UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

DOPAMINERGIC GENES PREDICTIVE OF UNMOTIVATED CONFIRMATION BIAS ARE NOT PREDICTIVE OF MOTIVATED CONFIRMATION BIAS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

DOUGLAS J. BRYANT
Norman, Oklahoma
2020

DOPAMINERGIC GENES PREDICTIVE OF UNMOTIVATED CONFIRMATION BIAS

ARE NOT PREDICTIVE OF MOTIVATED CONFIRMATION BIAS

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF PSYCHOLOGY

BY THE COMMITTEE CONSISTING OF

Dr. Andrea Vincent, Chair

Dr. Mike Buckley

Dr. Lauren Ethridge

Dr. Erin Freeman

Dr. Cecil Lewis

Dr. Michael Markham

TABLE OF CONTENTS

Abstract

Confirmation bias is persistently devastating to rational judgment and decision-making. Previous research supports cognitive and behavioral distinctions between two types of confirmation bias: motivated confirmation bias and unmotivated confirmation bias. Motivated confirmation bias is a member of the larger class of motivated reasoning biases. These often occur when one's individual or group identity is tied up in certain beliefs or propositions that command one's assent. Prior research has shown that even when individuals possess cognitive problem-solving skills such as high numeracy, these skills offer no benefit to rational thinking or judgment in the face of motivated reasoning problem sets. Prior research has also shown that dopaminergic genes DRD2, DARPP-32, and COMT are predictive of susceptibility to unmotivated confirmation bias; however, the role of these genes in motivated confirmation bias had yet to be tested. The present investigation examined the possible connection. Participants were 200 university students who completed questionnaires and tasks assessing motivated confirmation bias, numeracy, political philosophy and party identification. Logistic regression modeled the association of these measures with accuracy on a bias detection task. Numeracy predicted accuracy; however, genotypes and political measures did not. These results suggest that distinct genetic determinants are responsible for motivated and unmotivated confirmation bias. Further, the findings replicated previous research demonstrating that accuracy is much diminished in the motivated scenario compared to an unmotivated control. However, contrary to this earlier work, the current findings suggest that numeracy confers a benefit in both motivated and unmotivated conditions, rather than just in motivated situations. Overall, these findings suggest continued research is needed to uncover the neurobiological determinants of motivated confirmation bias.

Keywords: Motivated confirmation bias, DRD2, DARPP-32, COMT, numeracy.

Dopaminergic Genes Predictive of Unmotivated Confirmation Bias are Not Predictive of Motivated Confirmation Bias

Biases and heuristics are cognitive shortcuts or tools we use to navigate complex environments. They are often used when there is not enough information available to make a more educated decision, not enough time to weigh out evidence for or against, not enough severity of consequences to justify thinking over testing, or not enough likelihood that one's belief is wrong to justify reexamining it. One of the two key features that make cognitive biases and heuristics so useful and often used is that they are highly reliable—i.e., they usually lead us to the correct solution. The other key feature behind their utility is that they are fast—even automatic. As a consequence, and added benefit, biases and heuristics require less cognitive resources than deliberative consideration of alternatives.

These and other benefits of cognitive biases and heuristics are often overlooked because they are typically studied in the contexts in which they go wrong—when they lead to suboptimal beliefs, decision-making, and outcomes. While cognitive biases and heuristics usually reach optimal outcomes, they also fail in certain cases where the outcome is substantial. For this reason, researchers in the judgment and decision-making field of cognitive psychology have sought to discover ways in which to overcome biases and stop reliance on heuristics where circumstances dictate. These are typically circumstances in which the outcome of a decision is significant enough to warrant the work and resources of more deliberative thought.

Strategies and Interventions for Augmenting Cognition

There are two main approaches to augmenting cognition in the face of biases and heuristics: training- or teaching-based interventions meant to change the way one thinks, or instead, accepting the way one thinks and redesigning user interfaces and information

presentation to reach different outcomes. This latter approach is a human factors approach. It uses cognitive psychology to design products and systems that optimize user behavior. One example of this approach is seen in the presentation of health risk information (see Garcia-Retamero and Cokely, 2013 for review). Garcia-Retamero and Cokely (2011) developed an intervention (an informational brochure) to augment sexually transmitted infection (STI) risk awareness. Inaccurate risk perceptions are often due to cognitive biases and heuristics, including the availability heuristic, representativeness heuristic, anchoring, and framing effects, among others. The availability heuristic is used in this context when one judges the likelihood of an outcome based on the ease with which examples come to mind (Schwarz *et al.*, 1991). The representativeness heuristic is used when one judges that the likelihood of an outcome is predicted by outcomes of a similar type or outcomes that followed similar antecedent conditions, e.g. thinking that a medical condition will improve on its own because previous medical conditions did so (Read & Grushka-Cockayne, 2011). Anchoring bias occurs in the context of risk perception when perceptions of present and future risks are biased by information about the frequency and severity of past risks (Lieder, *et al.*, 2018). Such information need not be first-hand and may even be false. Framing effects bias risk perception when information is provided such that it limits perceptions of, and responses to, a problem (Druckman, 2001). Framing effects were famously demonstrated by Tversky and Khaneman (1981) with a contagious disease response paradigm. Participants read a problem that stated the United States would soon face an outbreak of an Asian disease and 600 fatalities were expected. Participants then chose one of two response programs based on their expected outcomes. These responses were presented differently in two conditions: a “saved” condition, and a “die” condition. In the “saved” condition, participants read that under program A, 200 lives would be saved, while under

program B, there was a 1/3 probability that all 600 people would be saved, and a 2/3 probability that none would be saved. In the "die" condition, participants read that under program C, 400 people would die, while under program D, there was a 1/3 probability that no one would die, and a 2/3 probability that all would die. Across the two conditions, Programs A and C, and Programs B and D, were probabilistically equivalent, e.g. 200 people would be saved (Program A), 400 people would die (Program C). However, in the "save" condition, 72% of participants chose Program A, whereas in the "die" condition, only 22% of participants chose the equivalent program: Program C. The "save" condition was framed such that it promoted risk-averse responses. The "die" condition was framed such that it promoted risk-taking responses.

Similarly, Garcia-Retamero and Cokely (2011) showed that framing effects had different consequences on sexual behavior; specifically, gain-framed messages were more successful in fostering STI preventative behaviors, whereas loss-framed messages were more successful in fostering illness-detection behaviors like health screenings. More importantly, they showed that the two framing conditions could be made not only *equally*, but also *more* behaviorally effective simply by adding visual aids. Other successful interventions are often similarly simple. For example, Hales and Pronovost (2006) demonstrated the utility of procedural checklists in error avoidance in the face of cognitive challenges exacerbated by stress and fatigue. They can help, for example, in overcoming inaccurate initial diagnoses that lead to confirmation biases from which information contradicting the initial diagnosis is ignored, and confirmatory information is sought (Mendel *et al*., 2011).

A more controversial intervention to mitigate cognitive bias is incentivizing certain behaviors. Smith and Walker (1993) incentivized normative or neutral (Nash equilibrium) auction bids with a payoff of $250.00. They found that the payoff opportunity increased the

amount of cognitive work participants undertook to reason out the normative bid. In a review of monetary incentives in experimental research, however, Camerer and Hogarth (1999) found the opposite, as well as that such incentives introduced several experimental confounds. Hogarth *et al*. (1991) also found that financial incentives merely encourage people to pursue their strategies, biased or otherwise, with greater resolve.

Another type of cognitive bias intervention is the social influence or "nudge" approach advocated by Thaler and Sunstine (2009) wherein one seeks to change biased and incorrect beliefs and their resulting behaviors. An example of this approach is influencing beliefs and behaviors about smoking by implementing public messaging that suggests the majority of people do not smoke and that even those who do smoke want to stop smoking (Marteau *et al*., 2011). Other examples from Marteau *et al*. (2011) include countering over-consumption of alcohol by using smaller serving glasses at restaurants and bars, and countering obesity by making side salads, rather than French fries, the regular side item. In the financial compliance domain, Castro and Scartascini (2013), sought to improve rates of tax compliance in a developing municipal economy in Argentina by modifying beliefs in three areas: levels of enforcement of tax compliance, level of equity in tax compliance, and levels of fairness in the tax system itself. They found no effect with public messaging about equity and fairness; however, they discovered that simply informing municipal residents of the legal consequences and associated fines for noncompliance increased compliance significantly. Altering beliefs in this way is a double-edged sword. While nudging can be used to overcome false and biased beliefs, it can also be used to create false and biased beliefs, even if for a common good, thereby raising ethical questions about the approach (Nys & Engelen, 2017). While young smokers' risk perceptions about smoking are likely inaccurate and biased by availability and similarity heuristics, nudging might

counter those by creating opposing biases beliefs, for example, that smokers are socially undesirable.

Training-based interventions are more varied. They include training in logic, statistics and bias awareness among many others (Babcock & Loewenstein, 1997; Fischoff, 1982; Niu *et al*., 2013). One of the hurdles for training interventions is that individuals and organizations can be resistant to debiasing efforts (Arkes, 2003). Most training regimens show improvement in cognitive performance in the short term and immediate context; however, they usually run into the same two hurdles: gains are unlasting and context dependent (Fong & Nisbett, 1991). In studies where participants are retested weeks or months after training, gains have either significantly diminished or returned to pre-training levels. The second challenge is that of domain transfer: gains in laboratory or classroom settings are diminished or undetectable when trainees face real-world problems (Hogarth, 2001; Kagel & Levin, 1986). Subjects with training in critical thinking show little improvement in normative thinking (Niu *et al*., 2013). Mowen & Gaeth (1992) reported slight improvement in participants who were trained in the cognitive mechanisms underlying biased decision making; however, as with critical thinking, the material proved difficult to teach to nonexperts.

One approach that avoids the domain transfer problem is the use of domain-dependent strategies and cognitive "tricks" like using the wisdom of the crowd to inform one's judgment (Mannes *et al*., 2012). This only works in circumstances where the crowd consists of members who have relevant pieces of information about the problem, all of which will point toward the same solution, and false beliefs about the problem which point in many different directions, all of which will cancel one another out (Atanasov *et al*., 2017). Another such trick applies to the pick-a-door problem. When given a problem in which one must choose between three or more

options, for example, when one chooses Door Number One, Two, or Three, and has no basis on which to select one over the other, a person can improve their odds of successfully choosing the "correct" door merely by choosing one door and then changing one's answer (Herzog & Hertwig, 2009). This is counterintuitive, but it works because the likelihood that the first choice is correct is only 33%; while the likelihood the correct door is one of the other two doors is 67%. Thus, one necessarily improves one's odds by changing the original choice.

Aczel *et al*. (2015) showed that training in analogical thinking was beneficial in combating the domain transfer problem. Analogical thinking is the use of one problem with a known solution as an analogy by which to solve another problem with an unknown solution. In their experiment, participants were placed in three conditions: a no training control group, a bias awareness training group, and an analogical thinking training group. Only the analogical training group showed significant improvement over pretraining testing when retested one month later. The retest covered the same biases as the pretest; however, it set those biases in different domains than those used on the pretest. The authors reasoned that analogical reasoning was effective in addressing domain transfer because it requires an understanding of the underlying logic that is common across problems in distinct domains. Vendetti *et al*. (2014) relied on the same reasoning in developing an experimental intervention that used analogical reasoning problems themselves as a tool to promote a broader analogical reasoning mindset to overcome limitations in domain transfer. As with Aczel *et al.* (2015), improvements were statistically significant and effect sizes were modest. In sum, numerous interventions to counter cognitive biases have been developed and tested without great success. This suggests some deeper understanding of the cognitive and neurological mechanisms underlying biases may be warranted.

Confirmation Bias

Confirmation bias is a much-studied cognitive bias that is persistently devastating to rational judgment and decision-making. Cognitively, it can take many forms, most commonly noticing and remembering confirmatory information while not noticing or not remembering contrary information. Behaviorally, confirmation bias is seen in how a person differentially interacts with confirmatory and contrary information. In consuming highly polarizing political information, for example, it is seen in seeking out information that confirms one's beliefs and avoiding information that contradicts one's beliefs. Interactions with confirmatory information can feel pleasurable, while interactions with contrary information can feel painful. Neurobiologically, imaging studies show that interactions with information that changes strongly held political beliefs are accompanied by activity in self-monitoring and emotion related areas (Kaplan *et al.*, 2016).

There is not a single, agreed upon, definition of confirmation bias, and some experts use the term to refer to a class of biases (Nickerson, 1998). Charness & Dave (2017) give a threefold definition of confirmation bias that includes seeking, interpreting, and using information to support one's prior beliefs. Yariv (2002) includes the phenomenon in which people update their beliefs in order to view their past actions more favorably. Nickerson (1998) includes a variety of phenomena including hypothesis-driven information seeking and interpretation, restriction of attention to a favored hypothesis, treating confirmatory evidence preferentially, and overweighting confirming evidence, among others. Nickerson (1998) also argues that unwitting or unconscious confirmation bias is an essential component of the concept (Fischoff, 1982). Yet, while many studies demonstrate unconscious confirmation bias; awareness of the bias does not appear to diminish it.

Klayman (1995) describes a class of biases held together by a common propensity of over-belief in one's preferred opinions. For Klayman, the class of confirmation bias includes the specific biases of positive hypothesis testing (Klayman & Ha, 1987), wherein one searches for belief-confirming evidence, and the availability heuristic, wherein the search of one's own memory more readily produces instances that confirm expectations. Importantly, search styles, like searching for confirming evidence rather than disconfirming evidence, are not necessarily biased or suboptimal in many circumstances (Friedrich, 1993). Nevertheless, they can produce bias when one stops testing examples and concludes that the positive cases are sufficient to support the desired conclusion. This phenomenon can be observing in the rule discovery paradigm first used by Wason (1960). Subjects were given sets of three numbers or "triples" (e.g. 2, 4, 6) and asked to identify the rule to which the numbers adhered, much like a pattern recognition question on a standardized test. Subjects tested their hypotheses by proposing additional triples and an experimenter indicated whether the new triple obeyed the rule. Subjects stopped when they were highly confident that they had discovered the rule. Wason found that the participants tended to propose triples that fit their hypothesized rule. In doing so, they sought confirmation of their hypotheses, and they expected that the triples they proposed would fit the rule. This strategy, also called positive testing, was suboptimal, not least because Wason's rule was broader than most suspected—namely, any sequence that increases. Any one triple (e.g. 2, 4, 6), however, conforms with many rules: the numbers are positive, the numbers are single digit, the numbers are even, the numbers increase by two, etc. Thus, a participant might propose several examples, find that they all are correct instances of their hypothesized rule, and then falsely conclude that they have enough evidence to identify the rule. In this case, participants often propose examples such as 8, 10, 12. The experimenter confirms that these numbers fit the

rule, and so the participant concludes the rule is “increases by two.” The more general the rule, the more difficult it is to discover it by positive testing alone. Rules such as “the numbers are Arabic numerals, the numbers are numbers, the numbers are any three things and needn’t be numbers are all,” would be very difficult to discover using only positive tests as one’s search strategy.

