COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 45 (2021) e13018

ISSN: 1551-6709 online
DOI: 10.1111/cogs.13018

Motivated Reasoning in an Explore-Exploit Task

Zachary A. Caddick, Benjamin M. Rottman

Department of Psychology, University of Pittsburgh

Received 2 June 2020; received in revised form 18 June 2021; accepted 24 June 2021

Abstract

The current research investigates how prior preferences affect causal learning. Participants were tasked with repeatedly choosing policies (e.g., increase vs. decrease border security funding) in order to maximize the economic output of an imaginary country and inferred the influence of the policies on the economy. The task was challenging and ambiguous, allowing participants to interpret the relations between the policies and the economy in multiple ways. In three studies, we found evidence of motivated reasoning despite financial incentives for accuracy. For example, participants who believed that border security funding should be increased were more likely to conclude that increasing border security funding actually caused a better economy in the task. In Study 2, we hypothesized that having neutral preferences (e.g., preferring neither increased nor decreased spending on border security) would lead to more accurate assessments overall, compared to having a strong initial preference; however, we did not find evidence for such an effect. In Study 3, we tested whether providing participants with possible functional forms of the policies (e.g., the policy takes some time to work or initially has a negative influence but eventually a positive influence) would lead to a smaller influence of motivated reasoning but found little evidence for this effect. This research advances the field of causal learning by studying the role of prior preferences, and in doing so, integrates the fields of causal learning and motivated reasoning using a novel explore-exploit task.

Keywords: Causal learning; Dynamic; Economic decision making; Explore-Exploit; Motivated reasoning

1. Introduction

> "*I'm not saying there won't be a little pain … we might lose a little bit … but we're gonna have a much stronger country when we are finished.… So, we may take a hit, and you know what, ultimately we're going to be much stronger for it.*"
> - President Donald Trump (Factbase, 2018)

Correspondence should be sent to Zachary A. Caddick, University of Pittsburgh, 3420 Forbes Avenue, Pittsburgh, PA 15213, USA. E-mail: zac21@pitt.edu

"*The tariffs are beginning to have some impact in a negative way so I hope that we make some progress quickly on some of these other fronts, in particular with China.… If the end result of this is better trading relationships with all of these countries, particularly if it happens sooner rather than later, I think it would be great*."

-Senate Republican Leader Mitch McConnell (Shepardson, 2018)

"*Trump Tariffs Are Short-Term Pain Without Long-Term Gain, Economists Say: Nearly three-fourths of economists in WSJ [Wall Street Journal] survey said they expect short-term trade costs to outweigh any long-term benefits*."

-Wall Street Journal Article (Torry, 2019)

Humans are often faced with the task of evaluating the efficacy of an action or policy in dynamic settings, which can be very challenging. For example, when a politician decides to implement a new economic policy (e.g., tariffs), assessing the true impact of the policy is likely to be very difficult because other factors in the economy also change over time, and because one's expectations about how fast the policy will work and the short versus long-term impacts of the policy could lead different people to focus on different evidence. For another example, when a patient is assessing whether a medication is working, it is also very complicated because medications have complicated profiles of how quickly and long they work for and whether they produce short-term or long-term side effects. For an example especially relevant to the current moment in time, when a governor is assessing whether easing social distancing rules led to a subsequent improvement in the economy and/or subsequent increased coronavirus disease of 2019 (COVID-19) infections, it is complicated because it is unclear how long it will take for these outcomes to occur, and the counterfactual (e.g., what would have happened if social distancing was eased earlier or later within the same community) is unavailable for comparison.

Even when learning simple stable relations (e.g., the cause has a probabilistic but unchanging weakly positive influence on the effect), prior beliefs and expectations have strong impacts on the assessment of the strength of the relation from very positive to very negative (Alloy & Tabachnik, 1984; Fugelsang & Thompson, 2000, 2003; Goedert, Ellefson, & Rehder, 2014). However, in dynamic situations like those mentioned above, the task is considerably harder, and individuals may rely on additional temporal expectations for navigating the task (e.g., Buehner & McGregor, 2006; Hagmayer & Waldmann, 2002). Furthermore, in many situations, an individual might have strong preferences or engage in wishful thinking or "motivated reasoning" (e.g., hoping that easing social distancing and mask-wearing policies will not lead to a spike in COVID-19 infections, or hoping that implementing new tariffs would not hurt the economy, regardless of their belief), which could bias their interpretations of the evidence. Yet there is surprisingly little research at the intersection of learning from experience and motivated reasoning, particularly in dynamic situations such as assessing economic policies, which was our goal. In the rest of the introduction, we first discuss motivated reasoning, then learning from experience in dynamic tasks, and finally propose a set of hypotheses that we tested in three studies.

1.1. Motivated reasoning

Often when we reason about information, we already have prior preferences pertaining to the subject-matter. Individuals tend to more easily confirm information that is congruent with prior preferences (Nickerson, 1998) and reject information that is incongruent with prior preferences (Kunda, 1990). For example, Taber and Lodge (2006) found that individuals who had strong preferences about gun control or affirmative action were more likely to devalue arguments that were incongruent to their preference,1 regardless of their quality. This two-pronged process is known as motivated reasoning, where individuals are both more likely to accept information that confirms a prior belief as well as reject information that disconfirms a prior belief.

The current research specifically examines motivated reasoning within how people learn cause-effect relations, and in particular when people need to learn from experience. Although much of the recent work on motivated reasoning does not explicitly involve causality, some of the formative work on motivated reasoning studied how people assess causal claims. For a paradigmatic example, Kunda (1987) found that people tend to believe that their own attributes will lead to positive outcomes and reject the possibility that their attributes might lead to negative outcomes. In the first study, Kunda provided a description of a hypothetical person who had one of two attributes. Participants rated how likely the person was to get divorced based upon this attribute. When the attribute (the cause) matched an attribute of the participant, they were less likely to view this attribute as leading to divorce (the effect). Study 2 was similar but examined attributes predictive of success in graduate school and found that individuals who did not want to go to graduate school (lack of motivation) were less likely to engage in preferential reasoning. Study 3 examined how participants evaluate scientific evidence. Participants read a scientific article stating that caffeine consumption leads to poor health outcomes for women. Women who drank a lot of coffee found the evidence less convincing than those who drank only a little or none; however, for men, there was no difference in the ratings of convincingness, presumably since the evidence was only relevant to women and hence men had no motivation to engage in biased reasoning.

Despite the fact that some of the foundational work on motivated reasoning involved causal reasoning (see Kunda, 1987, 1990), much of the research that followed has not focused on causality (e.g., see Campbell & Kay, 2014; Hart & Nisbet, 2011; Kaplan, Gimbel, & Harris, 2016; Klaczynski, 1997; Nyhan & Reifler, 2010; Paharia, Vohs, & Deshpandé, 2013). Additionally, often there is no presentation of statistical evidence between a potential cause and outcome. Instead, much of the research on motivated reasoning has focused on how people confirm or reject evidence for reasons aside from the data itself such as the news outlet it was reported through or the qualifications of the author of a scientific study, which is known as the credibility heuristic (Kahan, Braman, Cohen, Gastil, & Slovic, 2010). People also prefer information that comes from sources with similar ideological preferences over those with competing preferences (Marks, Copland, Loh, Sunstein, & Sharot, 2018).

The closest motivated reasoning study to ours (Kahan, Peters, Dawson, & Slovic, 2017) had participants make causal assessments from data presented in a 2×2 contingency table of cross-sectional data. The contingency table presented evidence about cities that either banned

handguns in public or not and whether there was an increase or decrease in crime. Despite being presented with objective numbers, participants were more likely to make correct inferences about the influence of handgun policies on crime when the data supported their previously held preferences about handguns.

Inspired by the political discourse around predicting and assessing policies, in the current study, we sought to integrate research on motivated reasoning into a paradigm in which participants assess the impact that their choices have on an outcome, similar to many task paradigms in which people learn from experience.

1.2. Learning from experience

A core question addressed by the fields of causal learning, multiple cue learning, and reinforcement learning is how people learn the relations between multiple cues or causes and an outcome and therein come to choose causes that bring about a desirable outcome.

Within the field of causal learning, research has studied how people think that the causes combine together, how they learn the unique impact of each cause, and how they learn about many stable or unstable causes simultaneously (e.g., Derringer & Rottman, 2018; Lucas, Bridgers, Griffiths, & Gopnik, 2014; Spellman, 1996). Research on multiple cue learning (e.g., Speekenbrink & Shanks, 2010) is similar in that the participant needs to learn how the multiple cues together predict an outcome. Within the field of reinforcement learning (e.g., multi-armed bandit tasks, Iowa Gambling Task), research has studied how people learn the outcomes associated with mutually exclusive options, specifically, how they choose among the various options to both learn about them as well as select the option that they think will produce the best outcome, and how they balance exploration and exploitation in dynamic situations (e.g., Bechara, Damasio, Tranel, & Damasio, 2005; Schulz, Konstantinidis, & Speekenbrink, 2017; Steyvers, Lee, & Wagenmakers, 2009).

