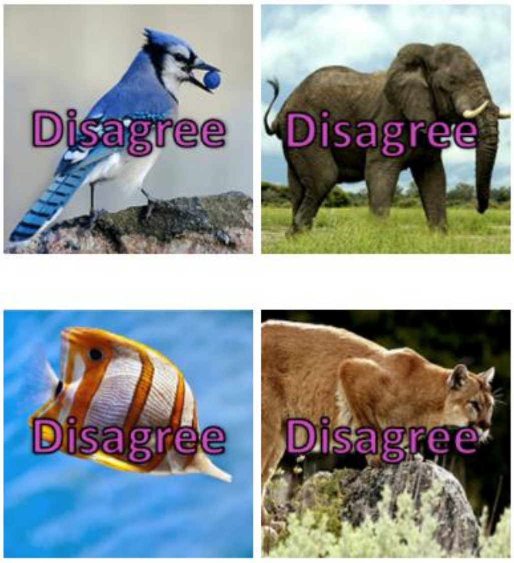
In general, is the below statement true for you personally?

I pay close attention to my finances

Again, after you give your answer you will be shown whether the four previous participants agreed or disagreed ***with your answer****.*

Here is another example of what this will look like. If a participant disagreed with you, the word 'Disagree' will appear in purple overlaid on the participant's animal icon. In this example all of the participants disagreed with you (in reality, some participants may disagree and some may agree).

Below you can see whether our previous participants agreed or disagreed **with your previous answer.**



Appendix 2

Demographic Information

Study 1

(Participants who completed the entire experiment)

Participants reported:

Gender and age: 34 females and 63 males, aged 20–58 years M = 34.81, SD = 9.59.

Ethnicity: 78% White, 6% Black, 6% Hispanic, 7% Asian, 2% Other.

Whether English was their first language: 98% said “yes”, “2%” said no.

Highest level of education completed:

High School Diploma	37%
2 Year Degree	22%
4 Year Degree	33%
Postgraduate/Professional Degree	7%
Other	1%

The approximate amount of income they earned in 2016:

Under $5,000	11%
$5,000-$10,000	8%
$10,001-$15,000	6%
$15,001-$25,000	7%
$25,001-$35,000	16%
$35,001-$50,000	22%

$50,001-$65,000	8%
$65,001-$80,000	10%
$80,001-$100,000	6%
Over $100,000	5%

Subjective socio-economic position on a 10-point scale (from 1 = “Worst off” to 10 = “Best off”; *M* = 6.54, *SD* = 1.62, Range = 3-10). A one-sample t-test showed that the mean was significantly different from the midpoint of the scale (*t*(96) = 9.34, *p* < .001), with participants reporting higher than average subjective socio-economic position.

Political ideology (on a sliding scale from 0 = “Liberal” to 1 = “Conservative”; *M* = .41, *SD* = .29, Range = 0-1). A one-sample t-test showed that the mean was significantly different from the 0.5 midpoint of the scale (*t*(96) = -2.99, *p* = .004), suggesting our sample was ideologically left of centre.

Interest/involvement in US politics (from 0 = “Not at all” to 100 = “Completely”; *M* = 60.96, *SD* = 27.40, Range = 0-100). A one-sample t-test showed that the mean was significantly different from the 50% midpoint of the scale (*t*(96) = 3.94, *p* < .001), with participants reporting greater interest and involvement than not.

Trust in other people that they interact with in daily life (from 1 = “Very little” to 7 = “Very much”; *M* = 4.92, *SD* = 1.60, Range = 1-7). A one-sample t-test showed that the mean was significantly different from the midpoint of the scale (*t*(96) = 8.73, *p* < .001), with participants reporting high levels of trust in the people they interact with.

Study 2

Gender and age: 47 females and 54 males, aged 18–63 years M = 37.59, SD = 10.92.

Ethnicity: 81% White, 5% Black, 5% Hispanic, 9% Asian, 0% Other.

English was first language: 99% said “yes”, “1%” said no.

Highest level of education completed:

High School Diploma	29%
2 Year Degree	25%
4 Year Degree	34%
Postgraduate/Professional Degree	13%
Other	0%

The approximate amount of income they earned in 2016:

Under $5,000	11%
$5,000-$10,000	8%
$10,001-$15,000	5%
$15,001-$25,000	20%
$25,001-$35,000	8%
$35,001-$50,000	13%
$50,001-$65,000	18%
$65,001-$80,000	11%
$80,001-$100,000	4%
Over $100,000	3%

Subjective socio-economic position on a 10-point scale (from 1 = "Worst off" to 10 = "Best off"; *M* = 6.49, *SD* = 1.76, Range = 1-10). A one-sample t-test showed that the mean was significantly different from the midpoint of the scale ($t(100) = 8.46$, $p < .001$), with participants reporting higher than average subjective socio-economic position.