Some authors consider positive hypothesis testing to be an example of confirmation bias because in using it, the individual is attempting to confirm their own beliefs or bias. Klayman & Ha (1989) among others, argue that the use of positive hypothesis tests does not necessitate that one is trying to “prove” the tested hypothesis. Subjects have even used positive hypothesis tests to seek to disconfirm a hypothesis, as in attempting to show the any-ascending-sequence rule as false by offering an extreme triple like “-100, 0, 105” (Klayman & Ha, 1989). In their view, it is not the positive test search strategy only that makes for confirmation bias; rather, it is the conclusion that one has sufficient information from this search strategy to stop collecting data.

Positive hypothesis testing is, nevertheless, often an inferior strategy because it can only uncover false positives—triples one expects to work but which fail. Negative hypothesis testing or seeking to disconfirm, is usually a more optimal strategy because it can uncover false negatives—triples one expects to fail but which work. This reveals more information about the underlying rule, and it does so more quickly and efficiently (Klayman, 1995).

The feature that defines an optimal strategy is the environment; specifically, whether the environment is one in which false positives or false negatives are more consequential and more in need of discovery. A common example of a case where false positives are more consequential and hence positive testing is more optimal, is that of car buying. Buying a lemon is a greater risk than missing out on reliable cars (Friedrich, 1993).

Klayman (1995) also argued that the well-educated are not immune from instances of confirmation bias. In a study by Lord, Ross, and Lepper (1979), death penalty proponents and adversaries evaluated the same body of evidence and each concluded that the evidence supported their positions. More, the evidence used in the study, which was mixed, further entrenched each side. Both appeared to seize on the confirming pieces alone and disregard information that detracted from their views. Plous (1991) made a similar finding with advocates and detractors of new technologies when interpreting a technology's performance data. Skeptics of a new technology interpreted the performance data as further evidence for their skepticism of the technological product. Advocates of the new technology interpreted the same performance data as supporting their own view.

Koehler (1993) made a similar finding among scientists. In Koehler's study, the participants were a group who advocated for the scientific study of parapsychology and another group that aimed to refute parapsychological claims. The participants were tasked with evaluating the methodology of a number of studies that reached different conclusions about parapsychology—a matter logically distinct from the claims themselves. Nevertheless, each side tended to more readily approve of the methodologies of studies whose conclusions were consistent with their beliefs, and more readily criticized the methodologies of studies that reached conclusions that opposed their beliefs. The study is an example of a more general phenomenon; scientists are less skeptical of evidence that supports their beliefs than they are of evidence that is inconsistent with their beliefs. For example, professional audiences who read peer-reviewed studies do not equally discount studies in which the data are believed to be flawed in some way. Rather, confirmatory flawed data are more readily believed than flawed data that contradicts one's prior beliefs (Gorman, 1986, 1989).

Two Types of Confirmation Bias

Prior research supports cognitive and behavioral distinctions between two types of confirmation bias: motivated confirmation bias and unmotivated confirmation bias (Kunda, 1990; Nickerson, 1998). Motivated confirmation bias is a member of the larger class of motivated reasoning biases (Haidt, 2001; Kunda, 1990). These often occur when an individual or group identity is tied up in certain beliefs or propositions that command assent. However, confirmation bias can operate independently of any of the properties of motivated reasoning, such as the desire to prove or defend a hypothesis or fear of a competing hypothesis. Both Wason's rule discovery task (1960) (discussed above) and card selection task (1968) serve as examples of a common preference for positive or confirmatory test strategies over disconfirming strategies. However, this occurs in the absence of a motivated reasoning component. Even though subjects who used positive testing as a search strategy sought to confirm a specific hypothesis, they had no loyalty to such a hypothesis, and the failure of the hypothesis was no threat to their identity, broader goals, or values. Thus, the rule discovery task shows that confirmation bias need not be motivated to prove a hypothesis in order to count as confirmation bias. This is true even if one agrees with Klayman and Ha (1989) that use of the positive test strategy alone is insufficient to count as confirmation bias, and that some conclusion about having made a sufficient search to discover the rule is required.

More recent investigations of confirmation bias also demonstrate that it need not take the motivated form. In fact, unmotivated confirmation bias is often studied because it is easier to induce and control experimentally. Work from Doll *et al*. (2011) used a probability task in which playing cards with novel symbols were played head-to-head. Participants were instructed to select between two playing cards the one with the highest probability of being the winning card.

Cards had an assigned probability of winning prior to the task, yet participants were given inaccurate instructions about which cards were most likely to win. Although feedback from each trial was immediate—that is, the participants saw whether the cards they selected won or lost—some participants did not learn the true probabilities observed via these outcomes and instead persisted in complying with the erroneous instructions. Other participants were successful at learning throughout the task despite the instructions and adjusted their play to achieve better outcomes. In other words, some participants were more susceptible to confirmation bias than others. However, none of the participants in this task had any vested interest in proving that certain cards turn out to be more probable winners than others. The outcome did not engage any emotional commitments such as their individual or group identities.

Somewhere between these extreme cases of unmotivated and motivated confirmation bias are cases where motivation creeps into the task process. Scherer *et al*. (2012) demonstrated a phenomenon known as post-prediction selection bias, wherein making a prediction biases subsequent information searches. Scherer *et al*. (2013) expanded on this and showed that a completely arbitrary hypothesis, on a topic on which a person knows nothing, is also sufficient to bias future information searches. An arbitrary hypothesis can become a preferred hypothesis motivating confirmation bias merely by the participant predicting that the hypothesis is true. In their experiment, participants viewed two paintings, displayed side-by-side, and then indicated which painting's *original form* they thought was better liked by college students. The "original form" description was included to make the choice even more arbitrary. Participants then selected articles to read from a list provided by the experimenters. As expected, participants selected articles consistent with their prediction. Since the predictions were arbitrary and uninformed, the two options are said to be hedonically neutral, *i.e*. unmotivated. They further

argued that motivation for bias in this case was merely the desire to be correct in one's prediction.

Similarly, though perhaps more easily classified as motivated confirmation bias, is the phenomenon in which people seek to support and defend the beliefs and actions of their past selves (Hart *et al*., 2009). In doing this, people seek to view their past selves in a positive light and may even seek to avoid confronting the consequences of what would occur if they were wrong. This may take the form of selecting sources of information that support prior decisions such as taking one job over another or purchasing one house over another. More consequential are failures to confront information that contradicts prior decisions such as political decisions and civil actions (Bronfman *et al.*, 2015; Lerman & Acland, 2020), wrongful convictions (Rossmo & Pollock, 2019), conclusions in strategic intelligence analysis (Whitesmith, 2019), and medical diagnoses (Elston, 2020).

A Social Conception of Motivated Confirmation Bias

More recently, researchers in a variety of disciplines including social psychology and communications have studied motivated confirmation bias from a social perspective. Several studies have examined the effects of modern, individualized, news consumption and how news is shared within ideologically homogeneous groups (Athey *et al*., 2017; Duffy & Ling, 2020; Masta & Shearer, 2018; Törnberg, 2018), both of which can lead to confirmation bias (Garrett, 2017). Del Vicario *et al*. (2016, 2017) go further, finding that information consumers congregate into like-minded communities that nurture confirmation bias. In legacy forms of news delivery such as evening news broadcasts and physical newspapers, consumers were exposed to more heterogenous messages. Exposure to diverse perspectives has been shown to moderate bias (Guilbeault *et al*., 2018). Ling (2008, 2014) showed that news consumption via mobile

communication devices, such as smart phones, intensifies social cohesion and likely also exacerbates information silos. Ling (2020) argued that both cognitive and social conceptions should be combined in future studies of confirmation bias. Some researchers have taken this approach, using social network analysis and cognitive measures (Gašević *et al*., 2019; Houghton *et al*., 2015). However, as Ling (2020) notes, these studies can be methodologically difficult, not least because motivated confirmation bias is difficult to measure and control. Nevertheless, this is precisely the type of confirmation bias that has real world consequences.

The Perils of Motivated Confirmation Bias

Motivated confirmation bias has extensive public and private costs. For example, it affects the public understanding of science on a variety of polarized issues. These include vaccination (Meppelink *et al*., 2019), climate change (Druckman, 2015), gun control (Kahan, *et al*., 2017), and COVID-19 (Garcia-Alamino, 2020), among others. Since people can select and cultivate not only their own news sources, but also their own health and science information sources, their choices are easily influenced by confirmation bias. Worse, these choices can further entrench false beliefs. Knobloch-Westerwick *et al*. (2020) showed that people are selective about the news and information to which they expose themselves and tend toward that which agrees what they already believe. Meppelink *et al*. (2019) reached the same conclusion with regard to beliefs about vaccinations. In their study of nearly 500 parents of small children, parents indicated their beliefs about vaccination through an online survey and then selected from a reading list of 10 articles on vaccination. Unsurprisingly, people chose to read articles consistent with their beliefs. Participants also evaluated two preselected readings (one pro-vaccination, one anti-vaccination) on three dimensions: credibility, usefulness, and convincingness. Unsurprisingly again, participants rated the reading that was consistent with

their beliefs higher on all three aspects. More worrisome, Meppelink *et al.* (2019) also measured participants' health literacy using the Newest Vital Signs (NVS: Fransen *et al.*, 2014), a validated health literacy measure, and found that the most health literate participants were also the most biased in their reading selections and perceptions.

Research has demonstrated a number of factors that are correlated with biased information search and selective exposure across a variety of contexts including health literacy and others. These include dogmatism, authoritarianism, and anxiety (Hart *et al.*, 2009), less positive affect, higher need for cognition, and greater cognitive reflection (Knobloch-Westerwick *et al.*, 2020). Two of particular importance to the current study, problem solving skills and facility with numerical information, are discussed below.

Problem Solving Skills Can Exacerbate Confirmation Bias

Prior research has shown that even when individuals possess cognitive problem-solving skills, these skills may not offer any benefit to rational thinking or judgment in the face of motivated reasoning problem sets (Kahan *et al.*, 2017; Meppelink *et al.*, 2019). Kahan *et al.* (2017) demonstrated differences in rationality on a motivated confirmation bias detection task. In the task, participants interpreted a data table and indicated what conclusion it supported. In the control, *i.e.* unmotivated, condition, participants were told the data table was about the effectiveness of a rash treatment cream. In the treatment, *i.e.* motivated, condition, participants were told the data table was about the effectiveness of a ban on carrying concealed weapons to reduce crime. Apart from these instructions, the data tables, and hence the numerical tasks, were identical. Kahan demonstrated two major findings. First, numeracy, or facility with numerical information (an analog of literacy), was predictive of accuracy in the control condition, but not in the treatment condition. This finding suggests that the advantage in reasoning that numeracy

typically conveys is not operable in motivated reasoning scenarios. Second, in the treatment condition, accuracy was predicted by whether the data table a given participant interpreted was one that was congruous with that participant's political beliefs. Thus, if the data table showed that the gun control measure reduced crime, liberals were more likely than conservatives to interpret the table correctly. Conversely, if the data table supported the conclusion that the gun control measure was ineffective, conservatives were more likely than liberals to interpret it correctly.

Kahan *et al.* (2017) designed the experiment to evaluate two competing hypotheses: the science comprehension thesis (SCT) and the identity-protective cognition thesis (ICT). The SCT holds that people fail to understand scientific findings because they lack the intellectual ability to do so. The SCT would hold, for example, that many people reject the scientific evidence for climate change because climate science is complex and difficult to understand. The ICT, on the other hand, holds that most of the science presented to the public at large can be readily understood by its audience. Instead, the more determinative factor in whether an individual accepts or rejects scientific claims is whether those claims threaten one's individual or group identity. The ICT would hold, for example, that many people reject the scientific evidence for climate change because they see those claims as threatening the truth of their cultural, religious, and/or political beliefs. ICT is a hypothesis about motivated confirmation bias; it holds that the public's rejection of scientific claims is a defensive response to perceived threats. When confirmation bias is at work, individuals avoid such threatening information and instead seek out information that confirms their prior beliefs (Nickerson, 1998).

Kahan's findings supported the ICT hypothesis; while numeracy was predictive of accuracy in the control condition, with higher numeracy increasing the likelihood of a correct

response, the advantage conveyed by numeracy was not detectable in the treatment condition. Consistent with the ICT hypothesis, accuracy in the treatment condition was predicted by participants' political identities (party affiliation and political philosophy) and whether the information participants were judging confirmed or contradicted their political outlook.

The Neurobiology of Confirmation Bias

Kappes *et al*. (2020) showed that people are differentially sensitive to the strength of others' opinions, appropriately moderating their own views when others' opinions confirm their own beliefs, but not so when others' opinions contradict their own beliefs. While climate scientists' belief in climate change has strengthened over time, for example, the percentage of the U.S. population that believes in climate change has simultaneously decreased (Funk & Kennedy, 2016). Kappes *et al*. (2020) argued that this phenomenon is due to differences in the posterior medial prefrontal cortex, where neural sensitivity is reduced in the face of disconfirming opinions.

In a neuroimaging study, Kaplan *et al*. (2016) presented 40 politically liberal participants with arguments that contradicted their political positions, as well as arguments that contradicted some non-political positions. When participants viewed political counterevidence, as compared to non-political counterevidence, neural activity was increased in the precuneus, posterior cingulate cortex, medial prefrontal cortex, inferior parietal lobe, and anterior temporal lobe. By contrast, when viewing arguments that countered their non-political beliefs, participants showed increased neural activity in the dorsolateral prefrontal cortices and orbitofrontal cortices. There were also differences between participants who more strongly resisted counterevidence and those who changed their minds. Belief perseverance was positively correlated with activity in the left dorsomedial prefrontal cortex and negatively correlated with activity in the left orbitofrontal

cortex. Since participants rarely changed their political beliefs in the face of counterevidence, belief change was assessed using non-political arguments. Participants who changed their minds showed less neural activity in two areas: the dorsal anterior insular cortex, and the amygdala.

Talluri *et al*. (2018) examined perceptual sensitivity in the selection and accumulation of post-choice information in a perceptual task. Participants viewed an initial dot-motion stimulus and stated the direction of motion they perceived. They then viewed a second dot-motion stimulus and indicated their perception of the overall direction of motion—their perception of the direction of motion across both stimuli. Participants were more sensitive to motion in the second stimulus when the direction of motion was consistent with their stated prior perception. This was true whether their prior perception was correct or incorrect. Modeling supported the hypothesis that rendering a choice after the first stimulus modulated weighting of sensory neurons so as to increase sensitivity to choice-consistent perceptions.

Using a similar paradigm, Rollwage *et al*. (2020) demonstrated how a better understanding of the neurobiology of confirmation bias can inform interventions to overcome such bias. They examined the selection and accumulation of post-choice information in a perceptual task that involved identifying whether dots moving on a screen were indicative of motion to the left or right. Participants also indicated their level of confidence in their choice. Following exposure to additional evidence indicating the same direction of motion as the first exposure, participants gave a final determination of motion directionality. Participants who were wrong in their initial choice but highly confident were less likely to change their minds after viewing additional evidence than were participants who were wrong but less confident. measured Neural activity was measured using magnetoencephalography (MEG) to develop neural models.