Though these fields use somewhat different paradigms, the questions they address are highly overlapping. In the following sections, we highlight two key features of learning from experience that are especially relevant to situations in which motivated reasoning is at play.

1.2.1. Function learning and ambiguity

When learning the relation between a cue or a choice and an outcome, one of the challenging tasks is learning the "form" or the "function" that relates the two (e.g., Lucas, Griffiths, Williams, & Kalish, 2015; Schulz, Tenenbaum, Duvenaud, Speekenbrink, & Gershman, 2017). Prior research has shown that people more readily learn positive than negative functions and linear than nonlinear functions (e.g., Brehmer, 1971, 1974; Busemeyer, Byun, DeLosh, & McDaniel, 1997; Koh & Meyer, 1991). Not only are there many (in fact infinite) potential functions or classes of functions, but from limited and noisy data, the function may be ambiguous—it may not be clear which of multiple functions is the best fitting function. Faced with ambiguity, people often make use of prior beliefs or explanations to try to make sense of ambiguous data that could be explained in multiple ways (e.g., Luhmann & Ahn, 2007; Marsh & Ahn, 2009).

One of the types of functions that we studied—causes that exhibit negative short-term outcomes but positive long-term outcomes (or vice versa) after repeatedly using the cause—are especially ambiguous. Suppose a learner tries such a cause and notices that quickly after trying it, it seems to produce a strong negative outcome. In this case, a learner may decide to stop using it and may never even experience the long-term benefit; or suppose a learner tries a cause that produces a short-term benefit and continues to use it and later experiences a long-term negative outcome. The learner may be able to detect this long-term negative outcome, or they might instead attribute the negative outcome to something else changing over time. We also studied cases in which a cause exhibits a positive or negative effect, but it takes some repeated usage to produce the maximal influence; initially, the cause produces a small effect but over time it produces a bigger effect. These cases are less ambiguous than the cases in which the short-term and long-term outcomes are opposites. Still, they are ambiguous in the sense that if they are only tested for a short amount of time, the learner will not realize how beneficial or harmful they actually are.

One of the goals of the current study was to examine the impacts of how people learn about more versus less ambiguous functions in the context of political motivation. We hypothesized that motivation combined with ambiguity could make people think that they understand the impact of a policy when in reality they may be latching on to only one, perhaps short-term, perspective.

1.2.2. Active learning and explore-exploit paradigms

Another aspect of learning from experience that has become a focus of considerable research especially in reinforcement learning and causal learning is active learning. In active learning paradigms, a participant does not passively observe cues and outcomes but instead learns to control the outcome through making choices. In active learning situations, the learner needs to navigate both exploring various options and exploiting the options that they think are best.

There are many different sorts of active-learning and explore-exploit tasks. In static bandit tasks, some options are always better than other options, so the goal is to explore the various options and settle on which option is best (Gershman, 2018; Steyvers et al., 2009). In dynamic or restless bandit tasks, one option may initially be better than another, but over time, the second may become better (e.g., Speekenbrink & Konstantinidis, 2015; Yi, Steyvers, & Lee, 2009), and this may or may not be signaled by contextual cues (e.g., Schulz, Konstantinidis, & Speekenbrink, 2018). In another type of bandit, one option is always better than another (similar to static), but the baseline fluctuates over time (similar to dynamic; Rottman, 2016).

In the current study, we built off of another explore-exploit paradigm sometimes called the "Harvard Game" in which one option is better for the short-term, and the other option is better for the long term (see Sims, Neth, Jacobs, & Gray, 2013, for a review). This paradigm is known to be difficult; participants often exhibit "melioration" in which they primarily implement the version of the policy that produces the better short-term outcome but is sub-optimal in the long run. We decided to build our task off of this paradigm because many real-world policies have different short- versus long-term impacts.

In many of these different bandit problems previously mentioned, the optimal strategy is often impossible to calculate analytically and/or rest on very particular assumptions about the task. In many real-world settings, the number of possible functional forms or types of dynamics is so large that it would not be possible to specify an optimal strategy. Situations with high uncertainty about how the causes function open the door for prior beliefs about which causes are better and motivations for wanting certain causes to be better to shape the exploration process. Thus, another goal of the current research was to understand how political motivations can impact active learning in a challenging dynamic explore-exploit task.

1.3. Relations between motivated reasoning and learning from experience

Historically, there have long been debates around whether a particular type of learning or reasoning pattern is best explained as motivated reasoning versus whether it could potentially be explained as rational Bayesian updating (Kunda, 1990; Nisbett & Ross, 1980). Though the debate about motivation versus rational reasoning has existed for many years, more recently, researchers from the tradition of optimal Bayesian learning theory have argued that many of the phenomena typically explained as motivated reasoning could potentially be explained as rational (e.g., Gershman, 2019; Tappin, Pennycook, & Rand, 2020). Even belief polarization in which two groups of people with opposing beliefs become even more polarized upon experiencing the same evidence can, in theory, be rational (Jern, Chang, & Kemp, 2014).

Some have gone even further to argue that the dichotomy between motivated reasoning versus rational updating should be dissolved. In particular, Kruglanski, Jasko, and Friston (2020) proposed that "motivated reasoning" and "active learning" are highly interrelated because, they argue, "all thinking is motivated." In research on active learning such as in an explore-exploit tasks, we often assume that the only motivation is to learn the best option as quickly as possible in order to exploit it. Kruglanski et al. would call this a motivation for *certainty*. They point out that reinforcement learning algorithms typically assume that the goal is to minimize uncertainty. However, in many real-world situations, people may prefer to be "blissfully ignorant"—they may be motivated to maximize uncertainty (e.g., avoid going to the doctor after experiencing a worrying symptom in order to avoid learning that they may have a serious illness). Kruglanski et al. also point out that in addition to potentially having motivations for certainty versus uncertainty, these motivations also exist at two levels: specific and nonspecific. A motivation for nonspecific certainty means wanting to know the outcome but not having a preference among the possible outcomes. In contrast, a motivation for specific certainty means wanting to know the outcome and having a preferred outcome. Kruglanski et al.'s point is that all information seeking or avoidance behavior, including the motivation to learn as quickly as possible (which is the assumption in traditional explore-exploit tasks), are driven by different types of motivations.

In the current work, our goal was to understand how prior political beliefs, preferences, or motivations, impact learning in an explore-exploit task that resembles testing different economic policies. We say beliefs, preferences, or motivations, to acknowledge that any effects could be driven through what has typically been considered less rational preferences or motivations or through a more rational process of updating prior beliefs; our goal was not to try

to separate the two. For example, a person might prefer an increase in border security funding due to a belief that it would be good for the economy, or they might prefer an increase in border security funding for other reasons (e.g., security), even if they do not necessarily think that it would improve the economy—they might even prefer an increase in border security funding despite believing that it would hurt the economy. In fact, due to the difficulty mentioned above of delineating rational versus irrational belief updating, we chose to study situations in which participants have preferences and beliefs in the same direction in order to study directionally motivated reasoning but to be agnostic about whether such motivated reasoning is rational or not. For concision, we use the term "political preferences" to refer to both preferences and beliefs, which were aligned.

How might political preferences impact learning in an explore-exploit task? At a general level, Kruglanski et al. (2020) describe motivation as instilling a sense of doubt or quelling the doubt. In regards specifically to information-seeking behavior and search, motivations can lead people to stop searching or to continue to seek further information (e.g., Ditto, Scepansky, Munro, Apanovitch, & Lockhart, 1998; Kruglanski, 2004). Research on information search in studies in which participants only have a non-directional motivation for accuracy, not a directional motivation (e.g., to uphold a political belief), can provide insight into potential learning that may be influenced by directional motivations.

Consider "confirmation bias" or what Klayman and Ha (1987; Klayman, 1995) call a positive test strategy. In a situation in which an individual is seeking to both learn the outcomes of one's actions in order to bring about desired rewards, a positive test strategy involves primarily choosing the actions that one already believes to produce the desired outcome (Klayman & Ha, 1987, p. 222). Research on how people actively test causal systems in order to uncover the causal structure has revealed a type of positive test strategy in which people sometimes conduct tests that produce many changes in the system (Coenen, Rehder, & Gureckis, 2015) despite not always being the most efficient way of narrowing down the set of causal structures.