Political ideology (on a sliding scale from 0 = “Liberal” to 1 = “Conservative”; *M* = .45, *SD* = .32, Range = 0-1). A one-sample t-test showed that the mean was not significantly different from the midpoint of the scale (*t*(100) = -1.64, *p* = .11).

Interest/involvement in US politics (from 0 = “Not at all” to 100 = “Completely”; *M* = 66.91, *SD* = 26.24, Range = 0-100). A one-sample t-test showed that the mean was significantly different from the midpoint of the scale (*t*(100) = 6.48, *p* < .001), with participants reporting higher than average interest and involvement in politics.

Trust in other people that they interact with in daily life (from 1 = “Very little” to 7 = “Very much”; *M* = 5.04, *SD* = 1.47, Range = 1-7). A one-sample t-test showed that the mean was significantly different from the midpoint of the scale (*t*(100) = 10.53, *p* < .001), with participants reporting high levels of trust in the people they interact with.

Study 3

(Participants included in the analysis)

Gender and age: 43 females, 9 males; mean age = 18.79, SD = 0.75.

Country of residence: 65% UK, 35% non-UK.

Study 4

(Participants included in the analysis)

Gender and age: 45 females, 5 males; mean age = 18.78, SD = 0.68.

Country of residence: 60% UK, 40% non-UK.

Study 5

(Participants included in the analysis)

Gender and age: 12 males, 38 females; mean age = 32.32 years, SD = 12.87.

Country of residence: 98% UK, 2% non-UK.

Study 6

(Participants included in the analysis)

Gender and age: 113 females, 13 males, and 2 who said “other” when asked about their gender, mean age = 18.57, SD = 0.88.

Country of residence: 30% UK, 70% non-UK.

Debrief Questions

Study 1

In the debrief, participants (who completed the entire experiment) were asked:

To report what they thought was the purpose of the study in an open-answer format: 57 participants reported that they did not know what the purpose of the study was or provided an incorrect answer (e.g., “To determine memory recall”). 40 participants provided answers that were related to a goal or sub-goal of the study (i.e., answers that mentioned testing relationship between similarity and influence or competence).

Whether they found any rule(s) to decide if each object was a blap, and what the rules were: 51 participants reported a rule, 44 reported that they did not know, one said it seemed random, and one did not answer the question.

How sure they were that their rule(s) were correct (from 0 = Not at all confident, to 100 = Very confident; *M* = 39.38, *SD* = 26.72, Range = 0-100). A one-sample t-test showed that the mean was significantly different from the midpoint of the scale (*t*(96) = -3.91, *p* < .001), with participants reporting low levels of confidence in their rule(s) for categorising blaps.

How many blaps there were in the task, as a percentage (*M* = 59.80, *SD* = 17.51, Range = 5-95). A one-sample t-test showed that the mean was significantly different from the midpoint of the scale (*t*(96) = 5.52, *p* < .001), with participants believing that a greater than average percentage of the items in the task were blaps.

How well they learned about the accuracy of the four sources in the blap task: 24 participants thought they learned about the accuracy of all four, 66 thought they learned about some of the sources, 4 thought they didn't learn about any of the sources and 3 reported that they did not know.

How well they learned about the political opinions of the four sources: 37 participants thought they learned about all four, 49 thought they learned about some of the sources, 7 thought they didn't learn about any of the sources and 4 reported that they did not know.

To rate each source on a number of dimensions:

	Similar-Accurate	**Dissimilar-Accurate**	**Similar-Random**	**Dissimilar-Random**
Competence in the blap task (see main text)	71.93 (19.26)	64.11 (23.58)	67.54 (18.49)	53.40 (23.09)
Consistency of performance in the blap task	73.47 (18.92)	64.11 (21.90)	69.46 (15.98)	56.45 (22.14)
Political views (from 0 = "Liberal" to 1 = "Conservative")	.44 (.30)	.55 (.33)	.49 (.27)	.54 (.27)
Consistency of political views	77.64 (16.11)	67.78 (25.61)	74.42 (16.59)	62.11 (22.78)
Trust in source	73.34 (19.53)	54.75 (25.18)	67.94 (20.72)	46.05 (24.75)
Similarity (see main text)	77.06 (14.88)	30.19 (21.54)	70.60 (18.04)	35.07 (22.90)

Numbers presented in the table are means with standard deviations in parentheses

General impressions and any other comments for each source in open-answer format: the open-text answers tended to match up with the quantitative measurements.