Rollwage *et al*. concluded that high-confidence choice selection alters later neural processing of additional evidence that privileges choice-consistent data.

The tasks used by Talluri *et al*. (2018) and Rollwage *et al*. (2020) were perceptual and thus not ones that would invoke a motivated reasoning bias. Thus, it is possible that the underlying neural mechanisms are at work in both simple, or unmotivated, confirmation bias, and more complex motivated confirmation bias. Moreover, since perception of one's own confidence level is metacognitive—a perception of one's own cognitions—and also an indicator of post-choice confirmation bias, Rollwage *et al*. suggest that metacognitive interventions may be effective in identifying and preventing confirmation bias.

Palminteri *et al*. (2017) showed that prediction error valence—whether positive or negative—has differential effects on learning. Participants were placed into two learning groups in which they answered multiple choice questions. Partial feedback consisting of indicating to participants whether their selected option was correct or incorrect was provided to one group. The second group was given complete feedback which indicated correctness or incorrectness for each possible answer choice. In factual learning, participants were more sensitive to, and learned more readily from, positive prediction errors. Thus, participants were more likely to remember the answer to a question if they thought they answered it incorrectly and were surprised by feedback that they were correct. In counterfactual learning, however, participants were more sensitive to negative prediction errors. In such cases they were more likely to remember an incorrect answer to a question as being incorrect if they did not choose that answer and were surprised by feedback stating they were correct in not choosing it. In either case, while the feedback type biased learning in different directions, participants still showed improved learning when their prior beliefs were confirmed.

Prior research has also shown that dopaminergic genes in the dopamine reward system are predictive of individual differences in susceptibility to unmotivated confirmation bias (Doll *et al.*, 2011). Specifically, the findings of Doll *et al.* support the conclusion that individual differences in dopaminergic genes play a causal role in individual cognitive differences—and these in turn play a causal role in individual behavioral differences.

Doll *et al.* identified single nucleotide polymorphisms (SNPs) predictive of individual differences in susceptibility to unmotivated confirmation bias. These SNPs are located in genes that affect dopamine regulation in the striatum (DARPP-32 & DRD2) and prefrontal processing efficiency (COMT). Striatal dopamine regulation is thought to affect confirmation bias because striatal dopamine is key to reward prediction error, the mechanism that likely goes awry in confirmation bias. While these findings implicate dopaminergic genes as a causal influence in unmotivated confirmation bias, the role of these genes in motivated confirmation bias has yet to be tested.

Reward Prediction Error

A reward prediction error (RPE) is a miscalculation of the likely result of an outcome. Rewards are best understood in their classical conditioning context; they strengthen the probability that a rewarded behavior will increase in frequency. Punishments, by contrast, weaken the probability that a punished behavior will increase in frequency. The reward prediction component of RPE is a prediction about the amount of reward that follows a behavior. This can occur in two ways. The reward can be greater than predicted (positive reward), or the reward can be less than predicted (negative reward). In some cases, the difference between the expected and actual reward is easy to measure. In a gambling task experiment, participants might win a variable number of marshmallows when they select a winning card. The RPE, in this case,

could be measured as the difference between the number of marshmallows expected and the number received. In other cases, predicted reward and actual reward are more difficult to measure. This is the case when the predicted and actual reward are amounts of dopamine released by dopaminergic neurons, or the felt effects of dopamine modulation. If the release of dopamine as a reward is perceived as pleasure or another positive affective state, then RPE might be perceived as a feeling of more or less pleasure than expected. This is how RPEs can be rewarding, encouraging the repetition of a behavior, or punishments, discouraging the repetition of a behavior. The former occurs when the error is an underestimation of the reward (reward is more than expected); the latter occurs when the error is an overestimation of the reward (reward is less than expected).

RPEs are encoded by dopaminergic neurons; however, dopaminergic neurons do much more than RPE. The various roles these neurons play can be distinguished by timescale. The signaling associated with RPE is phasic, occurs approximately 100 ms after stimulus onset, and has a latency period of approximately 150 ms (Schultz, 2017). Other dopamine functions, including reward more broadly, occur over seconds and longer. In addition to these phasic activities, dopaminergic neurons also play an essential role in maintaining homeostatic dopamine levels through tonic dopamine release. This tonic activity is what is disrupted in Parkinson's disease (Schultz, 2017).

Dopaminergic neurons that encode RPE do so by encoding estimated value—typically the averaged value of recent past rewards that followed a specific action (Bayer & Glimcher, 2005). The averaged value is also the expected value, and any prediction error is the difference between the expected average and actual reward. If there is an RPE, the information from the error is used to update a new average for reward prediction. In this way, learning about expected

rewards is encoded. This simple model cannot account for all learning, however, especially learning complex tasks. Consider a game in which a subject is rewarded with a marshmallow after eating a chocolate, and rewarded with a chocolate after eating a marshmallow, but not rewarded if the subject ever eats two chocolates in a row or two marshmallows in a row. The optimal strategy is to switch back and forth between the rewards; however, dopaminergic encoding of RPE can likely only explain an approach or avoidance behavior for each reward, and not a more complex switching strategy (Bayer & Glimcher, 2005). This is because the RPE gives feedback after single actions and not series of actions. It will always signal an approach or avoid response after each action. Combining series of actions requires more complex, higher-order, learning mechanism (Bayer & Glimcher, 2005). Thus, RPEs are optimal for predicting binary (approach, avoid) and immediate probabilistic outcomes.

That dopaminergic neurons play a role in encoding a predicted value, also called a utility value, was worked out theoretically through neurocomputational modeling (Bayer & Glimcher, 2005). This theoretical approach was followed by neurophysiological experiments to test a series of theoretical models. Bayer and Glimcher (2005) found support for the encoding of utility values in reward prediction models through neurophysiological experiments with midbrain dopaminergic neurons. Interestingly, the behavior of these neurons could only account for one type of RPE; namely, the RPE that occurs when a reward is greater than expected (positive RPE). In other words, they did not account for RPEs that occur when a reward is less than expected. Thus, the firing pattern of these neurons that encodes the difference between predicted and actual rewards is only observed when actual rewards are greater than expected. The authors speculated that the encoding of negative RPEs may be accomplished by another system.

More specifically, Bayer and Glimcher (2005) found a correlation between the spike frequency of dopamine neurons and a calculated RPE function that considers the average of the last seven rewards. In their experiment, these rewards were varying amounts of juice given to primates. The trials occurred back-to-back over the course of several minutes. The primates were trained in a delayed saccade task and had to work out what amount of time post-stimulus would achieve a maximum reward. The behavior of these neurons in response to negative RPEs was quite distinct. They rectified—which is to say, their activity followed a ramp function—the importance of which is that their activity was not sensitive to, and did not differentially encode, the exact difference between varying averaged rewards and predicted rewards (Bayer & Glimcher, 2005). Instead, this activity was distinctive of all negative RPEs.

Further evidence that positive and negative RPEs are encoded by different mechanisms comes from human neuroimaging studies in which these processes have been shown to be correlated with activity in different brain areas. A meta-analysis of thirty-five fMRI studies of human subjects by Garrison *et al.* (2013) found activity in the midbrain and ventral striatum correlated with positive RPE. Negative or aversive RPEs were found to correlate with increased activity in a more confined area of the left ventral striatum (Garrison *et al.*, 2013). This, of course, does not imply that the mechanisms for encoding these distinct RPEs are different; it is possible that the mechanisms are the same but carried out by redundant systems in different locations. Thus, dopaminergic neurons in the confined portion of the left ventral striatum may still distinguish between different negative reward prediction areas by spike frequency.

Nevertheless, such compartmentalization appears to be part of an integrated hierarchy of processing structures. Both positive and negative RPEs are correlated with activation of the anterior cingulate cortex (ACC) (Garrison *et al*., 2013), a known point of integration for a variety

of processes from distinct neural substrates. Within the ACC, specific regions show activity that correlates with only positive, only negative, and both positive and negative RPEs. Positive RPEs, for example, showed activation in the pregenual ACC, while both positive and negative RPEs showed activation in the anteromedial ACC. This may be evidence that distinct systems for positive and negative RPEs have inputs into distinct regions of the ACC and the information from these inputs is combined in yet another region of the ACC.

Since RPEs help us to refine our expectations of the environment's response to our behaviors, they are tremendously important in reward-based learning. Unsurprisingly, learning is easily disrupted when the systems responsible for recalculating these expectations are disrupted (Murray *et al*., 2008).

SNPs Associated with Susceptibility to Confirmation Bias

The SNPs examined by Doll *et al*. were in genes that code for the DARPP-32 protein, the COMT enzyme, and D_2 dopamine receptors, all elements involved in dopamine regulation. The polymorphisms, shown to be predictive of individual differences in susceptibility to confirmation bias, likely have their effects on cognition and behavior via RPE calculations in the striatum and, in the case of COMT, possibly PFC dopamine regulation at neurons with striatal projections.

DRD2

Dopamine receptor D_2 (DRD2) is the gene that codes for the D_2 dopamine receptor subtype principally expressed in the striatum (Camps *et al.*, 1989). The DRD2 SNP of interest is C957T (rs6277). Although a synonymous mutation, the resulting mRNA occurs in different conformations for 957C and 957T. 957T mRNA is degraded at a higher rate than 957C mRNA. The result is that T alleles produce less D_2 receptors (Hirvonen *et al.*, 2004, 2005, 2009). Behaviorally, because there is less D_2 affinity, T allele carriers are less sensitive to RPEs in

which dopamine drops below baseline. Their takeaway from negative RPEs is avoidance rather than reward recalculation.

Doll *et al.* (2011) showed that participants with the T allele at C957T were less accurate than C/C homozygotes when given incorrect instructions in a probabilistic card task. This finding is supported by other work showing that T alleles are part of a mechanism responsible for under-learning from punishment (negative RPEs), *i.e.*, diminished learning of information inconsistent with prior beliefs. This occurs because confronting information that challenges one's beliefs causes dopamine levels to drop below baseline, and D_2 receptors are essential for detecting this drop (Doll *et al.*, 2011). In such an occurrence, one experiences a negative RPE; however, one fails to learn from it and recalibrate future reward predictions. Additionally, Frank (2005), Hikida *et al*. (2010), and Shen *et al*. (2008) have demonstrated that instead of recalibrating to better predict future outcomes, individuals with T alleles at C957T show improvements in avoidance learning. Thus, in the face of RPEs, learning still takes place in these individuals; however, what is learned is not a recalibration of probabilistic outcomes that maps onto the world. Rather, it is an avoidance of stimuli that produce the negative feelings caused by the drop in dopamine levels produced by negative RPEs (punishments).

DRD2 variants have been examined in connection with a variety of cognitive functions. Zhang and Zhang (2016) showed associations between multiple SNPs in DRD2 and individual differences in performance at insight problem solving. A number of recent studies have also demonstrated support for a link between DRD2 variants, including at C957T, and neurological and psychiatric disease (Nkam *et al*., 2017). C957T variants are associated with both schizophrenia and cognitive and neurological deficits accompanying schizophrenia (Zai *et al*., 2017), such as diminished working memory capacity (Schwarz *et al*., 2016) and inhibitory

control (Liu *et al*., 2014). For this reason, it was theorized that C957T polymorphisms may play a contributory causal role in schizophrenia; however, a metanalysis of 17 independent studies recently found no link between C957T and the cognitive components of executive function: working memory, response inhibition, and cognitive flexibility (Klaus *et al.*, 2019). Even among healthy adults, however, differences in reward learning performance among T/T homozygotes at DRD2 C957T, and C allele carriers have been observed (Byrne *et al*., 2016).

DARPP-32

The DARPP-32 polymorphism of interest, though synonymous, results in different transcription rates for the DARPP-32 protein leading to more or less available synaptic dopamine (Scheggi *et al*., 2018). The SNP has been widely studied and its two alleles differentially associated with schizophrenia and biopolar disorder (Yoshimi *et* al., 2008), suicide (Feldcamp *et al*., 2008), amygdala volume and anger (Reuter *et al.*, 2009), and neurological connectivity in associative emotional learning (Ćurčić-Blake *et al.*, 2012) among others. DARPP-32 is an intracellular protein encoded in the PPP1R1B gene that has been shown to affect reward learning. When presynaptic dopamine crosses the synapse and stimulates postsynaptic D_1 dopamine receptors, phosphoryl groups in the postsynaptic cell attach to DARPP-32 proteins at two sites, Thr-34 and Thr-75. DARPP-32 is then dephosphorylated by D_2 receptor stimulation. Such protein phosphorylation is a pervasive and essential means of functional regulation; it enables signal transduction that is both fast and reversable, enabling the phasic bursts typical of the dopamine reward system's RPEs, and thereby strengthening synaptic plasticity, and by extension, reward learning (Stipanovich *et al*., 2008; Svenningsson *et al*., 2004). Here, phosphorylation also inhibits protein phosphatase-1 (PP-1) from attaching at these sites and reducing the strength of synaptic connections (Munton *et al.*, 2004). Frank *et al*. (2007, 2009)

found that T alleles at rs907094 are positively correlated with increased learning from positive RPEs. This may occur because T alleles at this SNP are part of a haplotype that is associated with greater DARPP-32 expression (Meyer-Lindenberg *et al*., 2007). Similarly, Doll *et al.* (2011) found that while T alleles at rs907094 promote learning from positive RPEs, they are negatively correlated with learning from negative RPEs.

In contrast to findings from Doll *et al*. (2011), a study by Collins and Frank (2012) found no relationship between DARPP-32 polymorphisms and reinforcement learning. Byrne *et al.* (2016) hypothesized that this difference may be due to differences in cognitive processing associated with differences in the two study designs. Doll *et al*. (2011) used a probability selection task in which participants learned and categorized novel symbols. Collins and Frank (2012) used known categories and category members in their task. Byrne *et al*. (2016) reasoned that the DARPP-32 polymorphisms may play distinct rolls in these two types of learning and categorization.

DARPP-32 also modulates striatal plasticity with different effects on reward learning (Doll *et al*., 2015). DARPP-32 polymorphisms are also associated with individual differences in working memory, prefrontal activity, and differences in model-based and model-free learning (Deserno *et al*., 2015). A recent study demonstrated that DARPP-32 T alleles in the striatum increased model-free learning over C alleles; however, C alleles were predictive of better performance in model-based learning (Doll *et al.*, 2016).

COMT

Catechol-o-methyltransferase (COMT) is an enzyme that degrades catecholamines including extracellular dopamine. The COMT gene codes for the enzyme; however, a nonsynonymous polymorphism, Val158Met (rs4680), has been shown to contribute to individual

differences in dopamine regulation (Gogos *et al*., 1998; Huotari *et al*., 2002; Matsumoto *et al.*, 2003). Val158Met is the most studied of the SNPs examined here. Its alleles are differentially associated with focus during working memory tasks (Heinz & Smolka, 2006), emotional resilience (Smloka *et al.*, 2005), and a variety of findings related to risks of schizophrenia, ADHD, and substance abuse, among others. Differences at this SNP are associated with different basal dopamine levels and, by extension, the availability of D_1 receptors in the prefrontal cortex (PFC) (Slifstein *et al*., 2008). The Met version of COMT operates less efficiently in breaking down synaptic dopamine such that Met carriers have four times as much available dopamine as Val carriers. This leaves more synaptic dopamine available to activate postsynaptic neurons in the PFC. The result is sustained PFC activation and improved executive functioning, working memory for abstract rules (Durstewitz *et al.*, 2010; Durstewitz & Seamans, 2008) and high-order cognitive faculties in reward learning (Frank *et al*., 2007, 2009).