Another aspect of rational information search that has been studied extensively is the "control of variables" strategy, which involves systematically testing a single variable at a time rather than making confounded changes to a system. Because testing variables in a controlled and systematic way is not intuitive for children (see Zimmerman, 2007, for a review), it is a core standard for science education (National Academy of Sciences, 2013, p. 52). Still, in certain instances such as sparsity (e.g., only one of many causes actually influences an effect), making multiple changes at once to a system to see if any of them make a difference, and then conducting subsequent tests to determine which one makes the difference can be more efficient (Coenen, Ruggeri, Bramley, & Gureckis, 2019). This example of testing multiple variables simultaneously could be viewed as both an example of positive testing (i.e., trying to make the outcome happen) as well as an example of when conducing confounded tests can be useful.

In sum, positive testing and conducting confounded testing are two information search strategies that are sometimes (but not always) suboptimal and are two general information search patterns that we investigated. We hypothesized that in a situation in which people not only have a motivation for accuracy but also have a political motivation (they think that some policies are better than others based on prior beliefs and/or they simply prefer some policies

more than others), the political motivation may exacerbate positive testing and confounded testing. In the next section, we explain more details of the current study and then propose specific hypotheses for how a political motivation may play out in an explore-exploit setting.

2. Summary of studies and hypotheses

Inspired by the examples of economic policies at the beginning of the introduction and how they can be highly ambiguous, we designed an explore-exploit paradigm in which the goal was to try to identify and implement the best policies. To accomplish this goal, we created a task in which participants learned about six policies. On each trial, they could choose between two different versions (e.g., increasing vs. decreasing border security funding) of each of the six policies. After observing the outcome of each trial, participants could decide to change any of the six policies for the next trial. Two of the policies worked fairly quickly; after a couple of trials of using the policy, it had its maximum impact. Two of the policies exhibited the temporal tradeoff of the short-term versus long-term costs versus benefits; these policies had more ambiguous impacts on the economy. In addition, there were also two policies for which the different versions of the policies made no difference, which we call "non-causal" policies.

We analyzed six sets of goals and questions. For concision, we list out these general goals in bullet form below, and then elaborate them afterward:

- Characterize testing and learning (Studies 1–3).
- Impacts of political preferences on testing during the learning phase (Studies 1–3).
- Impacts of political preferences on final judgments after the learning phase (Studies 1–3).
- Main effects and interactions of function ambiguity on the above questions (Studies 1–3).
- Impact of having strong versus neutral preferences on the above questions (Study 2).
- Whether prior knowledge of potential causal functions improves learning and reduces the effect of preferences (Study 3).

First, we sought to **characterize participants' testing habits and learning curves** during the learning phase to set up a context for subsequent questions. We analyzed the following three questions: (1) When did participants make changes to the policies? (2) How frequently did participants make a change and then hold other policies constant for a period of time to wait to see the impact of the change? (3) Did participants learn to exploit the optimal policy?

Second, we asked a set of questions about the impact of **having political preferences on participants' testing during the learning phase**. Specifically, we tested if political preferences would have an impact even when those preferences are technically irrelevant to this hypothetical task and participants were incentivized for accuracy. We analyzed the following questions: (1) Did participants make confounded changes, and if so, were confounded changes related to having preferences? (2) Was there evidence of positive testing—whether participants tested their preferred policies more or earlier than their non-preferred policies

and were more likely to make choices that optimized the output if the choice was congruent with their preferences?

Third, we analyzed the consequences of **political preferences on participants' judgments about the efficacy of the policies after the learning phase**. We analyzed the following questions: (1) Were participants' judgments about the efficacy of the policies after testing biased by their political preferences. (2) Could participants accurately identify the functional form of the policies, and was this biased by their political preferences?

Fourth, after demonstrating that participants had more difficulty learning the more ambiguous policies (mismatching short-term and long-term effects) than less ambiguous ones (matching short-term and long-term effects), we tested whether the motivated reasoning effects would be **exacerbated for the policies that were more ambiguous**. This hypothesis assumes that when a cause-effect relation is more ambiguous, it could reasonably be interpreted in multiple ways, allowing more room for a bias to seep in.

Fifth, in addition to the above questions, Study 2 compared causal judgments when participants had **strong versus neutral prior preferences**. The main question was whether having strong preferences on average (across preferences that happen to be right and preferences that happen to be wrong) leads to more biased testing and less accurate judgments, compared to when participants are more open-minded (have neutral preferences).

Sixth, in addition to the above questions, Study 3 tested whether causal learning and judgments are affected by having **more versus less knowledge** of the potential functional relations between the causes and effect. We hypothesized that having more knowledge about the potential ways that the causes could influence the effect would lead to better strategies for testing the policies overall, and particularly benefit learning about the policies that have different short- and long-term effects.

3. Study 1: Preference and ambiguity

3.1. Method

3.1.1. Participants

Fifty people participated via MTurk. Participants were paid $6.50 for participation (which amounted to approximately $8–10/h). In addition, participants could earn up to $3.00 in bonuses contingent upon performance and were informed of their bonus total after the completion of the study.

3.1.2. Design

Each participant learned about six economic policies. Each policy had two options that participants chose between. For example, for the policy of border security funding, the two options were "increasing border security funding" and "decreasing border security funding."

As explained below, with pretesting, we selected policies for which each individual participant had very strong preferences that one option was better for the economy and the other was

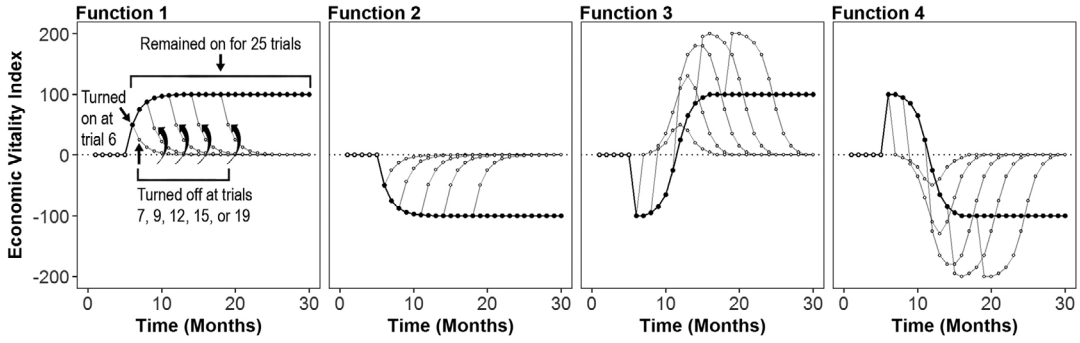

Fig 1. Illustrations of the payoff functions. Note: The first five trials in every graph represent an input of "off" (white dots). The solid black line shows the function if it is turn on at Trial 6 and left on until Trial 30. The gray lines show the pattern economic output if the function is turned "off" on Trial 7, 9, 12, 15, or 19, instead of being left on. Functions 1 and 2 are the "low ambiguity" (short-term and long-term effects match). Functions 3 and 4 are the "high ambiguity" (short-term and long-term effects are mismatched).

worse. For example, one participant might believe that increasing border security funding is better for the economy, and another participant might believe the opposite.

Independently from participants' preferences, we randomly assigned the six policies to one of six "payoff functions" (Fig. 1). The functions determined how each policy choice affected the economic output, which we called the "Economic Vitality Index" or EVI for short. The two options were randomly assigned to either be the better or worse states of the function. Thus, participants' preferences about the influence of a policy could either be preference-congruent (e.g., believing that more funding for border security is better for the economy, and indeed it was better) or preference-incongruent (e.g., believing that more funding for border security is better for the economy, but in fact it was worse). In addition, for two of the policies, the options made no difference.

The EVI was a sum of the six payoff function outputs (Fig. 1), plus a constant of 700, and a noise function. The noise function is a randomly generated Gaussian distribution with a mean of 0 and a standard deviation of 27. This degree of noise was selected to make the task hard but not impossible. If none of the policies were changed for a long period of time, the EVI would remain stable with the only fluctuation being due to the noise.