Which source they preferred for blap questions: 42 participants reported they preferred the Similar-Accurate source, 25 the Dissimilar-Accurate, 23 the Similar-Random, 7 the Dissimilar-Random.

Which source they avoided for blap questions: 8 participants reported that they avoided the Similar-Random, 15 the Similar-Accurate source, 19 the Dissimilar-Accurate, 55 the Dissimilar-Random.

How they made decisions about which source to choose for blap questions in open-answer format: participants generally reported that they chose the source that seemingly performed the best in the learning stage.

Whether they chose a source that they thought would be wrong so that they could do the opposite: 15 participants reported that they used this strategy.

To what extent they believed that the responses from the sources were those of previous participants?" (from 0 = "Did not at all believe it" to 100 = "Completely believed it"). Despite the question being the last in the funneled debriefing, and thus the most specific and closed-ended, the ratings revealed only mild suspicion rates with a mean score not significantly different from the mid-point of the scale (*M* = 44.44, *SD* = 33.96, *t*(96) = -1.61, *p* = .11).

Finally, participants were thanked and asked if any of the instructions were unclear and if they had any final comments for the researchers.

Study 2

(Participants who completed the entire experiment)

Participants reported:

What they thought was the purpose of the study in an open-answer format: 72 participants reported that they did not know what the purpose of the study was or provided an incorrect answer. 29 participants provided answers that were related to a goal or sub-goal of the study (i.e., answers that mentioned testing relationships between similarity and influence or competence).

Whether they found any rule(s) to decide if each object was a blap, and what the rules were: 52 participants reported a rule, 35 reported that they did not know, 3 said it seemed random, 8 said that they tried to remember specific examples and 6 provided answers that did not address the question.

How sure they were that their rule(s) were correct (from 0 = Not at all confident, to 100 = Very confident; *M* = 41.09, *SD* = 26.53, Range = 0-100). A one-sample t-test showed that the mean was significantly different from the midpoint of the scale (*t*(100) = -3.38, *p* < .001), with participants reporting low levels of confidence in their rule(s) for categorising blaps.

How many blaps there were in the task, as a percentage (*M* = 58.52, *SD* = 18.08, Range = 10-100). A one-sample t-test showed that the mean was significantly different from the midpoint of the scale (*t*(100) = 4.74, *p* < .001), with participants believing that a greater than average percentage of the items in the task were blaps.

How competent they thought they were at the blap task (from 0 = Very incompetent, to 100 = Very competent; *M* = 44.78, *SD* = 23.16, Range = 0-100). A one-sample t-test showed that the mean was significantly different from the midpoint of the scale (*t*(100) = -2.26, *p* = .026), with participants believing they were worse than average at guessing which shapes were blaps.

How well they learned about the accuracy of the four sources in the blap task: 26 participants thought they learned about the accuracy of all four, 61 thought they learned about some of the sources, 8 thought they didn't learn about any of the sources and 6 reported that they did not know.

How well they learned about the political opinions of the four sources: 36 participants thought they learned about all four, 58 thought they learned about some of the sources, 3 thought they didn't learn about any of the sources and 4 reported that they did not know.

To rate each source on a number of dimensions:

	Similar-Accurate	**Dissimilar-Accurate**	**Similar-Random**	**Dissimilar-Random**
Competence in the blap task (see main text)	67.80 (20.01)	62.55 (20.25)	61.48 (18.79)	56.67 (20.04)
Consistency of performance in the blap task	68.50 (18.51)	62.03 (20.58)	61.92 (17.63)	59.40 (18.25)
Political views (from 0 = "Liberal" to 1 = "Conservative")	.50 (.28)	.52 (.33)	.46 (.28)	.55 (.28)
Consistency of political views	72.09 (15.10)	64.55 (21.52)	69.03 (17.84)	62.71 (19.74)
Trust in source	69.05 (21.27)	52.83 (26.20)	61.25 (20.35)	50.09 (23.77)
Similarity (see main text)	71.99 (16.84)	29.98 (22.53)	66.89 (21.79)	34.74 (21.26)

Numbers presented in the table are means with standard deviations in parentheses

General impressions and any other comments for each source in open-answer format: the open-text answers tended to match up with the quantitative measurements.