Unlike the SNPs in DRD2 and DARPP-32, the Val158Met SNP in the COMT gene does not have significant effects on striatal dopamine (Gogos *et al.*, 1998); however, there is evidence that the effects of COMT in the PFC play a role in striatal dopamine regulation through neuronal projections from the PFC to the striatum (Krugel *et al*., 2009).

Doll *et al.* (2011) found that Met allele carriers were more susceptible to unmotivated confirmation bias. In a probabilistic card selection task in which participants were given incorrect instructions about which cards were most likely to be winners, Met allele carriers continued to play according to the incorrect probabilities in the instructions and did not switch strategies based on the observed outcomes in the task trials. Doll *et al.* reasoned that this susceptibility to confirmation bias was the downside of the improved working memory associated with Met carriers. While improved working memory might allow these participants to

keep instructions in mind throughout multiple trials with feedback, it is unclear why improved working memory could only hold those instructions in mind, and not the feedback from the trials themselves. Thus, Doll *et al.*'s finding regarding COMT and confirmation bias may require additional or alternative explanation.

Like DARPP-32, COMT polymorphisms also play differential roles in model-based and model-free learning. Doll *et al.* (2016) showed that prefrontal Met alleles predicted better performance on a model-based learning task than Val alleles, such that Met allele carriers outperformed Val/Val homozygotes. They reasoned that Met alleles amplified model-based choice performance by augmenting working memory, a cognitive process which occurs in the prefrontal cortex and is dopamine dependent (Otto *et al.*, 2013).

Despite numerous studies showing an effect of COMT polymorphisms on cognitive abilities, a meta-analysis of 58 studies recently found no such effect (Geller *et al.*, 2017) and concluded the effect was either too small to detect or was not there at all. The study differed from a 2008 meta-analysis (Barnett *et al.*) in that Geller *et al.* only included studies of healthy adults; however, both studies failed to find the much-discussed association between Val158Met polymorphisms and individual differences in cognitive function.

Apart from cognitive functions as traditionally conceived, the COMT Val158Met polymorphism has been implicated in emotional decision-making, a relationship that de Souza Costa *et al.* (2016) showed to be mediated by sex. Using the Iowa Gambling Task (IGT), de Souza Costa *et al.* showed Val/Val homozygotes performed better in the latter half of 100 selections in the task only when those Val/Val genotypes belonged to women. Learning in the IGT typically shows up in later trials because most learning takes place in early trials. Decision-making in later trials is considered emotional decision-making because the task is probabilistic

and the decisions are therefore risky. By contrast, early trials which take place prior to or during early learning are considered to be ambiguous. De Souza Costa *et al*. reasoned that the sex differences may be due to an increased focus on a long-game payoff in men, and an increased sensitivity to individual losses in women. Regardless of sex, Met alleles at Val158Met have been associated with a negativity bias such that resistance to negative emotional states is reduced (Gao *et al*., 2016). These Met alleles are also associated with anxiety disorders (McGrath *et al*., 2004) and depression (Åberg *et al*., 2011).

The Val158Met polymorphism is also predictive of individual differences in susceptibility to framing effect biases (Quinn *et al*., 2019). Specifically, Val carriers were more susceptible to framing effects. Participants answered questions in one frame, then answered the same questions differently framed. Val carriers were more likely to change their responses to the questions when they were reframed. The authors hypothesized that the Met allele, and its accompanying increased dopamine in prefrontal synapses, may enhance consistent choice selection.

Current Study

In order to discover whether the dopaminergic genes predictive of unmotivated confirmation bias are also predictive of motivated confirmation bias, we used the motivated confirmation bias detection task developed by Kahan *et al.* (2017) and its unmotivated control task in conjunction with genotyping the relevant SNPs in DRD2, DARPP-32, and COMT genes. This approach allowed us to rely on a peer-reviewed paradigm for detecting motivated confirmation bias established in a sample of 1111 participants. In doing so, we sought to replicate Kahan's findings as well as to replicate and extend the findings of Doll *et al*. by

evaluating whether motivated bias is associated with the same genes found to be predictive of unmotivated confirmation bias.

Hypotheses

Given that RPE underlies motivated confirmation bias, we hypothesized that the same SNPs identified by Doll *et al*. (2017), DRD2 (rs6277), DARPP-32 (rs907094), and COMT (rs4680), as predictive of individual differences in susceptibility to unmotivated confirmation bias would also be predictive of individual differences in susceptibility to motivated confirmation bias. It was further hypothesized that political identity would predict bias; specifically, that in the treatment condition, participants whose stimuli affirmed their political beliefs would be more accurate, and those whose stimuli contradicted their political beliefs would be less accurate. Political ideology was not expected to have any effect on accuracy in the control (unmotivated) condition. Finally, it was hypothesized that in the control condition, numeracy and accuracy would show a strong, positive, correlation, such that increased numeracy would be associated with improved accuracy. In the treatment condition, however, it was expected that this relationship would not hold. Rather, it was hypothesized that Kahan's findings would be replicated and motivated confirmation bias would wipe out any advantage in accuracy that numeracy would otherwise convey.

It is possible that genotype might interact with numeracy to influence cognitive performance, such as susceptibility to bias. The dopaminergic genes examined here have been associated with various cognitive processes and traits in multiple studies. These include executive functioning such as working memory, response inhibition, and cognitive flexibility, emotional decision-making, and framing-effects-induced bias, among others. However, it is also the case that several of these purported links between the genetic polymorphisms under

examination here, and various cognitive properties, have not shown an effect in meta-analyses. Therefore, while we conducted analyses to detect any such interactions, we did not form a hypothesis regarding them. Similarly, Doll *et al*. (2015) found that COMT polymorphisms at C957T had differential effects on model-based and model-free learning.

Methods

Participants and Procedure

Participants were 200 university students, 128 males and 72 females, mean age = 18.97, who participated in this experiment for class credit. The study protocol was approved by the University of Oklahoma Institutional Review Board and Institutional Biosafety Committee. All participants completed informed consent and HIPAA forms before beginning the experiment. Following randomization to study groups, the participants completed the primary experimental task and other study measures. Participants were randomly assigned to either the control (n = 96) or experimental (n = 104) group which were distinguished according to the version of the primary experimental task that was administered. All study questionnaires were delivered in-person via computer.

The primary experimental task was a motivated bias detection task (MBDT) developed by Kahan *et al.* (2017). The MBDT is a data interpretation task designed to elicit and detect motivated confirmation bias in the treatment condition when compared to the control condition. Control group participants were asked to interpret a data table to determine the effectiveness of a hand cream to treat a skin rash while the experimental group participants were asked to interpret a data table indicating the effectiveness of a concealed-carry weapons ban at reducing crime. Each study group was further randomly subdivided into whether the data table presented in the scenario was in support of the effectiveness or ineffectiveness of the skin cream/weapons ban.

The resulting four scenarios are presented in Figures 1-4. Each participant responded to only one scenario. As shown in the figures, the numerical data and their positions in the table were unchanged across the scenarios but the column labels differed. When properly interpreted on the basis of the data alone, the varying positions of these labels should result in a single "correct" judgment to be made with regard to effectiveness. Accuracy of the resulting decision of the participant served as the primary dependent variable for all analyses.

After completion of this task, participants completed a nine-question numeracy scale and a demographics questionnaire. The numeracy scale combined questions from separate validated numeracy scales such as Weller *et al.*'s (2012) scale and the Cognitive Reflection Test (CRT) (Liberali *et al.*, 2012) and was originally combined and used by Kahan *et al.* (2017). The same scale was used here to replicate Kahan's findings. Like other numeracy scales, Kahan's combined scale measures participants' facility with quantitative data and their ability to apply it to solve practical problems. Use of this measure allows for analyses exploring whether numeracy confers any benefit when motivated reasoning is at play. Each question in the numeracy scale was marked correct or incorrect and participants were scored from 0-9 based on their number of correct answers.

The demographics questionnaire included two questions about political identity. The first was a 7-point party identity Likert scale ranging from strong democrat to strong republican. The second was a 5-point political ideology Likert scale ranging from very liberal to very conservative. These scales were also used in Kahan *et al.* (2017) and were used here to replicate those findings. The purpose of capturing political ideology is to assess differences in accuracy on the MBDT when considering alignment between participants' political outlook and their

randomly assigned table. Kahan found, for example, that data tables were more likely to be interpreted correctly when they confirmed participants' prior political beliefs.

At the conclusion of the computer-based surveys, participants provided saliva samples for DNA analysis. Each participant used one Oragene®•Discover OGR-600 from DNA Genotek. Participants indicated the last time they had any food or drink. They then waited at least 30 minutes from that time before providing a saliva sample. They then washed their hands with soap and water and put on disposable gloves. Once gloved, participants took their saliva collection kits and spit into the funnel until they produced enough saliva to reach the fill line. Once complete, participants handed their samples to the experimenter. The experimenter verified the amount of saliva was sufficient, replaced the tube funnel with a tube cap, shook the capped tube for five seconds, and verified and wrote the participant number on the tube. After handing over their saliva samples, participants disposed of their gloves and washed their hands again with soap and water before departing the lab.

Molecular Biology

Reagents

The prepIT®•L2P (PT-L2P), TaqPath™ ProAmp™ Master Mix and all other reagents were purchased from Thermo Fisher Scientific (Waltham, MA).

Nucleic Acid Extraction

DNA extraction from the saliva samples was conducted in a Biosafety Level II (BSL2) lab at Oklahoma Medical Research Foundation (OMRF). The extraction protocol used was DNA Genotek "Laboratory protocol for manual purification of DNA from 0.5 mL of sample." Saliva samples were mixed by inversion and gentle shaking for 10 seconds and then incubated in a 50°C water bath for one hour. After incubation, collection tubes were wiped clean and 500 μL of

each sample was placed into a separate 1.5 mL microcentrifuge tube. Twenty μL of prepIT®•L2P (PT-L2P) reagent was added to each microcentrifuge tube. The combined solution was then vortexed for five seconds. The microcentrifuge tubes were next incubated on ice for 10 minutes and then centrifuged at room temperature for five minutes at 15,000 x g. 500 μL of supernatant from each tube was then transferred into separate microcentrifuge tubes and the pellets were discarded. 600 μL of room temperature 100% ethanol was added and gently mixed by 10 inversions. After sitting at room temperature for 10 minutes, samples were centrifuged again for two minutes at 15,000 x g. The supernatant was removed and discarded. The DNA pellet was then washed with 250 μL of 70% ethanol. After one minute, the ethanol was removed and discarded. 100 μL of TE solution was added to dissolve the DNA pellet and the tubes were vortexed for 5 seconds. Samples incubated overnight at room temperature and were then stored at 4°C until testing.

Real-Time PCR for DNA Genotyping

DNA genotyping was conducted at the University of Oklahoma's Biology Core Molecular Laboratory. Each sample was tested to determine its genotype for three SNPs: DRD2 gene polymorphism (rs6277), DARPP-32 gene polymorphism (rs907094), and COMT gene polymorphism (rs4680). DNA concentration was established via Qubit fluorometer 2.0 (Thermo Fisher Scientific).

Samples were tested using TaqMan™ SNP Genotyping Assays for each of the three genes and used TaqPath™ ProAmp™ Master Mix (Catalog no.: A30865; Thermo Fisher Scientific). DRD2 (rs6277) genotypes were tested using TaqMan SNP Assay (Identification no.: C_11339240_10; Thermo Fisher Scientific). The assay contained two minor groove binding fluorescent probes, a VIC™-labeled probe, which detected the C allele at C957T, and a FAM™-

labeled probe, which detected the T allele. For each 96-well 10 μL Fast reaction, 5 μL TaqMan® Master Mix and 0.5 μL assay working stock were premixed via vortex and briefly centrifuged. Purified DNA samples were diluted with nuclease-free water and added to the plate wells. Two wells in each plate served as no-call controls. The remaining wells filled with 4.5 μL diluted DNA sample and 5.5 μL assay reaction mix. An adhesive film was added to the plate and the plate was briefly centrifuged.

DARPP-32 (rs907094) genotypes were tested using TaqMan SP Assay MTO Human SM (Identification no.: C_7452370_1; Thermo Fisher Scientific). The assay contained two minor groove binding fluorescent probes, a VIC™-labeled probe, which detected the C allele and a FAM™-labeled probe, which detected the T allele. Assay reaction mix, diluted DNA, and plate preparation followed the same procedure as that for DRD2 above.

COMT (rs4680) genotypes were tested using TaqMan Drug Metabolism Assay (Identification no.: C_25746809_50); Thermo Fisher Scientific). The assay contained two minor groove binding fluorescent probes, a VIC™-labeled probe, which detected the Val allele, and a FAM™-labeled probe, which detected the Met allele. Assay reaction mix, diluted DNA, and plate preparation followed the same procedure as that for DRD2 above.

Samples were amplified on a BioRad CFX96 Touch™ Real-Time PCR Detection System. A pre-PCR plate read was first run to identify background fluorescence. This background reading was subtracted from the PCR runs that followed. For DRD2 and DARPP-32 amplification, samples were heated to 95°C for 10 minutes to activate polymerases, then followed 40 cycles of denaturation at 95°C for 15 seconds and annealing at 60°C for 1 minute. For COMT amplification, samples were heated to 95°C for 10 minutes to activate polymerases, then underwent 50 cycles of denaturation at 95°C for 15 seconds and annealing at 60°C for 90

seconds. Failed reactions were repeated. Genotypes for all samples were determined. Quality control for the genotype calls was established using negative controls to show a zero point for comparison. Samples were also run twice, as a positive control, to ensure each run produced the same results.

Data Analysis

Chi-square and independent samples *t*-tests were used to compare demographic characteristics between study groups. Chi-square tests were also used to compare allele frequencies between treatment and control groups. Binary logistic regression analysis was conducted to examine the effects of genotype, group assignment (treatment or control), and table assignment (effective or ineffective) on the accuracy of judgments made in data table interpretation. Additional logistic regression analyses were conducted to examine secondary aims related to numeracy and political identity. Data analyses were generated using SPSS software, Version 24 of IBM SPSS Statistics for Mac (IBM Corporation., Armonk, NY, USA). Statistical significance was determined at the 0.05 level.

Results

Group Differences

Of the 96 participants in the control group, 55 (57.3%) received the data table which when correctly interpreted indicated effectiveness of the skin cream and 41 (42.7%) ineffectiveness. Similarly, in the experimental group, approximately half of the 104 participants, 47 (45.2%) interpreted a table indicating the gun-control measure was effective and 57 (54.8%) ineffective.