3.1.3. Economic functions

Functions 1 and 2 were "clear" in that the policies made a change relatively quickly, and the change lasted as long as the policy was used (Fig. 1). For Function 1, after the cause was turned from "off" to "on," it quickly produced an increase in the EVI. Function 2 was simply the opposite of Function 1 (negative coefficient signs); after the cause was turned from "off" to "on," it quickly produced a decrease in the EVI. The math behind these functions is based upon the idea of a decaying causal influence, similar to radioactive decay or a medication half-life. For example, imagine that the cause is a drug, which decays in half after each trial. At the end of Trial 1 after starting to take the medicine, 50% of the drug remains. At the end

of Trial 2 of taking the medicine, 50% remains from Trial 2, and 25% remains from Trial 1, producing a 75% effect. At the end of Trial 3, 12.5% remains from Trial 1, 25% remains from Trial 2, and 50% remains from Trial 3, and so forth. If the drug is repeatedly taken, the effect approaches 100%. If and when the policy is turned off, the remaining effect from prior trials continues to decay. Eq. 1 (Function 1) shows this function where policy p can be either on (1) or off (0) for each trial t.

$$EVI_t = 0.5\,(EVI_{t-1} + 100p_t) \tag{1}$$

Functions 3 and 4 are "ambiguous" in that the short-term effects of the policy are opposite to the long-term effects (Fig. 1). For Function 3, when the policy is turned from "off" to "on," it immediately has a negative influence on the EVI but eventually has a positive influence. Function 4 is the opposite; it initially has a positive influence but eventually has a negative influence. Functions 3 and 4 are similar to the function used in the melioration literature (e.g., Sims et al., 2013). These functions are somewhat analogous to a fixed-income security (e.g., treasury bond, certificate of deposit), and can also be viewed as somewhat analogous to the decision to buy versus rent a home. Importantly, Functions 3 and 4 have two defining features. First, there is a buy-in cost, which reduces the current EVI (analogous to spending money on the bond, certificate of deposit (CD), or a down-payment for a house, reducing one's current cash level). Second, there is a defined rate of return over time, and the cumulative return is larger than the initial cost, in this case twice as large. This means that if one keeps on buying the investment (using the policy) over and over again, initially the costs are substantial and one's cash deposits will be low. However, over time, as the dividends start to come in, one's cash level will be higher after repeatedly making the investment than if never investing at all. If one stops investing after having repeatedly invested, they will temporarily have an increased cash flow because of the incoming dividends, but over time, the benefits will taper away.

For Function 3, the investment function works such that when 100 EVI is invested, 200 EVI is returned over the following 10 trials. The rate of return on investment rises until it peaks five trials later and then decreases; this is why Function 3 p_{t-5} has the largest coefficient (50). The rate of return follows roughly a normal distribution from t–9 to t–1, which means that the cumulative payoff, if left on, is sigmoidal (Fig. 1).2 If this function is kept "on," then eventually every trial will return 200 EVI (a net gain of 100 EVI). Upon being set to "off," investments will no longer be made and only past investments will be returned (if any investments were made in the last 10 trials). Function 4 is the inverse to Function 3 (with negative instead of positive coefficients) and represents a policy that has short-term benefits but long-term consequences.

$$t_{-1} + 10p_{t-2} + 20p_{t-3} + 40p_{t-4} + 50p_{t-5} + 40p_{t-6} + 20p_{t-7} + 10p_{t-8} + 5p_{t-9} \tag{2}$$

For Functions 5 and 6, neither of the two options have any impact on the EVI, so they are called "non-causal." In cases where Functions 1–4 were at asymptote (i.e., a constant output of either +100 or –100), a change to Function 5 or 6 would be associated with a "noise function" only.

3.1.4. Procedures and measures

3.1.4.1. Initial instructions. Participants were told to imagine that they had just been elected the leader of a large industrialized country. As the leader, they have the responsibility to make important decisions about economic policies with the goal of maximizing economic output. Before taking office, they must first evaluate a set of economic policies, which will shape their economic platform.

3.1.4.2. Initial policy preferences. In order to choose six economic policies for each participant for which they had strong preferences that one option was better than the other, participants rated all 33 policies (Appendix A) on two questions. One question was about their subjective preference for a particular policy option (see Supplement A), and the other question their objective belief about whether the policy would have a positive or negative impact on the economy (see Supplement B). For example, for the policy about border security, participants were asked "Would you prefer the government decrease or increase border security spending?" on a scale of 1 = strongly prefer decreasing to 7 = strongly prefer increasing, and they were also asked "Do you believe decreasing or increasing border security spending is better for the economy?" on a scale of 1 = strongly believe decreasing border security spending is better for the economy to 7 = strongly believe increasing border security spending is better for the economy. The reason for using both questions was because, in pilot testing, we found that even though for most items the two questions were highly related, for a few items, they were not. We omitted these items going forward and decided to use both questions and averaged them together.

After participants answered all 66 questions, the computer selected the six policies for which participants had the most extreme ratings measured as the extremity of the average of the two questions. Most participants had at least six policies that they rated maximally extreme (either a 1 or a 7 on both questions). These six policies formed the participants' policy platform and were used in the subsequent tasks.

3.1.4.3. Party color selection. Next, participants selected a color (purple, pink, orange, yellow, green, or brown) to represent their political party. Red and blue were omitted from the choices due to the strong association these colors have with the two main political parties in the United States. After selecting a color, the participant was presented with a color that represented the opposition party.

3.1.4.4. Economic learning task. The economic learning task was the primary task for the study. Participants' goal was to select economic policies that produced the highest economic output and correctly assess which policies were best for the economy. Participants were told that they will receive a payment bonus based upon their average economic output for their time in office, relative to other participants' performance on the task, with a range of zero to two dollars. The six payoff functions were randomly assigned to the six policies.

Participants were presented with the six policies, randomly ordered on the screen (Fig. 2). The screen presented the participants' preferred option with a square of the color of their party and the non-preferred option with a square of the color of the opposing party. Initially, each

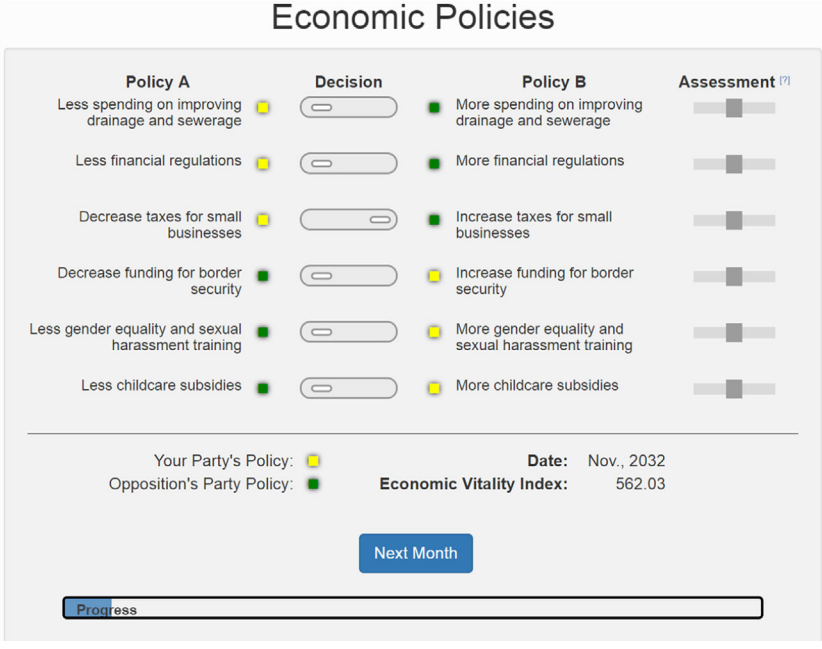

Fig 2. Screenshot of the economic learning task. On each trial, participants decide between Policy A and B for each of the six policies. In order to maximize the Economic Vitality Index on each trial, participants can update their beliefs about which policy is better using the assessment slider. Participants chose the color to represent their party, and this color indicates the participant's preferred policy stance as previously self-reported. The task lasts for 140 months where each trial is a month.

of the six policies was randomly set in either the "on" or "off" setting, which was framed as the policy selection of the prior administration. This random selection means that some of the prior policy decisions agreed with the participant's preference and some disagreed.

The screen also displayed the current "EVI," which is intended to be a made-up economic indicator similar to the gross domestic product or the stock market. All payoff functions were initially set such that they have already reached their asymptote (see Fig. 1) as if they have either been "on" or "off" for at least 20 trials.

Participants experienced 150 trials, and each trial represented 1 month in time. During the first 10 trials, the participants were told that they had not yet assumed power, so they just observed the six policies and observed the EVI of the prior administration. During these 10 trials, the six policies were held constant, and because the policies were already asymptote, the change in the EVI across the 10 trials was only due to the noise function.