Which source they preferred for blap questions: 41 participants reported they preferred the Similar-Accurate source, 24 the Dissimilar-Accurate, 24 the Similar-Random, 12 the Dissimilar-Random.

Which source they avoided for blap questions: 11 participants reported that they avoided the Similar-Random, 15 the Similar-Accurate source, 37 the Dissimilar-Accurate, 38 the Dissimilar-Random.

How they made decisions about which source to choose for blap questions in open-answer format: participants generally reported that they chose the source that seemingly performed the best in the learning stage.

Whether they chose a source that they thought would be wrong so that they could do the opposite: 28 participants reported that they used this strategy.

To what extent they believed that the responses from the sources were those of previous participants?" (from 0 = "Did not at all believe it" to 100 = "Completely believed it"). Despite the question being the last in the funneled debriefing, and thus the most specific and closed-ended, the ratings revealed only mild suspicion rates with a mean score not significantly different from the mid-point of the scale (*M* = 45.58, *SD* = 30.18, *t*(100) = -1.47 *p* = .15).

Finally, participants were thanked and asked if any of the instructions were unclear and if they had any final comments for the researchers.

Study 3

In the debrief, participants were asked:

To report what they thought was the purpose of the study in an open-answer format: 45 participants reported that they did not know what the purpose of the study was or provided an incorrect answer (e.g., "cognitive availability and decision making?"). 8 participants provided answers that were related to a goal or sub-goal of the study (i.e., answers that mentioned testing relationship between generosity and influence or competence).

To what extent they believed that the responses from the sources were those of previous participants?" (response options: 1 = "Certainly not previous participants", 2 = "Probably not previous participants", 3 = "Unsure", 4 = "Probably previous participants", 5 = "Certainly previous participants"). The ratings revealed only mild suspicion rates with a mean score suggesting that participants were slightly skeptical that the responses were from real people (M = 2.73, SD = 0.91, t(51) = -2.13 p = .038).

How they made decisions about which source to choose in the Choice Stage in open-answer format: participants generally reported that they chose the source that seemed to perform best in the learning stage.

Whether they chose a source that they thought would be wrong so that they could do the opposite: 22 participants reported that they used this strategy.

Finally, participants were thanked and asked if any of the instructions were unclear and if they had any final comments for the researchers.

Study 4

In the debrief, participants were asked:

To report what they thought was the purpose of the study in an open-answer format: 27 participants reported that they did not know what the purpose of the study was or provided an incorrect answer. 23 participants provided answers that were related to a goal or sub-goal of the study (i.e., answers that mentioned testing relationship between similarity and influence or competence).

To what extent they believed that the responses from the sources were those of previous participants?" (Response options: 1 = "Certainly not previous participants", 2 = "Probably not previous participants", 3 = "Unsure", 4 = "Probably previous participants", 5 = "Certainly previous participants"). The ratings revealed only mild suspicion rates with a mean score suggesting that participants were slightly skeptical that the responses were from real people (M = 2.48, SD = 0.95, t(49) = -3.86 p < .001).

How they made decisions about which source to choose in the Choice Stage in open-answer format: participants generally reported that they chose the source that seemed to perform best in the learning stage.

Whether they chose a source that they thought would be wrong so that they could do the opposite: 15 participants reported that they used this strategy.

Finally, participants were thanked and asked if any of the instructions were unclear and if they had any final comments for the researchers.

Study 5

In the debrief, participants were asked:

To report what they thought was the purpose of the study in an open-answer format: 32 participants reported that they did not know what the purpose of the study was or provided an incorrect answer. 18 participants provided answers that were related to a goal or sub-goal of the study (i.e., answers that mentioned testing relationship between similarity and influence or competence).

To what extent they believed that the responses from the sources were those of previous participants?" (Response options: 1 = "Certainly not previous participants", 2 = "Probably not previous participants", 3 = "Unsure", 4 = "Probably previous participants", 5 = "Certainly previous participants"). The ratings revealed only mild suspicion rates with a mean score suggesting that participants were slightly skeptical that the responses were from real people (M = 2.54, SD = 0.91, t(49) = -3.58 p < .001).

How they made decisions about which source to choose in the Choice Stage in open-answer format: participants generally reported that they chose the source that seemed to perform best in the learning stage.

Whether they chose a source that they thought would be wrong so that they could do the opposite: 15 participants reported that they used this strategy.

Finally, participants were thanked and asked if any of the instructions were unclear and if they had any final comments for the researchers.