Differences between study groups are presented in Table 1. As shown, there were no significant differences observed between treatment and control groups in age ($p = 0.19$) or

gender ($p = 0.53$). Further, no differences were observed in the frequency of party identity ($p = 0.42$), political philosophy ($p = 0.12$), or numeracy score ($p = 0.32$).

Overall allele frequencies are presented in Table 2. The distribution of DRD2 genotypes was significantly different between treatment and control groups ($p < 0.02$) due to the higher presence of homozygous T/T genotypes in the treatment group (37:17). There were no significant differences in the group distributions of DARPP-32 or COMT genotypes ($p > 0.05$).

Allele frequencies for all three genes were within Hardy-Weinberg equilibrium. The calculated Chi-squared values for differences between observed and expected allele frequencies for each gene were as follows: DRD2 $\chi^2 = .718$, $p > 0.05$; DARPP-32 $\chi^2 = 2.855$, $p > 0.05$; COMT $\chi^2 = 0.064$, $p > 0.05$.

We also examined group differences for a binary reclassification of the alleles according to the presence or absence of a particular allele. For DRD2, groups were created according to presence of the T allele, for DARPP-32 the C allele, and for COMT the Met allele. As an example, for DRD2 the two groups consisted of 1) the non-T (i.e., C/C) homozygous group and 2) a combination of the T/T homozygous group and the heterozygous (C/T) group. No differences were observed in the binary allele frequency between treatment and control groups for any of the three genes (all $p > 0.05$; Table 3).

MBDT Performance Predictors

Only 63/200 (31.5%) participants interpreted the data tables correctly. Participants in the control (i.e., unmotivated) condition were significantly more likely ($\chi^2 = 23.059$, $p < .001$) to answer the primary task question correctly ($n = 46$, 48%) than participants in the treatment group ($n = 17$, 27%).

Primary Analyses

A 3-predictor logistic regression model was used to test differential effects on the accuracy of MDMT task performance of DRD2 genotype, group assignment (treatment or control), and table assignment (effective or ineffective). None of the two- or three-way interactions were significant, thus the model was reduced to the main effects model.

This 3-predictor model successfully predicted accuracy as indicated by a significant overall likelihood ratio, $\chi_2(4) = 25.7$, $p < .001$, $R_2 = 12.1\%$, Nagelkerke's $R_2 = 16.9\%$. However, as shown in Table 4, only group assignment successfully predicted the odds of accurately interpreting the MBDT data table. Consistent with the basic group comparisons presented above, treatment group participants were significantly less likely (OR = .196, 95% CI: .098-.389) to answer correctly ($p < .001$). Neither DRD2 genotype nor table assignment significantly altered the odds of MBDT accuracy (ps > 0.05; Table 4). This analysis was replicated using the binary reclassification of DRD2 according to presence of the T allele. Results were unchanged and again only identified a significant effect of group (Table 5).

A second 3-predictor logistic regression model was used to test differential effects on the accuracy of MDMT task performance of DARPP-32 genotype, group assignment (treatment or control), and table assignment (effective or ineffective). None of the two- or three-way interactions were significant, thus the model was reduced to the main effects model.

This 3-predictor model successfully predicted accuracy as indicated by a significant overall likelihood ratio, $\chi_2(4) = 24.5$, $p < .001$, $R_2 = 11.5\%$, Nagelkerke's $R_2 = 16.2\%$. However, as shown in Table 4, only group assignment successfully predicted the odds of accurately interpreting the MBDT data table. Consistent with the basic group comparisons presented above, treatment group participants were significantly less likely (OR = .216, 95% CI: .111-.420) to answer correctly ($p < .001$). Neither DARPP-32 genotype nor table assignment significantly

altered the odds of MBDT accuracy (*p*s > 0.05; Table 4). This analysis was replicated using the binary reclassification of DARPP-32 according to presence of the C allele. Results were unchanged and again only identified a significant effect of group (Table 5).

Finally, a third 3-predictor logistic regression model was used to test differential effects on the accuracy of MDMT task performance of COMT genotype, group assignment (treatment or control), and table assignment (effective or ineffective). None of the two- or three-way interactions were significant, thus the model was reduced to the main effects model.

This 3-predictor model successfully predicted accuracy as indicated by a significant overall likelihood ratio, $\chi_2(4) = 25.5$, $p < .001$, $R_2 = 12.0\%$, Nagelkerke's $R_2 = 16.8\%$. However, as shown in Table 4, only group assignment successfully predicted the odds of accurately interpreting the MBDT data table. Consistent with the basic group comparisons presented above, treatment group participants were significantly less likely (OR = .211, 95% CI: .108-.411) to answer correctly ($p < .001$). Neither COMT genotype nor table assignment significantly altered the odds of MBDT accuracy (*p*s > 0.05; Table 4). This analysis was replicated using the binary reclassification of COMT according to presence of the Met allele. Results were unchanged and again only identified a significant effect of group (Table 5).

Post-hoc power analyses were conducted for each of these three main effects models and showed that the analyses were underpowered for the observed effect sizes.

Secondary Analyses

Having found that none of the three SNPs in the dopaminergic genes of interest were predictive of accuracy on the MBDT, we examined whether numeracy, political philosophy, or political party affiliation were predictive of accuracy. Since primary analyses indicated a group

effect (treatment or control) on the odds of accurate response, each of the secondary analysis predictor variables were examined separately in models that controlled for group assignment.

Numeracy

We next investigated the possibility of interactions between numeracy and genotype in order to discover if perhaps numeracy differentially acted on genotype to influence behavior. To do so, we created three two-way interaction terms representing the interaction of numeracy with each of the three genotypes and examined each in separate analyses. In these analyses, each genotype included three possibilities: two homozygote and one heterozygote. Since the behavioral effects detected by Doll *et al*. (2011) showed up when isolating certain alleles, we also conducted analyses recoding genotype factor in the interaction into just two options (for example, presence of C allele vs. T/T homozygote). None of these six analyses showed any significance for genotype. Finally, we included sex as a covariate. When we did so, no significant effect was found; however, the significant effect of numeracy that was previously found was lessened. A significant effect for sex can be shown by eliminating numeracy from the analysis; however, numeracy must be included as it accounts for more of the variability in performance than does sex.

Having found no significant interaction terms, these were dropped from the model in order to reduce complexity and increase power. A 2-predictor logistic regression model was used to test differential effects on the accuracy of MDMT task performance of numeracy (scored 0-9) and group assignment. The resulting model successfully predicted accuracy as indicated by a significant overall likelihood ratio, $\chi_2(2) = 38.1$, $p < .001$, $R_2 = 17.3\%$, Nagelkerke's $R_2 = 24.3\%$. Specifically, after controlling for group assignment, numeracy significantly altered the odds of

MBDT accuracy (p = .009); for every point increase in numeracy score, the odds of accurately interpreting the table increase by 38%.

Party Identity

A 2-predictor logistic regression model was also used to test the differential effects on the accuracy of MDMT task performance of party identity and group assignment. Political identity was assessed on a 7-point Likert scale from Strong Democrat to Strong Republican. No group x party identity interaction was indicated. The resulting model successfully predicted accuracy as indicated by a significant overall likelihood ratio, $\chi_2(7) = 29.0$, $p < .001$, $R_2 = 13.5\%$, Nagelkerke's $R_2 = 18.9\%$. However, after controlling for group assignment, party identity did not significantly alter the odds of MBDT accuracy (p = .533).

Political Philosophy

Logistic regression was also used to model the differential effects of political philosophy and group assignment on the accuracy of MDMT task performance. Political philosophy was assessed on a 5-point Likert scale from Very Liberal to Very Conservative. No group x political philosophy interaction was indicated. As with party identity, the resulting model successfully predicted accuracy, $\chi_2(5) = 29.9$, $p < .001$, $R_2 = 13.9\%$, Nagelkerke's $R_2 = 19.5\%$. However, after controlling for group assignment, political identity did not significantly alter the odds of MBDT accuracy (p = .191).

Discussion

Confirmation bias takes two cognitively and behaviorally distinct forms: motivated and unmotivated. Seminal work by Doll *et al.* (2011) and Kahan *et al*. (2017) showed that the influence of motivated confirmation bias is so strong as to erode the benefit of numeric competency in rational thought, and that SNPs important to dopamine regulation predict

susceptibility to unmotivated confirmation bias, respectively. In order to assess whether these cognitively and behaviorally separable phenomena are neurobiologically distinct, we sought to replicate Kahan's findings on motivated confirmation bias while also testing for effects from the dopaminergic SNPs identified by Doll.

Results of this study showed significant MBDT performance differences as a function of interpreting the motivated (i.e., treatment) versus the unmotivated (i.e., control) task condition. Despite the numerical task being identical in the two conditions, control participants were approximately five times more likely to answer correctly than their treatment counterparts. Only the context of the scenario presented—the subject matter to which the participants were told the data referred—was different. Given no other detected group differences, it is likely that these differences in performance are due to this difference in context, thereby supporting the conclusion that treatment participants were more biased—motivatedly so—in their reading of the MBDT data table.

These findings are consistent with Kahan *et al*. (2017) who showed that the MBDT was a difficult task overall, and that performance on the task was much worse when subjects were presented with the motivated bias scenario. Like Kahan, we attribute this difference in group performance to motivated confirmation bias. Doll *et al.* (2011) identified three dopaminergic genes, DRD2, DARPP-32, and COMT, that were predictive of unmotivated confirmation bias. Thus, we examined whether the same SNPs in dopaminergic genes underlying these reward prediction errors—and responsible for individual differences in susceptibility to *unmotivated* confirmation bias—would also account for differences in motivated confirmation bias and, at least partially, account for the observed performance differences.

The present investigation found no relationship between genotypes and accuracy on a motivated confirmation bias detection test. Separate analyses for each gene revealed odds of accurately interpreting the data table differed only by group assignment with no effect of genotype. In additional analyses, we also examined specific allelic groupings. In Doll *et al.* (2011), behavioral differences associated with the three genes of interest were detected when grouping by the presence or absence of an allele, thereby creating two groups or genotypes for each gene. In these groupings, heterozygotes were included with one or the other homozygous groups. Therefore, we sought to replicate this approach to determine whether the binary reclassification of the genes revealed an underlying effect not detected by our first analysis. Prior to conducting these analyses, we first examined the equality of the distribution of the binary gene classification in the control and treatment groups: DRD2 was classified based on the presence or absence of the T allele, DARPP-32 on the basis of the presence or absence of the C allele, and COMT on the basis of the presence or absence of the Met allele. Results showed no differences in the distributions of the genotypes across the groups; therefore, we re-examined our models using the binary allele groupings. Findings were consistent with our broader analysis; only group assignment successfully predicted the odds of accurately interpreting the MBDT data table.
If correct, these findings may suggest that distinct genetic determinants of the dopamine reward system underlie motivated and unmotivated confirmation biases. This is a lingering possibility in the young field of research into the neurobiology of cognitive biases, and it is due to the difficulty in studying the neurobiology of motivated confirmation bias—an exceedingly complex, real-world phenomenon (Rollwage *et al*., 2020).

Alternatively, the relationship between these dopaminergic polymorphisms and motivated confirmation bias may be more complex than considered here. An example of just

such an explanation was seen above in Byrne *et al*. (2016), who hypothesized that the opposing findings from Doll *et al*. (2011) and Collins and Frank (2012) on the association between DARPP-32 and reinforcement learning may have been due to differences in the task. One of those tasks involved classification of novel symbols, while the other involved classification of familiar symbols, a cognitive difference which might employ distinct neural processes.

Similarly, while Talluri *et al*. (2018) showed that post-choice perceptual sensitivity is increased for choice-consistent stimuli, Rollwage *et al.* (2020) showed that confidence level in the initial choice may better explain post-selection bias. In short, prior research suggests that there is more than one possible explanation why the dopaminergic polymorphisms predictive of unmotivated confirmation bias were not predictive of motivated confirmation bias in the present study. It may be that these polymorphisms do not participate in these two cognitively distinct biases. Alternatively, if they do participate in both types of confirmation bias, they may do so as part of more complex, distinct mechanisms that involve other variables not examined here. One such candidate variable for future research is confidence, as it may be more extreme in the motivated form of confirmation bias.

Doll *et al*. (2011) found that differences at these SNPs in the three genes of interest were predictive of individual differences in susceptibility to confirmation bias. Specifically, T allele carriers in DRD2 (rs6277), T allele carriers in DARPP-32 (rs907094), and Met allele carriers in COMT (rs4680), all demonstrated significantly more susceptibility to unmotivated confirmation bias. Were these same allelic differences part of a single neurobiological cause of both motivated and unmotivated confirmation bias, they would have been predictive of accuracy in the present study's MBDT. We reasoned that RPEs are the probable determinant of both types of confirmation bias, and as such, SNPs in dopaminergic genes whose function is related to RPEs in

unmotivated confirmation bias would also be related to motivated confirmation bias. On the other hand, motivated and unmotivated confirmation bias are cognitively and behaviorally separable. This may be due to distinct neurological mechanisms underlying these phenomena. While future research will have to identify probable genetic and neurobiological markers for susceptibility to motivated confirmation bias, it remains possible that these will regulate dopamine in RPE function.

Not having found main effects for any of the genes of interest, we conducted secondary analyses to examine whether numeracy, political party identity, or political philosophy affected the odds of correctly interpreting a motivated bias task. Kahan previously showed that congruency between political identity and data table assignment predicted accuracy, whereas incongruency predicted inaccuracy. In other words, when participants read a data table that confirmed their prior beliefs, they were more likely to interpret that table correctly than participants who read a table that contradicted their prior beliefs. Findings of the present study did not confirm this association – we found no effects for party identity or political ideology after controlling for group assignment.

Numeracy, or the facility with which one understands and is able to work with numbers, is believed to enhance one's ability to apply rational thinking or judgment in the face of a variety of problem sets. In the literature on numeracy and comprehending medical data for weighing health risks, for example, higher numeracy associated with getting medical treatment, adhering to instructions for the use of medications, understanding screenings risks, and better health outcomes overall (Reyna *et al.*, 2009). However, prior research has shown that even when individuals possess cognitive problem-solving skills such as high numeracy, these skills confer no additional benefit when motivated reasoning is at play. For example, Kahan *et al.* (2017)

showed that performance differed between motivated and unmotivated task conditions but found no effect of numeracy in the motivated scenario. The present study, by contrast, found that even after controlling for group assignment, numeracy was still associated with increased odds of accurately responding in the MBDT task. So, unlike Kahan *et al.*, who found that numeracy had no effect in the motivated condition, we showed that numeracy is predictive of accuracy in both motivated and unmotivated scenarios in certain populations.

Why does numeracy confer any benefit at all? In both Kahan's study and the control condition in the present study, numeracy conveyed an advantage that was predictive of accuracy. This is because numeracy is a measure of facility with numerical information (Peters *et al.*, 2006), and the tasks in both experiments were fundamentally numerical tasks—namely, interpreting data tables. This is consistent with the broader body of judgment and decision-making research showing that numeracy is predictive of optimal, rational outcomes in a variety of numerical tasks, from understanding risk in medical decision-making (Reyna *et al.*, 2009) to completing formal education and maintaining employment (Parsons & Bynner, 2005). Thus, numeracy typically conveys a performance advantage in narrow numerical tasks as well as broader endeavors that rely on numerical reasoning. The importance of Kahan *et al.*'s (2017) finding is that it showed this advantage was wiped away in the face of motivated confirmation bias.