After the 10th trial, participants were told that they had been elected to office and could set the policies for the next 140 trials however they choose by using the toggle switches in the "decision" column. After they set the policies as they wish, they pressed the "next month" button to go to the next trial, which revealed the EVI produced that month. At that point, they could again make changes to the policies. Additionally, throughout the task, participants were

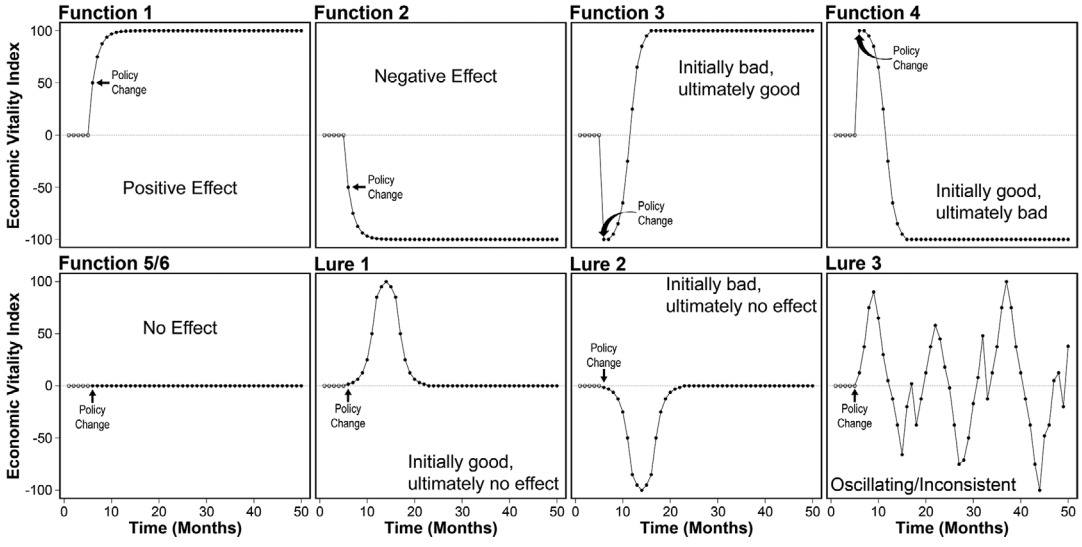

Fig 3. Function identification task payoff functions. Note: The above graphs were included as choices in the function identification task (Lure 3 was only included in Study 1 and removed for subsequent studies). The first five trials in every graph represent an input of "off" (white dots) before 45 inputs of "on" (black dots).

encouraged to use the slider in the "assessment" column to track which policy option, A or B, they thought was better. The slider scale was from –5 to +5 and was initially set to 0. Left means that Policy A was better, and right means that Policy B was better. After the last trial, participants were given one last opportunity to update their policy assessments.

3.1.4.5. Function identification. After the 150 trials were over, participants' understanding of how each policy works was tested by matching each of the six policies to a figure that presents eight possible functions (Fig. 3). These eight functions present the four unique functions from Fig. 1 plus the "non-causal" function, as well as three additional functions as lures. We also included textual descriptions of the influence of each policy. Instructions were provided stating that the graphs show different possibilities of what might happen if you switch from Option A (e.g., "decreasing border security funding") to Option B (e.g., "increasing border security funding") and to select the graph that they think would result from this policy change.

3.1.5. Individual differences

We initially had hypotheses about possible relations between individual difference measures (dogmatism, need for cognition, need for cognitive closure) and performance on the task, particularly around motivated reasoning. Though we measured these for Studies 1 and 2A, we found few reliable relations, so we stopped measuring them in future studies and do not report the results for concision.

3.2. Results

All data and analysis scripts are posted at https://github.com/caddickzac/Motivated-Reasoning-in-an-Explore-Exploit-Task.

We removed nine participants from our sample for making two or fewer policy changes throughout the entire learning task, which we viewed as a lack of engagement. In all, 41 participants submitted valid data for analysis.

For some analyses, we separated our analyses into two categories for causal and non-causal functions. The causal functions were Functions 1–4 that actually produce an effect and where one policy was better than another (e.g., Policy A > Policy B). The non-causal functions (Functions 5 and 6) had no impact regardless of which policy was chosen (i.e., Policy A = Policy B). For causal functions, a policy was called "preference-congruent" if the participant's preferred policy happened to be the optimal policy and was called "preference-incongruent" if the participant's preferred policy happened to be the suboptimal policy. For the non-causal functions, there is no such thing as preference congruence or incongruence because neither version of the policy is better than the other.

The results are separated into choices during the learning task, the influence of preference on choices during the learning task, judgments of policy efficacy after the learning task, and performance on the function identification task.

3.2.1. Choices during the learning task

In this section, we examined how participants tested the policies during the learning task. Some of these analyses are provided simply to provide evidence of learning context, but most serve a dual purpose of characterizing participants' testing habits more generally while also revealing differences in how participants tested the policies that they preferred versus the ones that they did not prefer. Because of the dual role of many of these analyses, they do not follow the order of the questions posed in the introduction. We first analyze choices made at the beginning of testing and then analyze choices throughout the learning task.

3.2.1.1. Switching multiple policies to the preferred option at the beginning of learning (Table 1, Figs. 4 and 5). One dramatic finding was that on Trial 1, many participants switched multiple of the policies initially set to their non-preferred option to their preferred option. Table 1 categorizes and tallies every change to each policy. On Trial 1, participants made a total of 83 changes to the policies, and 72 of these involved changing two or more policies simultaneously (confounded) instead of changing one policy at a time (controlled) toward the preferred option.

Fig. 4 shows the raw number of policies that were changed each trial, broken down by whether a confounded change or controlled change was made. This figure shows the spike in confounded changes on Trial 1. It also shows that after Trial 1, the majority of the changes were controlled, not confounded (see also Table 1). Furthermore, participants made more changes to the policies earlier on, especially during the first 40 or so trials, and then made fewer changes toward the end of the learning task. This pattern makes sense in that this is an

Table 1
Counts of testing instances by type

		First Trial		All Other Trials	
Study	*Testing Type*	*To Preferred*	*To Non-Preferred*	*To Preferred*	*To Non-Preferred*
1	Confounded	72	2	128	130
1	Controlled	8	1	339	383
2A	Confounded	69	9	117	113
2A	Controlled	13	0	370	410
2B	Confounded	206	13	566	571
2B	Controlled	38	9	1350	1424
3	Confounded	103	16	178	169
3	Controlled	9	1	527	575

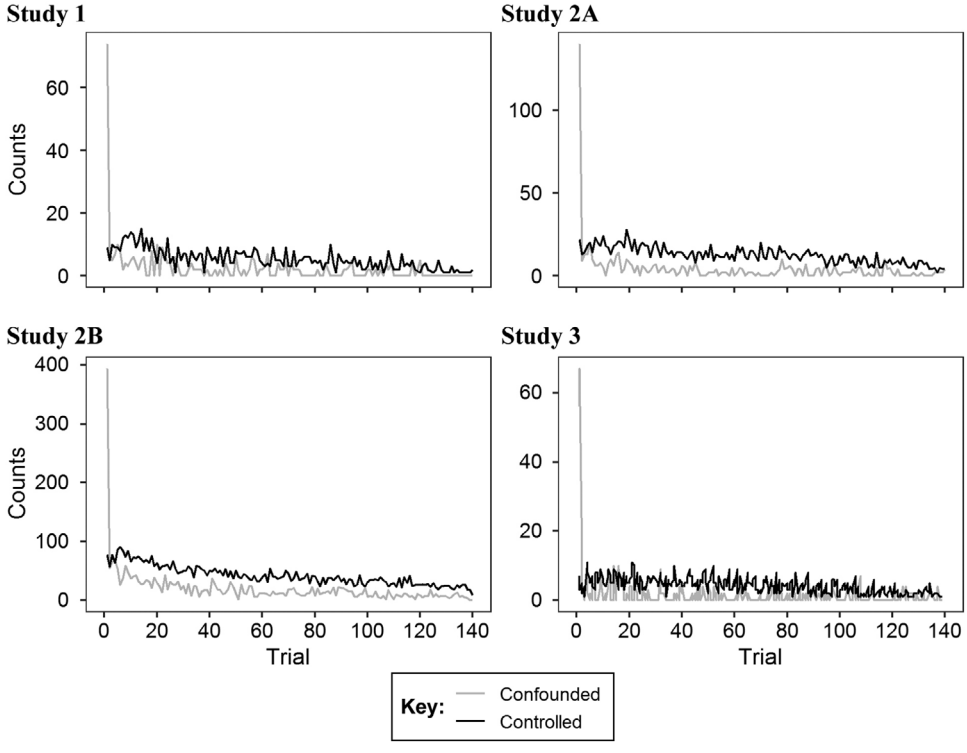

Fig 4. Number of controlled (changes to a single policy) and confounded changes (changes to two or more policies) per trial. Note that during Trial 1, participants mainly made confounded changes, and for the rest of the trials, participants mainly made controlled changes. Over the course of the study, participants made fewer policy changes as they settled on the policies they thought were best.