Study 6

In the debrief, participants were asked:

To report what they thought was the purpose of the study in an open-answer format: 83 participants reported that they did not know what the purpose of the study was or provided an incorrect answer. 45 participants provided answers that were related to a goal or sub-goal of the study (i.e., answers that mentioned testing relationship between similarity and influence or competence).

To what extent they believed that the responses from the sources were those of previous participants?” (Response options: 1 = “Certainly not previous participants”, 2 = “Probably not previous participants”, 3 = “Unsure”, 4 = “Probably previous participants”, 5 = “Certainly previous participants”). The ratings revealed only mild suspicion rates with a mean score suggesting that participants were slightly skeptical that the responses were from real people (M = 2.43, SD = 1.07, t(127) = -6.03 p < .001).

How they made decisions about which source to choose in the Choice Stage in open-answer format: participants generally reported that they chose the source that seemed to perform best in the learning stage.

Whether they chose a source that they thought would be wrong so that they could do the opposite: 47 participants reported that they used this strategy.

Finally, participants were thanked and asked if any of the instructions were unclear and if they had any final comments for the researchers.

Appendix 3

Political Statement Stimuli for Studies 1 and 2

The political statement stimuli were adapted from the following sources:

https://www.isidewith.com/political-quiz#

https://www.isidewith.com/polls

http://www.people-press.org/quiz/political-party-quiz/

Participants saw the following four practice stimuli and 80 statements from the Learning Stage stimuli:

Practice Stimuli
Increasing gun laws and regulations would not deter crime in the USA
Lowering the minimum voting age would help get young people interested in politics
A politician formerly convicted of a crime would likely make bad decisions in office
Remaining in NATO will help secure a peaceful future for the USA
Learning Stage Stimuli
The risks from offshore oil drilling are minimal
The Paris Climate Agreement disadvantages US businesses and workers
Assassinating suspected terrorists in foreign countries helps keep the world a safe place
Deporting immigrants who are potential threats will make America safer
A stricter USA immigration policy will improve social cohesion
Americans will flourish if immigrants who commit crimes are deported
Immigrants would fit in better if compelled to learn English
America could improve national safety with tighter border control
Immigrants take jobs away from people born in the USA
High immigration results in lower wages for US citizens
Illegal immigrants need to feel scared of being deported to prevent more coming to the USA
Immigrants abuse the welfare system
Women who get abortions usually don't understand the consequences of what they are doing
Gay marriage confuses children
Nuclear power is an unsafe method for generating energy
Cutting public spending will reduce national debt
Lowering the tax rate for corporations will reduce unemployment in America
Spending less on social welfare will motivate people to work
Low restrictions on access to welfare benefits encourages people to abuse the

system
Welfare recipients usually spend the money on drugs and alcohol
Labour unions hurt the economy
Ordinary people get a good proportion of the nation’s wealth
Encouraging private enterprise will improve the US economy
Private corporations educate political parties about important issues through lobbying
Assassinating suspected terrorists in foreign countries helps keep America safe
Terrorism would decrease if government surveillance were expanded to combat terrorism
Increased spending on the military will help to keep America safe
Allowing the police to monitor the phone calls and emails of criminals helps keep America safe
Giving nonviolent drug offenders mandatory jail sentences would reduce rates of delinquency
The bans on medicinal and recreational drugs protect people from themselves and others
Fracking is the safest way to keep oil prices low
Many of the claims about environmental threats are exaggerated
The death penalty deters people from committing crimes
The police would be more effective if they could access individual's private data
A worldwide American military presence helps maintain peace
A tough justice system keeps the crime rate low
Strong trade unions prevent industry goals from being achieved
Benefits for unemployed people are too high and discourage them from finding jobs
Most unemployed people don't try that hard to find a job
Patriotism improves social cohesion
Reducing civil liberties helps to maintain order in society
Newer lifestyles are contributing to a breakdown in society
Going to war is sometimes the only solution to international problems
Individual initiative needs to be incentivised to promote competition, even if this increases inequality
Criticising your country has an effect on your American identity
Society works better if people adhere to a simple unbending moral code
Society works better if all people are left to accomplish things on their own
Swift and severe punishment for criminals helps to maintain peace
Western civilization has brought more progress than all other cultural traditions
Social charities create dependency
Going to war with a country can actually improve outcomes for that country
Rewarding some more than others motivates competition
The police would be more effective if there were not so many rules preventing them from doing their jobs
Immigrants who work hard tend to find success in America
Building a wall around the southern border would reduce illegal immigration