Why, in confronting motivated confirmation bias, did numeracy offer a benefit to our participants, whereas for Kahan's, it did not? We suspect the difference lies in the populations from which our respective samples were taken. Kahan *et al.*'s participants were 1111 adults, with a mean age of 48 years, recruited through an online testing firm. Our participants were 200

university students with a mean age of 19 years who participated in a university setting for class credit. Age and/or setting may be critical determinants of performance on the MBDT.

Since the numerical tasks in the treatment and control conditions of the MBDT are identical, and only the explanations about subject matter to which the data refer are different, we conclude that the finding of a main effect for group assignment is the result of motivated confirmation bias. None of the potential genetic differences or political beliefs explored in this study explained this effect. That these latter predictors did not have any significant effects on outcome was most likely due to a non-normal, low kurtosis, distribution of party identity and political ideology. Figures 5 and 6 show the distributions of party identity and political philosophy broken down by correct and incorrect responses on the MBDT. In the sample, participants' party identity and political philosophy clustered around the center and away from the tails of the distribution. As a result, there may have been too few extreme political party affiliations and too few extreme political philosophies to show an effect on MBDT accuracy. We explored alternative classification of these scale scores which resulted in no difference in the findings.

Further, the sample largely consisted of university freshmen and sophomores enrolled in an introductory psychology course at a southern university. It is possible that in this demographic, views on gun control—the context in which the treatment group data were presented—are less variable than elsewhere in the United States, and not captured by measures of political party or political philosophy.

Limitations

We hypothesized that because both motivated and unmotivated confirmation bias likely operate as the result of RPEs, the same dopaminergic gene SNPs that are predictive of individual

differences in unmotivated confirmation bias would also be predictive of motivated confirmation bias. Results of the present study did not confirm this hypothesis and instead suggests that the neurobiological mechanisms that account for individual differences in susceptibility to motivated confirmation bias may differ and remain to be discovered. While alternative biomarkers have yet to be identified to account for such neurobiological differences, it is reasonable for future research to take one of two approaches. The first is to examine whether there are alternatives to RPEs that can serve as the underlying processes that produce confirmation bias. If there are not, future research should examine biomarkers that produce differences in dopamine regulation as alternatives. Like the polymorphisms in DRD2 and DARPP-32, these might directly affect dopamine availability in the striatum. Like COMT, however, prospective biomarkers might have only second- or third-order effects on the striatum while having their primary effects elsewhere (the PFC in the case of COMT) and then influence the striatum through neuronal projections from one region to another.

The sample size of 200 may not have been large enough to detect the putative effects of political party identity and political philosophy on accuracy in the MBDT, as effects sizes were lower than anticipated. A larger sample size from this population of 18-19-year-old university students might detect such effects. Alternatively, a broader and more age diverse population might be one in which political party affiliation and political philosophy are more indicative of beliefs and biases about gun control and gun issues.

One alternative approach to overcome the apparent disconnect between the detected bias on gun control, and the undetected bias connected to political measures, might be to include a scenario for which university students' views are more closely connected to their political commitments. Had we envisioned such a disconnect in the study sample, additional measures of

attitudes on gun control issues could have been included to further explore this possibility. These might be simple measures like Likert scales capturing one's support for gun control, or validated tools like the scale developed by Teske & Hazlett (1985) for measuring attitudes toward handgun control specifically. Such measures would give an indication of prior beliefs that could serve as a ground for motivated confirmation bias. If successful, they would show up in the control group as a difference in accuracy between participants whose prior beliefs and table assignment (gun control is effective at reducing violence, or gun control is ineffective at reducing violence) were aligned, and those whose beliefs and table assignment were incongruous.

Conclusions

The data presented here are evidence that the genetic and neurobiological determinants of motivated and unmotivated confirmation bias are distinct. Nevertheless, the biological mechanisms underlying motivated confirmation bias remain to be discovered. Future research should seek to replicate the present findings and identify biomarkers of interest in the neurobiological architecture of RPEs and dopamine regulation. That numeracy provides some benefit in the face of motivated confirmation bias, in at least some populations, is also a novel finding that warrants efforts at replication and further explanation. A better understanding of the precise conditions under which numeracy can be harnessed to combat motivated confirmation bias may provide additional tools in the effort to augment cognitive performance and improve rational judgment and decision-making.

Finally, this is a young field of research that is still developing research methods for examining complex real-world phenomena. Imaging work like that from Kaplan *et al.* (2016) can examine the neural correlates of emotionally and cognitively complex processes, like resistance to changing one's mind on political topics. However, imaging studies can only go so far. Kaplan

et al.'s work may be useful in identifying areas of activity where others may look for candidate mechanisms that explain such complex processes. It is also the case that findings from individual studies showing relationships between dopaminergic polymorphisms and relatively simple cognitive processes have not always survived examination in meta-analyses. Progress in our understanding of the neurobiology of simpler cognitive biases may be necessary to understanding complex real-world phenomena like motivated confirmation bias.

References

Åberg, E., Fandiño-Losada, A., Sjöholm, L. K., Forsell, Y., & Lavebratt, C. (2011). The functional Val158Met polymorphism in catechol-O-methyltransferase (COMT) is associated with depression and motivation in men from a Swedish population-based study. *Journal of Affective Disorders*, *129*(1-3), 158-166.

Aczel, B., Bago, B., Szollosi, A., Foldes, A., & Lukacs, B. (2015). Is it time for studying real-life debiasing? Evaluation of the effectiveness of an analogical intervention technique. *Frontiers in psychology*, *6*, 1120.

Arkes, H. R. (2003). The nonuse of psychological research at two federal agencies. *Psychological science*, *14*(1), 1-6.

Atanasov, P., Rescober, P., Stone, E., Swift, S. A., Servan-Schreiber, E., Tetlock, P., ... & Mellers, B. (2017). Distilling the wisdom of crowds: Prediction markets vs. prediction polls. *Management Science*, *63*(3), 691-706.

Athey, S., Mobius, M. M., & Pál, J. (2017). The impact of aggregators on internet news consumption. *Stanford University Graduate School of Business Research Paper Series*, *17*(8).

Babcock, L., & Loewenstein, G. (1997). Explaining bargaining impasse: The role of self-serving biases. *Journal of Economic perspectives*, *11*(1), 109-126.

Bayer, H. M. & Glimcher, P. W. (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron*, *47*(1), 129-141.

Bronfman, Z. Z., Brezis, N., Moran, R., Tsetsos, K., Donner, T., & Usher, M. (2015). Decisions reduce sensitivity to subsequent information. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1810), 20150228.

Byrne, K. A., Davis, T., & Worthy, D. A. (2016). Dopaminergic genetic polymorphisms predict rule-based category learning. *Journal of Cognitive Neuroscience, 28*(7), 959-970.

Camerer, C. F., & Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty, 19*(1-3), 7-42.

Camps, M., Cortes, R., Gueye, B., Probst, A., & Palacios, J. M. (1989). Dopamine receptors in human brain: Autoradiographic distribution of D2 sites. *Neuroscience, 28*(2), 275-290.

Castro, L., & Scartascini, C. (2013). Tax Compliance and Enforcement in the Pampas. *Evidence from a Field Experiment', Inter-American Development Bank Working Paper Series, Washington, DC, Inter-American Development Bank*.

Charness, G., & Dave, C. (2017). Confirmation bias with motivated beliefs. *Games and Economic Behavior, 104*, 1-23.

Collins, A. G., & Frank, M. J. (2012). How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. *European Journal of Neuroscience, 35*(7), 1024-1035.

Ćurčić-Blake, B., Swart, M., Ter Horst, G. J., Langers, D. R., Kema, I. P., & Aleman, A. (2012). Variation of the gene coding for DARPP-32 (PPP1R1B) and brain connectivity during associative emotional learning. *Neuroimage, 59*(2), 1540-1550.

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A, Caldarelli, G., Stanley, H. E., & Quattrociocci, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences 113* (3), 554-559.

Del Vicario, M., Scala, A., Caldarelli, G., Stanley, H. E., & Quattrociocchi, W. (2017). Modeling confirmation bias and polarization. *Scientific Reports 7*, 40391.

de Souza Costa, D., Bechara, A., de Paula, J. J., Romano-Silva, M. A., Correa, H., Lage, G. M., ... & Malloy-Diniz, L. F. (2016). Influence of COMT Val158Met polymorphism on emotional decision-making: A sex-dependent relationship? *Psychiatry Research*, *246*, 650-655.

Deserno, L., Huys, Q. J., Boehme, R., Buchert, R., Heinze, H. J., Grace, A. A., ... & Schlagenhauf, F. (2015). Ventral striatal dopamine reflects behavioral and neural signatures of model-based control during sequential decision making. *Proceedings of the National Academy of Sciences*, *112*(5), 1595-1600.

Doll, B. B., Hutchison, K. E., & Frank, M. J. (2011). Dopaminergic genes predict individual differences in susceptibility to confirmation bias. *Journal of Neuroscience*, *31*(16), 6188-6198.

Doll, B. B., Shohamy, D., & Daw, N. D. (2015). Multiple memory systems as substrates for multiple decision systems. *Neurobiology of Learning and Memory*, *117*, 4-13.

Druckman, J. N. (2001). The implications of framing effects for citizen competence. *Political behavior*, *23*(3), 225-256.

Druckman, J. N. (2015). Communicating policy-relevant science. *Political Science & Politics*, *48*(1), 58.

Duffy, A., and Ling, R. (2020). The gift of news: A shift from information to social in studies of news sharing. *Journalism Studies 21*(1), 72-87.

Durstewitz, D., & Seamans, J. K. (2008). The dual-state theory of prefrontal cortex dopamine function with relevance to catechol-o-methyltransferase genotypes and schizophrenia. *Biological Psychiatry*, *64*(9), 739-749.

Durstewitz, D., Vittoz, N. M., Floresco, S. B., & Seamans, J. K. (2010). Abrupt transitions between prefrontal neural ensemble states accompany behavioral transitions during rule learning. *Neuron*, *66*(3), 438-448.

Elston, D. M. (2020). Confirmation bias in medical decision-making. *Journal of the American Academy of Dermatology*, *82*(3), 572.

Feldcamp, L. A., Souza, R. P., Romano-Silva, M., Kennedy, J. L., & Wong, A. H. (2008). Reduced prefrontal cortex DARPP-32 mRNA in completed suicide victims with schizophrenia. *Schizophrenia Research*, *103*(1-3), 192-200.

Fischoff, B. (1982). Debiasing. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 422-444). Cambridge University Press.

Fong, G. T., & Nisbett, R. E. (1991). Immediate and delayed transfer of training effects in statistical reasoning. *Journal of Experimental Psychology: General*, *120*(1), 34.

Frank, M. J. (2005). Dynamic dopamine modulation in the basal ganglia: A neurocomputational account of cognitive deficits in medicated and non-medicated Parkinsonism. *Journal of Cognitive Neuroscience*, *17*(1), 51-72.

Frank, M. J., Moustafa, A. A., Haughey, H. M., Curran, T., & Hutchison, K. E. (2007). Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proceedings of the National Academy of Sciences 104*(41), 16311-16316.

Frank, M. J., Doll, B. B., Oas-Terpstra, J., & Moreno, F. (2009). Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation. *Nature Neuroscience*, *12*(8), 1062.

Fransen, M. P., Leenaars, K. E., Rowlands, G., Weiss, B. D., Maat, H. P., & Essink-Bot, M. L. (2014). International application of health literacy measures: Adaptation and validation of

the newest vital sign in The Netherlands. *Patient education and counseling*, *97*(3), 403-409.

Friedrich, J. (1993). Primary error detection and minimization (PEDMIN) strategies in social cognition: A reinterpretation of confirmation bias phenomena. *Psychological Review*, *100*(2), 298.

Funk, C., & Kennedy, B. (2016). The politics of climate. *Pew Research Center*, *4*.

Gao, X., Gong, P., Liu, J., Hu, J., Li, Y., Yu, H., ... & Zhou, X. (2016). COMT Val158Met polymorphism influences the susceptibility to framing in decision-making: OFC-amygdala functional connectivity as a mediator. *Human Brain Mapping*, *37*(5), 1880-1892.

Garcia-Alamino, J. M. (2020). Human biases and the SARS-CoV-2 pandemic. *Intensive & Critical Care Nursing 58,* 102861.

Garcia-Retamero, R., & Cokely, E. T. (2011). Effective communication of risks to young adults: Using message framing and visual aids to increase condom use and STD screening. *Journal of Experimental Psychology: Applied*, *17*(3), 270.

Garcia-Retamero, R., & Cokely, E. T. (2013). Communicating health risks with visual aids. *Current Directions in Psychological Science*, *22*(5), 392-399.

Garrett, R. K. (2017). The "echo chamber" distraction: Disinformation campaigns are the problem, not audience fragmentation. *Journal of Applied Research in Memory and Cognition 6*(4), 370-376.

Garrison, J., Erdeniz, B., & Done, J. (2013). Prediction error in reinforcement learning: a meta-analysis of neuroimaging studies. *Neuroscience & Biobehavioral Reviews*, *37*(7), 1297-1310.

Gašević, D., Joksimović, S., Eagan, B. R., & Shaffer, D. W. (2019). SENS: Network analytics to combine social and cognitive perspectives of collaborative learning. *Computers in Human Behavior*, *92*, 562-577.

Geller, S., Wilhelm, O., Wacker, J., Hamm, A., & Hildebrandt, A. (2017). Associations of the COMT Val158Met polymorphism with working memory and intelligence–A review and meta-analysis. *Intelligence*, *65*, 75-92.

Gogos, J. A., Morgan, M., Luine, V., Santha, M., Ogawa, S., Pfaff, D., & Karayiorgou, M. (1998). Catechol-O-methyltransferase-deficient mice exhibit sexually dimorphic changes in catecholamine levels and behavior. *Proceedings of the National Academy of Sciences*, *95*(17), 9991-9996.

Gorman, M. E. (1986). How the possibility of error affects falsification on a task that models scientific problem solving. *British Journal of Psychology*, *77*(1), 85-96.

Gorman, M. E. (1989). Error, falsification and scientific inference: An experimental investigation. *The Quarterly Journal of Experimental Psychology Section A*, *41*(2), 385-412.

Guilbeault, D., Becker, J., & Centola, D. (2018). Social learning and partisan bias in the interpretation of climate trends. *Proceedings of the National Academy of Sciences*, *115*(39), 9714-9719.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*(4), 814–834. https://doi.org/10.1037/0033-295X.108.4.814

Hales, B. M., & Pronovost, P. J. (2006). The checklist—a tool for error management and performance improvement. *Journal of critical care*, *21*(3), 231-235.

Hart, W., Albarracín, D., Eagly, A. H., Brechan, I., Lindberg, M. J., & Merrill, L. (2009). Feeling validated versus being correct: a meta-analysis of selective exposure to information. *Psychological bulletin*, *135*(4), 555.