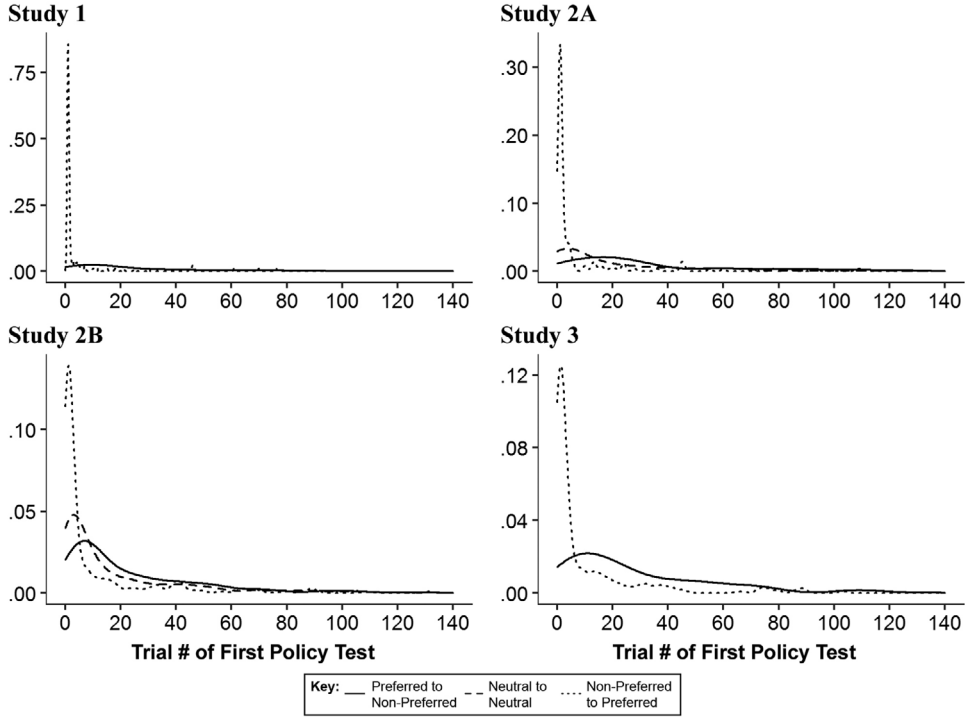

Fig 5. Density plots for the number of trials until testing by preference. Note: The Y-axis is the density probability estimation for first testing a policy. The X-axis is the trial range (1–140) for the learning task. All four studies show that participants quickly switched non-preferred policies to preferred and switched preferred policies to non-preferred later. Neutral policies tended to be switched after non-preferred policies and before preferred.

explore-exploit task, so as participants' beliefs about the optimal policies settle, they should gradually make fewer changes.3

We also ran an inferential statistical test to examine whether participants switched policies toward their preferred option earlier than toward the non-preferred option. If so, this would be evidence for a positive test strategy because a participant cannot learn anything about a given policy until a switch happens. Suppose that a policy was randomly set to the non-preferred option at the start. We hypothesized that very early on participants would tend to switch it to the preferred option. In contrast, we hypothesized that if a policy was randomly set to the preferred option at the start, that it would take longer for participants to switch it to their non-preferred option. Indeed, Fig. 5 shows a dramatic difference in how quickly such switches were made.

Because time until testing is positive and was skewed, a generalized linear model with a gamma distribution and an inverse link function was used to predict when a policy was first tested by policy preference at the start.4 A random intercept for subject and a random slope for policy preference at the start was included in the model. Participants switched non-preferred

Table 2
Average number of times that a participant made a change to an individual policy and held all other policies constant for a given number of trials

	Number of Trials System Held Constant							
Function Type	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8+*
Study 1								
Low ambiguity	0.87	0.51	0.43	0.30	0.21	0.06	0.06	0.43
High ambiguity	0.94	0.55	0.38	0.22	0.18	0.10	0.10	0.61
Non-causal	0.70	0.48	0.41	0.28	0.18	0.15	0.11	0.67
Study 2A								
Low ambiguity	0.66	0.56	0.46	0.36	0.27	0.16	0.11	0.49
High ambiguity	1.01	0.60	0.43	0.28	0.16	0.09	0.07	0.68
Non-causal	0.78	0.53	0.46	0.39	0.27	0.18	0.19	0.66
Study 2B								
Low ambiguity	1.28	0.55	0.38	0.25	0.15	0.14	0.08	0.45
High ambiguity	1.47	0.49	0.33	0.20	0.11	0.09	0.05	0.51
Non-causal	1.33	0.59	0.40	0.32	0.22	0.19	0.13	0.75
Study 3								
Low ambiguity	0.40	0.48	0.33	0.22	0.21	0.17	0.05	0.40
High ambiguity	0.66	0.51	0.22	0.21	0.15	0.13	0.08	0.56
Non-causal	0.24	0.34	0.37	0.33	0.24	0.16	0.12	0.76

policies to preferred earlier than they switched preferred to non-preferred ($\beta = –0.30$, $SE = 0.05$, $p < .001$).

In summary, we found that on Trial 1, participants made many confounded changes toward their preferred policy and that changing toward preferred occurred earlier than changes toward non-preferred are evidence of a positive test strategy. From the perspective of efficient learning, confounded changes are hard to justify because they do not reveal the unique contribution of each policy. However, if participants truly believed that their preferred policies were better for this artificial system, then one could argue that these findings could be an attempt to optimize the output. Regardless, these findings reveal an important pattern of information search when preference exists.

3.2.1.2. Changing a policy and holding others stable for periods of time (Table 2). The following analyses examine the testing choices throughout the learning process and not only right at the beginning of learning. In order to learn the influence of each policy, participants need to make controlled (unconfounded) changes to that policy. Furthermore, given that some of the policies need to be tested repeatedly to show their full influence, participants actually need to make a change to a single policy, and hold other policies constant for a number of subsequent trials, to accurately learn about these policies.

For the low ambiguity functions, participants do not need to hold other functions constant for very long to see the impact. For the low ambiguity functions, after one, three, and five

trials of testing, 50%, 88%, and 98% of the influence was experienced, respectively, because the function asymptotes fairly quickly (Fig. 1). However, for the high ambiguity functions, participants had to hold other policies constant for at least seven trials, which is when the long-term effect starts; at eight trials, 65% of the long-term trend is experienced.

We analyzed whether participants held other policies constant long enough to learn about the long-term implications of the policies. To do so, we measured how long a participant tested a single policy before making another change to the system. The length of a test is defined by the number of trials the participants held other policies constant after making a change to a single policy.

Table 2 presents descriptive statistics for this analysis. The numbers 1–8 show the average number of times over the 140 learning trials per participant that an individual policy was tested and other policies were held constant for a given number of subsequent trials. For example, the 0.87 in the top left cell means that, for each of the two low ambiguity policies, participants on average made a change to the policy and then made another change to the system on the subsequent trial almost once (technically 0.87 times) during learning. In addition, participants also made a change to the low ambiguity policies, held everything constant for one trial, and then made a change on the subsequent trial 0.51 times. They also made a change to these policies, held everything constant for two trials, and then made a change on the subsequent trial 0.41 times, and so forth. We tallied the number of each of these hold times up to seven separately and then aggregated together all changes that involved holds of eight or more trials.

Participants tended to make changes to policies and then not wait very long before making subsequent changes—the most frequent pattern was to make a change on the subsequent trial (1) or just wait two or three trials. The low hold times mean that it would be fairly easy for participants to learn the low ambiguity policies and non-causal policies because these do not need long hold times.

However, given the low hold times, it would be harder for participants to learn about the high ambiguity policies. For each high ambiguity policy, participants made a change to the policy and then held the system constant for at least eight trials only 0.61 times on average per policy. Furthermore, they made considerably more changes in which they held the system constant for only one, two, or three trials, so they were more exposed to the short-term rather than the long-term impacts of these policies.

In sum, this pattern of testing means that participants should be able to learn about the low ambiguity and non-causal policies. However, for the high ambiguity policies, they had considerably more evidence about the short-term impacts than the long-term impacts of the policies, so they may incorrectly assess the value of these policies.