It is unsafe to let Muslim immigrants enter the country until the government improves its ability to screen out potential terrorists
Illegal immigrants should not have access to government-subsidized healthcare
Health insurers exploit individuals who have a pre-existing medical condition
Requiring a photo ID before letting people vote would greatly impact election results
The death penalty is more than just a political tool
Internet service providers speed up access to popular websites (that pay higher rates) at the expense of slowing down access to other websites
Burning the American flag indicates a person may want to harm Americans
Privatization of veterans' healthcare will reduce the burden on society
The vast majority of offshore investing is perfectly legal
Forcing 18 year olds to provide at least one year of military service would make America safer
Formally declaring war on ISIS would make America safer
Local police could increase safety by increasing surveillance and patrol of Muslim neighborhoods
The military fly drones over foreign countries to gain intelligence and kill suspected terrorists
Markets suffer as a result of government interference
Allowing people who are against democracy to run in elections is a threat to democratic rights
The government's funding of planned parenthood improves child outcomes
Foreign Aid spending reduces worldwide suffering
The US did not decrease foreign aid spending during the last recession
Immigration gives a boost to the national economy
Immigrants help the US learn about beneficial new ideas
Homosexual couples have the same adoption rights as same sex couples
Businesses are generally more profitable if they have at least one woman on their board of directors
Listening to hate speech fuels extremist behavioural tendencies
Marijuana legalisation reduces violent crime
A diverse society is more creative than a homogeneous society
The US could increase national happiness by raising taxes on the rich
The government is legally obliged to ensure everyone is provided for
Providing everyone with a guaranteed basic income would reduce unhappiness in America overall
It is the government’s role to redistribute income to curb inequality
Human rights are often overridden to maintain national security
Forcing businesses to reduce carbon emissions will help protect the environment from climate change
The government do not focus on improving animal rights
The government do not use taxes to protect the environment
Economic growth typically harms the environment
Increasing government spending on public transport helps individuals and

businesses to make money
Allowing convicted criminals the right to vote improves election outcomes
Releasing non-violent criminals from jail is only a small threat and reduces overcrowding
Zero-hour employment contracts are detrimental to workers
The national minimum wage prevents businesses exploiting workers
The national living wage is increased by more than inflation annually
Incomes are less equally distributed than Americans think
Supporting minimum income rising more sharply than other income levels would help to reduce crime
It is the government’s responsibility to provide a job for everyone who wants one
Big business tends to benefit owners at the expense of workers
Inequality reduction would improve social cohesion in America
The poor are discriminated against by the police
The police use racial stereotyping
Traditional values stop people from embracing technological advancements
Providing free health care to all citizens would bankrupt America
Individuals have the legal freedom to believe whatever they want
Every assertion made by our highest political and military leaders is subject to scrutiny
Treating other countries as equals makes them more willing to cooperate
It is harder for ethnic minorities to pass a job interview in America than it is for caucasians
Women are disadvantaged in the workplace compared to men
Giving disabled children extra resources at school will help them to overcome their handicaps
Disabled people should not face disadvantage in the workplace
Adding "Gender Identity" to anti-discrimination laws will reduce discrimination
Increasing restrictions on the current process of purchasing a gun would reduce gun crimes
Obamacare helps to provide affordable medical care to those who would not receive it otherwise
Reducing interest rates on student loans would increase the USA's intellectual dominance
Accepting refugees from Syria does not pose any great threat to America
Foreign terrorism suspects should be given constitutional rights
Removing confederate monuments and memorials from public grounds will reduce feelings of discrimination
Providing 'trigger warnings' and 'safe spaces' for students will increase their focus on learning rather than worrying about threats
Increasing the spending on public transportation will solve transportation problems, like traffic jams
A state displaying the confederate flag on government property is a sign of racism and seperatism
Banning a Niqab, or face veil, at civic ceremonies infringes on individual rights and prevents people from expressing their religious beliefs