Heinz, A., & Smolka, M. N. (2006). The effects of catechol O-methyltransferase genotype on brain activation elicited by affective stimuli and cognitive tasks. *Reviews in the Neurosciences*, *17*(3), 359-368.

Herzog, S. M., & Hertwig, R. (2009). The wisdom of many in one mind: Improving individual judgments with dialectical bootstrapping. *Psychological Science*, *20*(2), 231-237.

Hikida, T., Kimura, K., Wada, N., Funabiki, K., & Nakanishi, S. (2010). Distinct roles of synaptic transmission in direct and indirect striatal pathways to reward and aversive behavior. *Neuron*, *66*(6), 896-907.

Hirvonen, M., Laakso, A., Någren, K., Rinne, J. O., Pohjalainen, T., & Hietala, J. (2004). C957T polymorphism of the dopamine D2 receptor (DRD2) gene affects striatal DRD2 availability in vivo. *Molecular Psychiatry*, *9*(12), 1060-1061.

Hirvonen, M., Laakso, A., Nagren, K., Rinne, J. O., Pohjalainen, T., & Hietala, J. (2005). C957T polymorphism of the dopamine D2 receptor (DRD2) gene affects striatal DRD2 availability *in vivo. Molecular Psychiatry*, *10*(9), 889-889.

Hirvonen, M., Laakso, A., Nagren, K., Rinne, J. O., Pohjalainen, T., & Hietala, J. (2009). C957T polymorphism of dompanien D2 receptor gene affects striatal DRD2 in vivo availability by changing the receptor affinity. *Synapse*, *63*(10), 907-912.

Hogarth, R. M. (2001). *Educating intuition*. University of Chicago Press.

Hogarth, R. M., Gibbs, B. J., McKenzie, C. R., & Marquis, M. A. (1991). Learning from feedback: Exactingness and incentives. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*(4), 734.

Houghton, R. J., Baber, C., Stanton, N. A., Jenkins, D. P., & Revell, K. (2015). Combining network analysis with cognitive work analysis: insights into social organisational and cooperation analysis. *Ergonomics*, *58*(3), 434-449.

Huotari, M., Santha, M., Lucas, L. R., Karayiorgou, M., Gogos, J. A., & Männistö, P. T. (2002). Effect of dopamine uptake inhibition on brain catecholamine levels and locomotion in catechol-O-methyltransferase-disrupted mice. *Journal of Pharmacology and Experimental Therapeutics*, *303*(3), 1309-1316.

Kagel, J. H., & Levin, D. (1986). The winner's curse and public information in common value auctions. *The American economic review*, 894-920.

Kahan, D. M., Peters, E., Dawson, E. C., & Slovic, P. (2017). Motivated numeracy and enlightened self-government. *Behavioural Public Policy*, *1*(1), 54-86.

Kaplan, J. T., Gimbel, S. I., & Harris, S. (2016). Neural correlates of maintaining one's political beliefs in the face of counterevidence. *Scientific Reports*, *6*, 39589.

Kappes, A., Harvey, A. H., Lohrenz, T., Montague, P. R., & Sharot, T. (2020). Confirmation bias in the utilization of others' opinion strength. *Nature Neuroscience*, *23*(1), 130-137.

Klaus, K., Butler, K., Curtis, F., Bridle, C., & Pennington, K. (2019). The effect of ANKK1 Taq1A and DRD2 C957T polymorphisms on executive function: A systematic review and meta-analysis. *Neuroscience & Biobehavioral Reviews*, *100*, 224-236.

Klayman, J. (1995). Varieties of confirmation bias. *Psychology of Learning and Motivation*, *32*, 385-418.

Klayman, J. & Ha, Y. W. (1989). Hypothesis testing in rule discovery: Strategy, structure, and content. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 596-604.

Knobloch-Westerwick, S., Mothes, C., & Polavin, N. (2020). Confirmation bias, ingroup bias, and negativity bias in selective exposure to political information. *Communication Research*, *47*(1), 104-124.

Koehler, J. J. (1993). The influence of prior beliefs on scientific judgments of evidence quality. *Organizational Behavior and Human Decision Processes*, *56*(1), 28-55.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*(3), 480–498. https://doi.org/10.1037/0033-2909.108.3.480

Krugel, L. K., Biele, G., Mohr, P. N., Li, S. C., & Heekeren, H. R. (2009). Genetic variation in dopaminergic neuromodulation influences the ability to rapidly and flexibly adapt decisions. *Proceedings of the National Academy of Sciences*, *106*(42), 17951-17956.

Lerman, A. E., & Acland, D. (2020). United in States of Dissatisfaction: Confirmation Bias Across the Partisan Divide. *American Politics Research*, *48*(2), 227-237.

Liberali, J. M., Reyna, V. F., Furlan, S., Stein, L. M., & Pardo, S. T. (2012). Individual differences in numeracy and cognitive reflection, with implications for biases and fallacies in probability judgment. *Journal of behavioral decision making*, *25*(4), 361-381.

Lieder, F., Griffiths, T. L., Huys, Q. J., & Goodman, N. D. (2018). The anchoring bias reflects rational use of cognitive resources. *Psychonomic bulletin & review*, *25*(1), 322-349.

Ling, R. (2008). *New tech, new ties*. Ling, R. S. (2008). *New tech, new ties*. Cambridge, MA: MIT press.

Ling, R. (2014). Mobile phones and digital Gemeinschaft: social cohesion in the era of cars, clocks and cell phones. In *Mobility and Locative Media* (pp. 39-52). Routledge.

Ling, R. (2020). Confirmation Bias in the Era of Mobile News Consumption: The Social and Psychological Dimensions. *Digital Journalism*, 1-9.

Liu, L., Fan, D., Ding, N., Hu, Y., Cai, G., Wang, L., ... & Xu, J. (2014). The relationship between DRD2 gene polymorphisms (C957T and C939T) and schizophrenia: a meta-analysis. *Neuroscience Letters*, *583*, 43-48.

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, *37*(11), 2098.

Mannes, A. E., Larrick, R. P., & Soll, J. B. (2012). The social psychology of the wisdom of crowds. In J. I. Krueger (Ed.), *Frontiers of social psychology. Social judgment and decision making* (pp. 227-242). Psychology Press.

Marteau, T. M., Ogilvie, D., Roland, M., Suhrcke, M., & Kelly, M. P. (2011). Judging nudging: can nudging improve population health? *Bmj*, *342*.

Masta, K. E., and E. Shearer. 2018. "News Use Across Social Media Platforms 2018." Pew Research Center, Journalism & Media.

Matsumoto, M., Weickert, C. S., Beltaifa, S., Kolachana, B., Chen, J., Hyde, T. M., ... & Kleinman, J. E. (2003). Catechol O-methyltransferase (COMT) mRNA expression in the dorsolateral prefrontal cortex of patients with schizophrenia. *Neuropsychopharmacology*, *28*(8), 1521-1530.

McGrath, M., Kawachi, I., Ascherio, A., Colditz, G. A., Hunter, D. J., & De Vivo, I. (2004). Association between catechol-O-methyltransferase and phobic anxiety. *American Journal of Psychiatry, 161*(9), 1703-1705.

Mendel, R., Traut-Mattausch, E., Jonas, E., Leucht, S., Kane, J. M., Maino, K., ... & Hamann, J. (2011). Confirmation bias: why psychiatrists stick to wrong preliminary diagnoses. *Psychological Medicine, 41*(12), 2651-2659.

Meppelink, C. S., Smit, E. G., Fransen, M. L., & Diviani, N. (2019). "I was right about vaccination": Confirmation bias and health literacy in online health information seeking. *Journal of Health Communication*, *24*(2), 129-140.

Meyer-Lindenberg, A., Straub, R. E., Lipska, B. K., Verchinski, B. A., Goldberg, T., Callicott, J. H., Egan, M. F., . . . Weinberger, D. R. (2007). Genetic evidence implicating DARPP-32 in human frontostriatal structure, function, and cognition. *The Journal of Clinical Investigation*, *117*(3), 672-682.

Mowen, J. C., & Gaeth, G. J. (1992). The evaluation stage in marketing decision making. *Journal of the Academy of Marketing Science*, *20*(2), 177-187.

Munton, R. P., Vizi, S., Mansuy, I. M. (2004). The role of protein phosphatase-1 in the modulation of synaptic and structural plasticity. *FEBS Letters, 567*(1), 121-128.

Murray, G. K., Corlett, P. R., Clark, L., Pessiglione, M., Blackwell, A. D., Honey, G., & Fletcher, P. C. (2008). Substantia nigra/ventral tegmental reward prediction error disruption in psychosis. *Molecular psychiatry*, *13*(3), 267.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*(2), 175-220.

Niu, L., Behar-Horenstein, L. S., & Garvan, C. W. (2013). Do instructional interventions influence college students' critical thinking skills? A meta-analysis. *Educational Research Review*, *9*, 114-128.

Nkam, I., Ramoz, N., Breton, F., Mallet, J., Gorwood, P., & Dubertret, C. (2017). Impact of DRD2/ANKK1 and COMT polymorphisms on attention and cognitive functions in schizophrenia. *PLoS One*, *12*(1), e0170147.

Nys, T. R., & Engelen, B. (2017). Judging nudging: Answering the manipulation objection. *Political Studies*, *65*(1), 199-214.

Otto, A. R., Raio, C. M., Chiang, A., Phelps, E. A., & Daw, N. D. (2013). Working-memory capacity protects model-based learning from stress. *Proceedings of the National Academy of Sciences*, *110*(52), 20941-20946.

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*(8), e1005684.

Parsons, S., & Bynner, J. (2005). *Does numeracy matter more?* London: National Research and Development Centre for Adult Literacy and Numeracy.

Peters, E., Västfjäll, D., Slovic, P., Mertz, C. K., Mazzocco, K., & Dickert, S. (2006). Numeracy and decision making. *Psychological Science*, *17*(5), 407-413.

Plous, S. (1991). Biases in the assimilation of technological breakdowns: Do accidents make us safer? *Journal of Applied Social Psychology*, *21*(13), 1058-1082.

Quinn, C. M., Borzello, M., Zahir, A., Kramer, J. H., & Chiong, W. (2019). COMT Val158Met polymorphism associated with greater susceptibility to framing effects in healthy older adults. *bioRxiv*, 790808.

Read, D., & Grushka-Cockayne, Y. (2011). The similarity heuristic. *Journal of Behavioral Decision Making*, *24*(1), 23-46.

Reuter, M., Weber, B., Fiebach, C. J., Elger, C., & Montag, C. (2009). The biological basis of anger: associations with the gene coding for DARPP-32 (PPP1R1B) and with amygdala volume. *Behavioural Brain Research*, *202*(2), 179-183.

Reyna, V. F., Nelson, W. L., Han, P. K., & Dieckmann, N. F. (2009). How numeracy influences risk comprehension and medical decision making. *Psychological Bulletin*, *135*(6), 943.

Rollwage, M., Loosen, A., Hauser, T. U., Moran, R., Dolan, R. J., & Fleming, S. M. (2020). Confidence drives a neural confirmation bias. *Nature communications*, *11*(1), 1-11.

Rossmo, D. K., & Pollock, J. M. (2019). Confirmation bias and other systemic causes of wrongful convictions: A sentinel events perspective. *Northeastern University Law Review 11*(2), 790-835.

Scheggi, S., De Montis, M. G., & Gambarana, C. (2018). DARPP-32 in the orchestration of responses to positive natural stimuli. *Journal of neurochemistry*, *147*(4), 439-453.

Scherer, A. M., Windschitl, P. D., O'Rourke, J., & Smith, A. R. (2012). Hoping for more: The influence of outcome desirability on information seeking and predictions about relative quantities. *Cognition*, *125*(1), 113-117.

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*(1), 106-112.

Schultz, W. (2017). Reward prediction error. *Current biology*, *27*(10), R369-R371.

Schwarz, E., Tost, H., & Meyer-Lindenberg, A. (2016). Working memory genetics in schizophrenia and related disorders: an RDoC perspective. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *171*(1), 121-131.

Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H., & Simons, A. (1991). Ease of retrieval as information: Another look at the availability heuristic. *Journal of Personality and Social psychology*, *61*(2), 195.

Shen, W., Flajolet, M., Greengard, P., & Surmeier, D. J. (2008). Dichotomous dopaminergic control of striatal synaptic plasticity. *Science*, *321*(5890), 848-851.

Slifstein, M., Kolachana, B., Simpson, E. H., Tabares, P., Cheng, B., Duvall, M., ... & Abi-Dargham, A. (2008). COMT genotype predicts cortical-limbic D1 receptor availability measured with [11 C] NNC112 and PET. *Molecular Psychiatry*, *13*(8), 821-827.

Smith, V. L., & Walker, J. M. (1993). Rewards, experience and decision costs in first price auctions. *Economic Inquiry*, *31*(2), 237-244.

Smolka, M. N., Schumann, G., Wrase, J., Grüsser, S. M., Flor, H., Mann, K., ... & Heinz, A. (2005). Catechol-O-methyltransferase val158met genotype affects processing of emotional stimuli in the amygdala and prefrontal cortex. *Journal of Neuroscience*, *25*(4), 836-842.

Stipanovich, A., Valjent, E., Matamales, M., Nishi, A., Ahn, J. H., Maroteaux, M., ... & Filhol, O. (2008). A phosphatase cascade by which rewarding stimuli control nucleosomal response. *Nature*, *453*(7197), 879-884.

Svenningsson, P., Nishi, A., Fisone, G., Girault, J. A., Nairn, A. C., & Greengard, P. (2004). DARPP-32: an integrator of neurotransmission. *Annual Review of Pharmacology and Toxicology*, *44*, 269-296.

Talluri, B. C., Urai, A. E., Tsetsos, K., Usher, M., & Donner, T. H. (2018). Confirmation bias through selective overweighting of choice-consistent evidence. *Current Biology*, *28*(19), 3128-3135.

Teske Jr, R. H., & Hazlett, M. H. (1985). A scale for the measurement of attitudes toward handgun control. *Journal of criminal justice*, *13*(4), 373-379.

Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth, and happiness*. Penguin.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *science*, *211*(4481), 453-458.

Törnberg, P. (2018). Echo chambers and viral misinformation: Modeling fake news as complex contagion. *PLoS one*, *13*(9), e0203958.

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, *12*, 129-140.

Wason, P. C. (1968). Reasoning about a rule. *Quarterly journal of experimental psychology*, *20*(3), 273-281.

Weller, J. A., Dieckmann, N. F., Tusler, M., Mertz, C. K., Burns, W. J., & Peters, E. (2013). Development and testing of an abbreviated numeracy scale: A Rasch analysis approach. *Journal of Behavioral Decision Making*, *26*(2), 198-212.

Whitesmith, M. (2019). The efficacy of ACH in mitigating serial position effects and confirmation bias in an intelligence analysis scenario. *Intelligence and National Security*, *34*(2), 225-242.

Yariv, L. (2002). I'll See it When I Believe it? A Simple Model of Cognitive Consistency. Cowles Foundation Discussion Paper No. 1352.