3.2.1.3. Never testing bias by preference (Fig. 6). Though most participants tested both versions of each policy, on average across all participants and all six policies, 7.72% of policies were never changed to test the version that was not selected at the start of the learning task. We hypothesized that participants might decide to leave policies that were initially set in their preferred state alone, never testing them, even though this would mean that they would

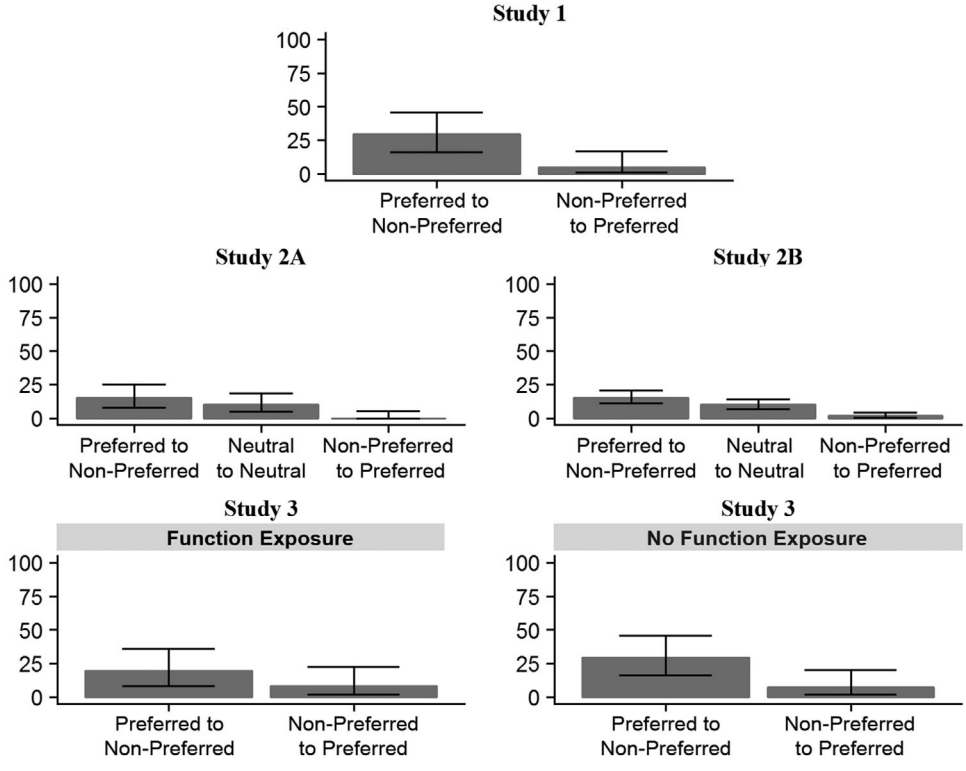

Fig 6. Percentage of policies never tested by initial policy state. Errors bars represent 95% confidence interval. Participants were more likely to never switch policies from preferred to non-preferred than non-preferred to preferred.

not have an opportunity to determine which version was actually more effective, which would presumably lower their bonus for the task.

To analyze this,5 we coded each participant as whether or not they failed to test at least one policy that was initially set to the preferred option, and whether or not they failed to test at least one policy that was initially set to the non-preferred option. We compared these using McNemar's test of paired proportions. Participants were more likely to have not tested a policy at all if the initial testing required switching a preferred policy to a non-preferred policy (29.27%),6 versus if the initial testing required switching a non-preferred policy to a preferred policy (4.88%), $\chi^2(1) = 6.75$, $p = .009$ (Fig. 6).

Never testing (Fig. 6) and the number of trials until testing (Fig. 5) are closely related. If a policy was initially set to the preferred option, participants were more likely to either take a long time before testing the non-preferred option or sometimes to never test it. In contrast, if the policy was initially set to the non-preferred option, participants fairly quickly switched it to the preferred option, and only very rarely did they never test the preferred option.

3.2.1.4. Percent of trials during which the preferred option was selected (Fig. 7). For

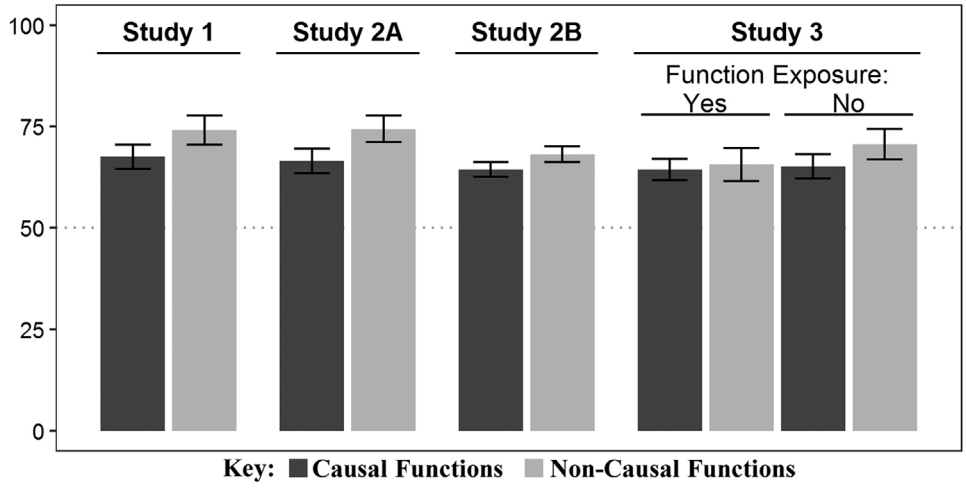

Fig 7. Percentage of trials and 95% CIs during which the preferred option was selected, compared to chance (50%). In all conditions, participants selected their preferred policies option more than 50% of the time.

the four causal policies for each participant, we calculated the average percentage of trials in which the policy was set to the participants' preferred option (out of 4 causal policies $\times$ 140 trials $=$ 560 observations per participant). If participants were not biased and simply tried to figure out which policy option was better, then they would try their preferred and non-preferred policy options equally. However, we hypothesized that they would try their preferred policy options more frequently than their non-preferred policy options. Using a one-sample *t*-test (see Fig. 7 for means), participants were more likely to select their preferred policies, compared to chance (50%), both for the causal functions $t(40) = 5.93$, $p < .001$, $d = 0.93$, and for the non-causal functions $t(40) = 6.79$, $p < .001$, $d = 1.06$.

3.2.1.5. Percentage of trials the optimal policy was selected by preference (Fig. 8). Because this task is an explore-exploit task, not a pure explore task, it is rational for participants to test the versions of the policies that they actually believe to be better. Fig. 8 shows the percentage of trials during which the optimal policy option was chosen. Fig. 8a shows learning curves; Fig. 8b shows the average over all 140 learning trials, which corresponds to our inferential statistical analyses.

We had three hypotheses. First, we expected that learning would be easier for the low ambiguity functions than the high ambiguity functions, so we expected that participants would more frequently test the optimal version of the policy for low ambiguity functions; both can be seen in Fig. 8a; by the end of learning, participants were above chance for the low ambiguity functions. In contrast, for the high ambiguity functions, they were below chance for the preference-incongruent functions and only a bit above chance for the congruent functions. This difficulty learning about the high ambiguity policies fits with the finding above that participants sometimes did not test these policies long enough to uncover their long-term influence.