The diversity due to affirmative action programs will give rise to innovation
When political candidates release their recent tax returns to the public, this increases their transparency
Increasing funding of health care for low income individuals would have a positive economic impact overall
Raising the tax rate for corporations will encourage companies to move to places with lower taxes
The Dakota Access pipeline will help the economy
Stricter smoking regulations will reduce the number of young people who start smoking
Allowing terminally ill patients to end their lives through assisted suicide will reduce unhappiness
The government should firmly control prices after wage increases
Positive discrimination is necessary to create a balance in the workplace and society
Abolishing the inheritance tax would result in increased inequality
Bringing essential public services and industries into state ownership would stop monopolies exploiting the public
Passing laws to protect whistle blowers would lead more people to come forward about wrong-doings
Increasing government spending will give a boost to the economy that is worth the extra debt
The government should regulate the price of life-saving drugs
Allowing the federal government to negotiate drug prices for Medicare will reduce health care costs
Making sure that everyone has an equal opportunity to succeed, will increase prosperity overall
Those with more resources have more obligations toward their fellow human beings
Abolishing the electoral college will lead to fairer elections

Charity Stimuli for Study 3

The charities were taken from a list of Britain's top 1,000 charities, ranked by donations (https://www.theguardian.com/news/datablog/2012/apr/24/top-1000-charities-donations-britain).

Participants saw the following two practice stimuli and 40 Learning Stage stimuli:

Practice Stimuli
THE PROSTATE CANCER CHARITY
AMNESTY INTERNATIONAL
Learning Stage Stimuli
CANCER RESEARCH

THE SAVE THE CHILDREN FUND
OXFAM
THE BRITISH RED CROSS SOCIETY
BRITISH HEART FOUNDATION
THE SALVATION ARMY
ACTIONAID
THE GUIDE DOGS FOR THE BLIND ASSOCIATION
AGE UK
WORLD WILDLIFE FUND (WWF)
BARNARDO'S
THE GREAT ORMOND STREET HOSPITAL CHILDREN'S CHARITY
THE MUSEUMS, LIBRARIES AND ARCHIVES COUNCIL
HELP FOR HEROES
WATERAID
YOUTH SPORT TRUST
ALZHEIMER'S SOCIETY
THE ROYAL SHAKESPEARE COMPANY
CATS PROTECTION
SOUTHBANK CENTRE
SHELTER, NATIONAL CAMPAIGN FOR HOMELESS PEOPLE LIMITED
ARTHRITIS RESEARCH UK
THE BRITISH DIABETIC ASSOCIATION
THE WORLD SOCIETY FOR THE PROTECTION OF ANIMALS
SAMARITAN'S PURSE INTERNATIONAL LIMITED
THE MULTIPLE SCLEROSIS SOCIETY OF GREAT BRITAIN AND NORTHERN IRELAND
THE GRAND CHARITY
AMNESTY INTERNATIONAL CHARITY LIMITED
THE NATIONAL FOUNDATION FOR YOUTH MUSIC
THE STROKE ASSOCIATION
PARKINSON'S DISEASE SOCIETY OF THE UNITED KINGDOM
BLIND VETERANS UK
LEUKAEMIA & LYMPHOMA RESEARCH
CONCERN WORLDWIDE (UK)
INTERNATIONAL FUND FOR ANIMAL WELFARE
WELLCOME TRUST
ACTION FOR CHILDREN
THE NATIONAL DEAF CHILDREN'S SOCIETY
EMERGE POVERTY FREE (WORLD EMERGENCY RELIEF)
CRISIS UK

Political Stimuli for Study 4

The political statements were adapted from questions on the website https://uk.isidewith.com/political-quiz.

Participants saw the following two pairs of practice stimuli and 40 pairs of Learning Stage stimuli:

Practice Stimuli	
The EU should impose a quota of migrants per country	The EU should not impose a quota of migrants per country
The government should raise the national minimum wage	The government should not raise the national minimum wage
Learning Stage Stimuli	
There should be fewer restrictions on current welfare benefits	There should be more restrictions on current welfare benefits
Homeowners should pay higher taxes on 'mansions' valued over £2m	Homeowners should not pay higher taxes on 'mansions' valued over £2m
The government should abolish the inheritance tax	The government should keep the inheritance tax
Disposable products (such as plastic cups, plates, and cutlery) that contain less than 50% of biodegradable material should be banned	Disposable products (such as plastic cups, plates, and cutlery) that contain less than 50% of biodegradable material should not be banned
I support the death penalty	I do not support the death penalty
The government should require children to be vaccinated for preventable diseases	The government should not require children to be vaccinated for preventable diseases
Social media companies should ban political advertising	Social media companies should not ban political advertising
Convicted criminals should have the right to vote	Convicted criminals should not have the right to vote
The government should regulate social media sites, as a means to prevent fake news and misinformation	The government should not regulate social media sites, as a means to prevent fake news and misinformation
I support the use of hydraulic fracking to extract oil and natural gas resources	I do not support the use of hydraulic fracking to extract oil and natural gas resources
The British Monarchy should be abolished	The British Monarchy should not be abolished
The government should increase spending on public transportation	The government should not increase spending on public transportation
Citizens should be allowed to save or invest their money in offshore bank accounts	Citizens should not be allowed to save or invest their money in offshore bank accounts