Yoshimi, A., Takahashi, N., Saito, S., Ito, Y., Aleksic, B., Usui, H., ... & Iwata, N. (2008). Genetic analysis of the gene coding for DARPP-32 (PPP1R1B) in Japanese patients with schizophrenia or bipolar disorder. *Schizophrenia Research*, *100*(1-3), 334-341.

Zai, G., Robbins, T. W., Sahakian, B. J., & Kennedy, J. L. (2017). A review of molecular genetic studies of neurocognitive deficits in schizophrenia. *Neuroscience & Biobehavioral Reviews*, *72*, 50-67.

Zhang, S., & Zhang, J. (2016). The association of DRD2 with insight problem solving. *Frontiers in psychology*, *7*, 1865

Table 1

Treatment and Control Group Statistics for Age, Sex, Party Identity, Political Philosophy, and Numeracy Score

	Treatment	Control	t/χ^2	*p*-value
N	104	96		
Age	19.19 (3.31)	18.72 (1.24)	1.32	.189
Sex (M/F)	44/60	28/68	3.74	.053
Political Philosophy			4.230	.376
Very Liberal	6 (5.8%)	4 (4.2%)		
Liberal	32 (30.8%)	27 (28.1%)		
Moderate	27 (26.0%)	34 (35.4%)		
Conservative	30 (28.8%)	34 (35.4%)		
Very Conservative	1 (0.01%)	5 (5.2%)		
Party Identity			.632	.123
Strong Democrat	4 (3.8%)	2 (2.1%)		
Democrat	18 (17.3%)	22 (22.9%)		
Independent Lean Democrat	20 (19.2%)	21 (21.9%)		
Independent	13 (12.5%)	13 (13.5%)		
Independent Lean Republican	18 (17.3%)	11 (11.5%)		
Republican	21 (20.2%)	25 (26.0%)		
Strong Republican	2 (1.9%)	10 (10.4%)		

Note. Means (SD) and *t*-tests were used for age and numeracy score; Frequencies and chi-square tests were used for gender, party identity, political philosophy.

Table 2

Allele Frequency (%) in Treatment and Control Groups for Single Nucleotide Polymorphisms in Three Genes of Interest: DRD2, DARPP-32, and COMT

Gene	Treatment	Control	χ_2	*p*-value
DRD2			8.176	.017
C/C	23 (22.1%)	29 (30.2%)		
T/T	37 (35.6%)	17 (17.7%)		
C/T	44 (42.3%)	50 (52.1%)		
DARPP-32			2.374	.305
C/C	50 (48.1%)	51 (53.1%)		
T/T	16 (15.4%)	8 (8.3%)		
C/T	38 (36.5%)	37 (38.5%)		
COMT			3.012	.222
VAL/VAL	24 (23.1%)	13 (13.5%)		
MET/MET	33 (31.7%)	34 (35.4%)		
VAL/MET	47 (45.2%)	49 (51.0%)		

Table 3

Frequency (%) of Allele Presence in Treatment and Control Groups for Each of Three Genes of Interest: DRD2, DARPP-32, and COMT

Presence of allele	Treatment	Control	χ^2	*p*-value
DRD2 T allele	81 (77.9%)	67 (69.8%)	1.699	.192
DARPP-32 C allele	88 (84.6%)	88 (91.8%)	2.350	.125
COMT Met allele	80 (76.9%)	83 (86.5%)	3.010	.083

Table 4

Logistic Regression Analysis of MBDT Accuracy for Each of Three Genes of Interest: DRD2, DARPP-32, and COMT

Model	β	SE β	Wald's χ_2	*df*	*p*	Odds Ratio
DRD2						
Constant	-.806	.171	22.125	1	<.001	
Group	-.8159	.175	21.686	1	< .001	.196
(0 = control, 1 = treatment)						
Table	.024	.165	.021	1	.884	1.049
(0 = ineffective, 1 = effective)						
TT allele (vs. CC)	.245	.257	.912	1	.340	1.203
CT allele (vs. CC)	-.305	.223	1.882	1	.170	.694
DARPP-32						
Constant	-.944	.213	19.67	1	<.001	
Group	-.766	.169	20.434	1	< .001	.216
Table	-.003	.164	.000	1	.986	.994
TT allele (vs. CC)	.287	.371	.595	1	.440	.709
CT allele (vs. CC)	.229	.258	.786	1	.375	1.186
COMT						
Constant	-.899	.183	24.039	1	<.001	
Group	-.779	.171	20.776	1	< .001	.211
Table	.008	.164	.002	1	.960	1.017
VV allele (vs. MM)	.018	.292	.004	1	.952	1.317
MV allele (vs. MM)	.241	.222	1.169	1	.280	1.646

Table 5

Logistic Regression Analysis of MBDT Accuracy for Each of Three Genes of Interest, DRD2, DARPP-32, and COMT, with Genotypes Regrouped for the Presence of an Allele of Interest.

Model	β	SE β	Wald's χ_2	*df*	*p*	Odds Ratio
DRD2						
Constant	-.817	.187	19.165	1	<.001	
Group	-.768	.169	20.659	1	< .001	.215
(0 = control, 1 = treatment)						
Table	.001	.163	.000	1	.995	1.002
(0 = ineffective, 1 = effective)						
TT & CT alleles (vs. CC)	-.089	.181	.241	1	.624	.837
DARPP-32						
Constant	-1.022	.279	13.409	1	<.001	
Group	-.764	.169	20.407	1	< .001	.217
Table	-.006	.164	.001	1	.972	.988
CC & CT alleles (vs. TT)	.209	.278	.566	1	.452	1.520
COMT						
Constant	-.867	.217	15.938	1	<.001	
Group	-.774	.170	20.782	1	< .001	.213
Table	-.002	.163	.000	1	.990	.996
MM & VM alleles (vs. VV)	.014	.217	.004	1	.947	1.029

Figure 1

Control Group Table for Effective Rash Cream

Medical researchers have developed a new cream for treating skin rashes. New treatments often work but sometimes make rashes worse. Even when treatments don't work, skin rashes sometimes get better and sometimes get worse on their own. As a result, it is necessary to test any new treatment in an experiment to see whether it makes the skin condition of those who use it better or worse than if they had not used it.

Researchers have conducted an experiment on patients with skin rashes. In the experiment, one group of patients used the new cream for two weeks, and a second group did not use the new cream.

In each group, the number of people whose skin condition got better and the number whose condition got worse are recorded in the table below. Because patients do not always complete studies, the total number of patients in each group is not exactly the same, but this does not prevent assessment of the results.

Please indicate whether the experiment shows that using the new cream is likely to make the skin condition better or worse.

	Rash Got Worse	Rash Got Better
Patients who did use the new skin cream	**223**	**75**
Patients who did not use the new skin cream	**107**	**21**

What result does the study support?

- ○ People who used the skin cream were more likely to GET BETTER than those who didn't.
- ○ People who used the skin cream were more likely to GET WORSE than those who didn't.

Note. Experimental task table and instructions presented to some participants in the control group. When interpreted correctly, this table indicates that the cream used to treat the rash was effective.

Figure 2

Control Group Table for Ineffective Rash Cream

Medical researchers have developed a new cream for treating skin rashes. New treatments often work but sometimes make rashes worse. Even when treatments don't work, skin rashes sometimes get better and sometimes get worse on their own. As a result, it is necessary to test any new treatment in an experiment to see whether it makes the skin condition of those who use it better or worse than if they had not used it.

Researchers have conducted an experiment on patients with skin rashes. In the experiment, one group of patients used the new cream for two weeks, and a second group did not use the new cream.

In each group, the number of people whose skin condition got better and the number whose condition got worse are recorded in the table below. Because patients do not always complete studies, the total number of patients in each group is not exactly the same, but this does not prevent assessment of the results.

Please indicate whether the experiment shows that using the new cream is likely to make the skin condition better or worse.

	Rash Got Better	Rash Got Worse
Patients who did use the new skin cream	**223**	**75**
Patients who did not use the new skin cream	**107**	**21**

What result does the study support?

- ○ People who used the skin cream were more likely to GET BETTER than those who didn't.
- ○ People who used the skin cream were more likely to GET WORSE than those who didn't.

Note. Experimental task table and instructions presented to some participants in the control group. When interpreted correctly, this table indicates that the cream used to treat the rash was ineffective.

Figure 3

Treatment Group Table for Effective Gun Control Policy

A city government is trying to decide whether to pass a law banning private citizens from carrying concealed handguns in public. Government officials are unsure whether the law will be more likely to decrease crime by reducing the number of people carrying weapons or increase crime by making it harder for law-abiding citizens to defend themselves from violent criminals.

To address this question, researchers divided cities into two groups: one consisting of cities that had recently enacted bans on concealed weapons and another that had no such bans. They then observed the numbers of cities that experienced decreases in crime and those that experienced increases in crime in the next year.

In each group, the number of cities that saw an increase in crime and the number that saw a decrease in crime are recorded in the table below. Because cities do not always continue or enforce a newly implemented ban, the total number of cities in each group is not exactly the same, but this does not prevent assessment fo the results.

Please indicate whether cities that enacted a ban on carrying concealed handguns were more likely to have a decrease in crime or instead were more likely to have an increase in crime than cities without bans.

	Increase in crime	Decrease in crime
Cities that did ban carrying concealed handguns in public	**223**	**75**
Cities that did not ban carrying concealed handguns in public	**107**	**21**

What result does the study support?

- ○ Cities that banned carrying concealed handguns in public saw a DECREASE IN CRIME
- ○ Cities that banned carrying concealed handguns in public saw an INCREASE IN CRIME

Note. Experimental task table and instructions presented to some participants in the treatment group. When interpreted correctly, this table indicates that the ban on carrying concealed handguns in public was effective in reducing crime.

Figure 4

Treatment Group Table for Ineffective Gun Control Policy

A city government is trying to decide whether to pass a law banning private citizens from carrying concealed handguns in public. Government officials are unsure whether the law will be more likely to decrease crime by reducing the number of people carrying weapons or increase crime by making it harder for law-abiding citizens to defend themselves from violent criminals.

To address this question, researchers divided cities into two groups: one consisting of cities that had recently enacted bans on concealed weapons and another that had no such bans. They then observed the numbers of cities that experienced decreases in crime and those that experienced increases in crime in the next year.

In each group, the number of cities that saw an increase in crime and the number that saw a decrease in crime are recorded in the table below. Because cities do not always continue or enforce a newly implemented ban, the total number of cities in each group is not exactly the same, but this does not prevent assessment fo the results.

Please indicate whether cities that enacted a ban on carrying concealed handguns were more likely to have a decrease in crime or instead were more likely to have an increase in crime than cities without bans.

	Decrease in crime	Increase in crime
Cities that did ban carrying concealed handguns in public	**223**	**75**
Cities that did not ban carrying concealed handguns in public	**107**	**21**

What result does the study support?

- ○ Cities that banned carrying concealed handguns in public saw a DECREASE IN CRIME
- ○ Cities that banned carrying concealed handguns in public saw an INCREASE IN CRIME

Note. Experimental task table and instructions presented to some participants in the treatment group. When interpreted correctly, this table indicates that the ban on carrying concealed handguns in public was ineffective in reducing crime.

Figure 5

Accuracy Across the Distribution of Responses to Political Philosophy Scale

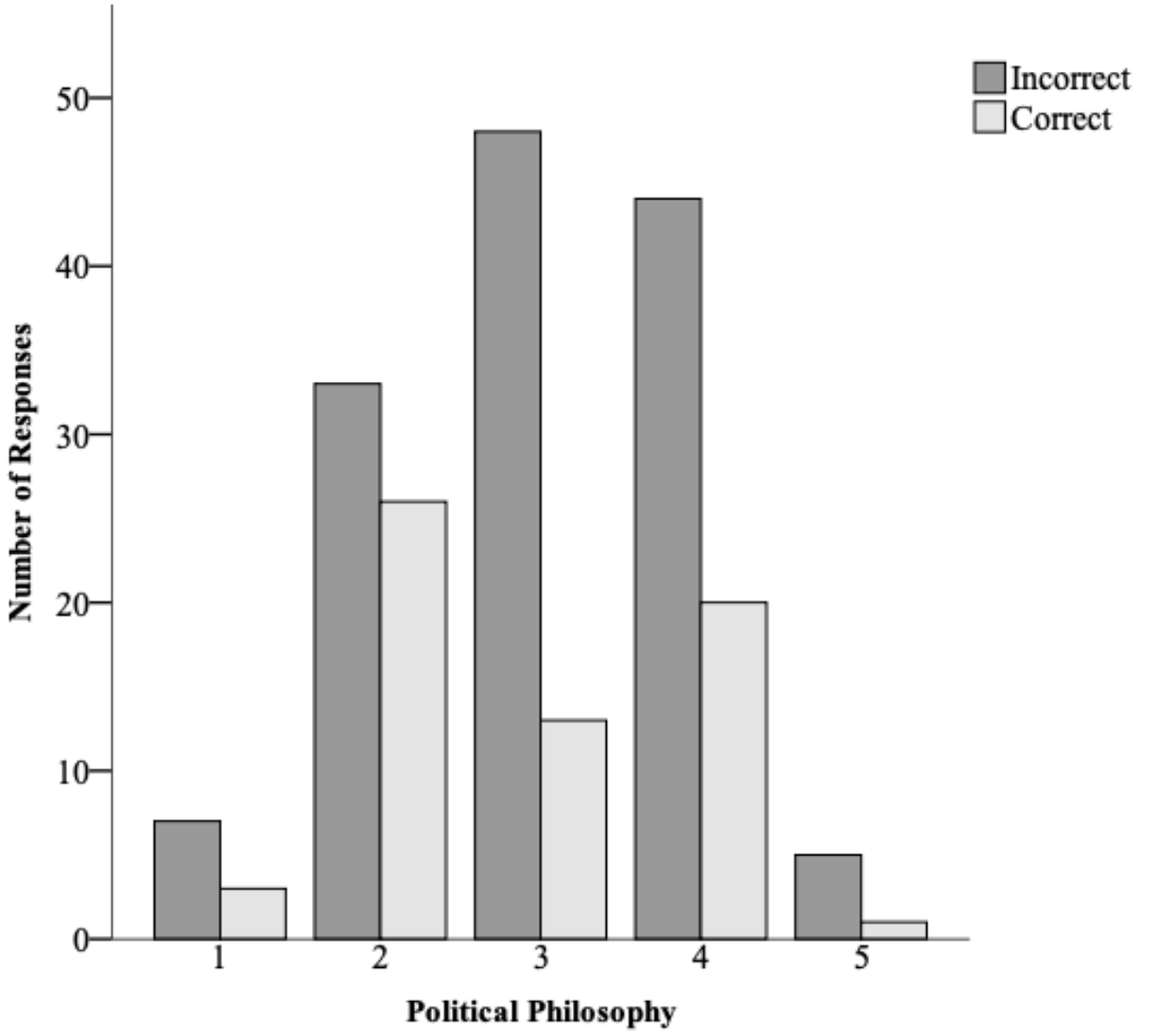

Figure 6

Accuracy Across the Distribution of Responses to Party Identity Scale