The UK should abolish the Human Rights Act?	The UK should not abolish the Human Rights Act?
The government should increase foreign aid spending	The government should decrease foreign aid spending
I support the use of zero hour contracts	I oppose the use of zero hour contracts
The U.K. should raise taxes on the rich	The U.K. should not raise taxes on the rich
I agree with the UK’s Brexit decision to withdraw from the European Union?	I disagree with the UK’s Brexit decision to withdraw from the European Union?
It should be illegal to burn the UK flag	It should not be illegal to burn the UK flag
I am in favour of decriminalising drug use	I am opposed to decriminalising drug use
The government should increase military spending?	The government should decrease military spending?
Terminally ill patients should be allowed to end their lives via assisted suicide	Terminally ill patients should not be allowed to end their lives via assisted suicide
Businesses should be required to have women on their board of directors	Businesses should not be required to have women on their board of directors
The government should be able to monitor phone calls and emails	The government should not be able to monitor phone calls and emails
I would support the return of a selective education system and the reintroduction of grammar schools	I would not support the return of a selective education system and the reintroduction of grammar schools
I support the use of nuclear energy	I do not support the use of nuclear energy
The London Underground should be considered an 'essential service' which would ban all future worker strikes	The London Underground should not be considered an 'essential service' which would ban all future worker strikes
The government should increase environmental regulations on businesses to reduce carbon emissions	The government should not increase environmental regulations on businesses to reduce carbon emissions
The UK should switch to a proportional representation voting system	The UK should not switch to a proportional representation voting system
The government should enact a stricter immigration policy	The government should not enact a stricter immigration policy
Labor unions help the economy	Labor unions hurt the economy
The national railway should be privatised	The national railway should not be privatised
There be more privatisation of the NHS	There be less privatisation of the NHS
The UK should renew its Trident nuclear weapons programme?	The UK should not renew its Trident nuclear weapons programme?
The government should make cuts to public spending in order to reduce the	The government should not make cuts to public spending in order to reduce the

national debt	national debt
My stance on abortion is pro-life	My stance on abortion is pro-choice
The UK should abolish university tuition fees	The UK should not abolish university tuition fees
The minimum voting age should be lowered	The minimum voting age should not be lowered
Foreign visitors should not have to pay for emergency medical treatment during their stay in the UK	Foreign visitors should have to pay for emergency medical treatment during their stay in the UK
Every 18 year old citizen should be required to provide at least one year of military service?	Every 18 year old citizen should not be required to provide at least one year of military service?

Value Stimuli for Studies 5 and 6

The values questions were adapted from questions used in the Hogan Motives, Values, and Preferences Inventory (Hogan & Hogan, 1996).

Art and literature are the highest forms of expression in life
I dislike people who think that because something is expensive it must be tasteful
I would like to be a writer
In my spare time I like to go to art museums or listen to classical music
I go to a lot of parties with my friends
It is important to stay in close contact with your friends
I enjoy group projects and working with others
I like to socialise and network with others
I like to spend my spare time helping others
Most of my friends help others who are in need
I never judge other people’s actions.
I value long-term relationships that lead to strong friendships
People are primarily motivated by money
I would like to be in business for myself
I don't like people who can't live within their means
A company’s main focus should be profits
I don’t like serious, strait-laced people
If I could afford it, I would spend my life taking holidays
The principal goal of life is enjoyment
I prefer creative, free-spirited people as my friends
Even in my spare time I like challenge and competition
I enjoy being in charge
The goal of life is to compete at something important and succeed
I worry about how my friends’ reputations will reflect on me
I would like a job that puts me in the public eye

I would like to associate with people who are famous
Organisations should make sure their star players get the best treatment
It is important to get individual recognition for work you do.
My friends keep up with recent advances in science
I don’t understand people who ignore data and facts
I believe progress is only possible through scientific research
I am really interested in how things work
Job security is more important than job satisfaction
I am not a thrill seeker
I don’t like unpredictable people
I try to live by the motto “look before you leap”
I know immediately when I have done something morally wrong
Even if something better comes along, I don’t like changing the way I do things
I am extremely careful in choosing the people with whom I associate
I dislike it when people break with established traditions