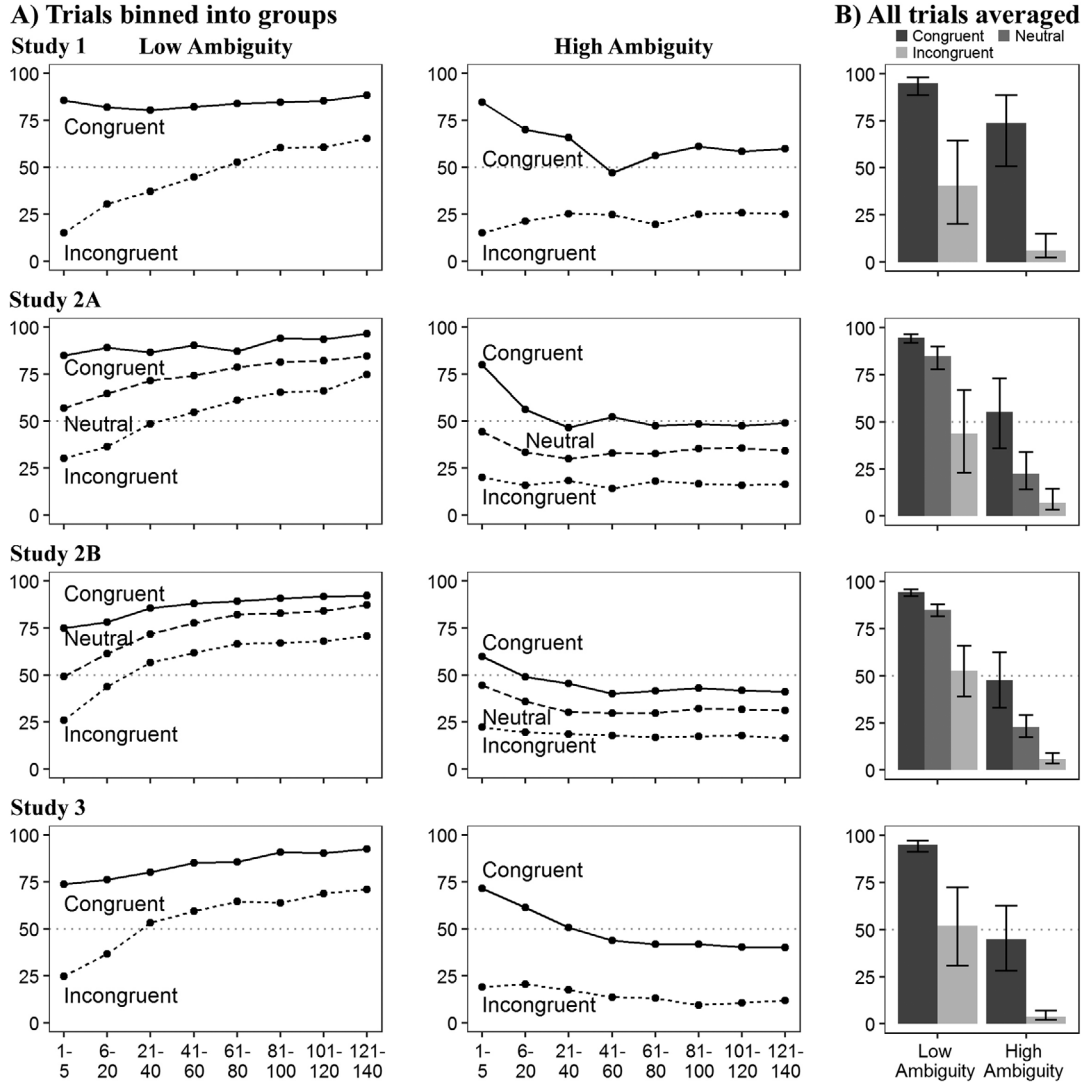

Fig 8. Percentage of trials the optimal policy was selected by preference and ambiguity. Note: Error bars represent 95% confidence intervals. Chance is 50%. (a) Represents learning curves. Over time participants tend to choose the optimal policy for the low ambiguity functions but are less successful for the high ambiguity functions. Additionally, participants were more likely to choose the optimal policy when it was preference-congruent (i.e., their preferred policy was optimal) than preference-incongruent (i.e., their preferred policy was not optimal); (b) collapses the data in (a) across the 140 trials.

Second, knowing that the participants tended to try their preferred policy options more than their non-preferred policy options, we hypothesized that participants would be more likely to test the optimal version of the policies when the optimal version was also their preferred version (preference-congruent), compared to when the optimal version was their

non-preferred version (preference-incongruent). Fig. 8a and b reveal that participants chose the optimal policy more often when it was preference-congruent.

Third, in the introduction, we had predicted that motivated reasoning could be exacerbated with the more ambiguous functions. Thus, we expected there to be an interaction between preference-congruence and ambiguity such that the difference between the percentage of testing the optimal policy option would be greater between preference-congruent versus incongruent functions for the high ambiguity functions than the low ambiguity functions.

For this analysis, we calculated the mean percentage of trials out of 140 that were set to the optimal choice for each policy (Fig. 8b). We then conducted a random effects regression with a by-subject random slope for congruence and ambiguity. To get the model to converge, we dropped the random correlations between slopes7 and also dropped the random slope for the interaction. Both predictors (and all other similar regressions in this manuscript) used effects coding with +0.5 preference-congruent and –0.5 for preference-incongruent and +0.5 for less ambiguous and –0.5 for more ambiguous. This analysis only includes the causal functions because non-causal functions cannot be categorized as preference-congruent or incongruent.

As expected, participants were more likely to select the optimal policy when it was also their preferred policy (preference-congruent as opposed to preference-incongruent; $\beta = 0.38$, $SE = 0.06$, $p < .001$) and if it was less ambiguous ($\beta = 0.26$, $SE = 0.06$, $p < .001$). However, contrary to our hypothesis, there was no significant interaction between preference-congruence and ambiguity ($\beta = –0.06$, $SE = 0.10$, $p = .497$).

3.2.1.6. Summary of choices during the learning task. The previous analyses revealed a number of specific ways in which motivated reasoning is revealed in searching for better policies. At the beginning of learning, participants frequently made multiple confounded changes to switch policies into their preferred state. During learning, participants tested their non-preferred policy options less frequently and were more likely to never test their non-preferred policy options, compared to their preferred policy options. These biases also meant that they were less successful at using the optimal policy when the optimal policy was preference-incongruent. The next section focuses on participants' judgments about the policies.

3.2.2. Judgments of policy efficacy after the learning task

3.2.2.1. Causal functions (Fig. 9). We had similar hypotheses about participants' final judgments of policy efficacy as for the previous section on the frequency of testing the optimal policy; we expected their final judgments to be more accurate for the low than high ambiguity policies, more accurate for the preference-congruent than incongruent policies, and we expected an interaction such that the effect of preference congruence would be magnified for high ambiguity policies.

The dependent variable was the error in the policy assessments. This was measured by taking the absolute value of the difference between the slider position from the ideal slider position. For example, if Policy B is in fact better (which corresponds to +5), and a participant sets the slider to +2, they are three points away from the correct answer. This computation is used to examine if the accuracy of participants' policy assessments differed by whether a

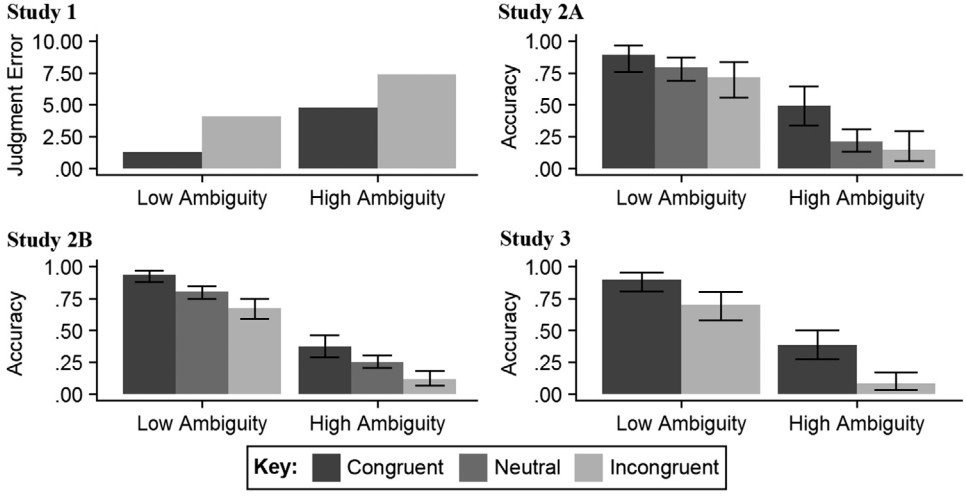

Fig 9. Accuracy of judgments of policy efficacy by congruence and ambiguity for causal functions. Participants were more accurate at identifying the optimal policy for low ambiguity than high ambiguity functions and for preference-congruent policies (i.e., the preferred policy was optimal) than preference-incongruent policies. Note: Study 1 is a measure of the error in judgment, whereas Studies 2A, 2B, and 3 are accuracy percentages. Error bars represent 95% confidence from a binomial test for each subgroup and did not account for repeated measures. There are no error bars for Study 1 because the data were analyzed differently using non-parametric tests.

preferred policy was optimal (or not). See Fig. 9 for descriptive results from all studies; note that future studies use a different dependent measure of accuracy.

Because we randomized whether a preferred policy was optimal or not, there was not necessarily one congruent and one incongruent policy for every function type. The below analysis was conducted at the user level; when multiple measurements were present, these judgments were averaged.8 We also used non-parametric tests due to violations of normality and left the test of the interaction between preference congruence and ambiguity to later studies.

As expected, a Wilcoxon rank sum found that participants' judgment error was lower in the preference-congruent condition, when their preferred policies happened to be optimal (median = 3) than in the preference-incongruent condition (median = 6), $U = 354.50$, $p < .001$, $r = .440$, 95% CI = 0.249–0.620). Additionally, participants were more accurate for the low ambiguity functions (median = 2) than for the high ambiguity functions (median = 5), $W = 94.50$, $p < .001$, $r = .49$, 95% CI = 0.32–0.65.

3.2.2.2. Non-causal functions (Fig. 10). We also examined how accurately participants assessed the non-causal functions and whether participants tended to select their preferred policy option as being better, despite neither policy option being better. To do this, participants' judgments were coded such that 0 represented an assessment that the preferred option produced a much better outcome than the non-preferred option (which was incorrect), 5 represented a correct assessment that there is no difference between the two options, and 10 represented an assessment that the non-preferred option produced a much better outcome

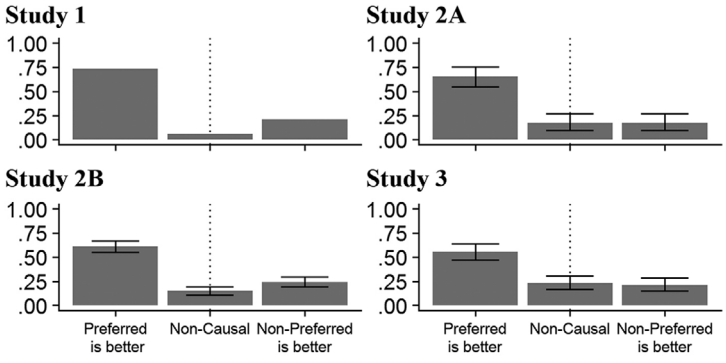

Fig 10. Judgments of policy efficacy after the learning task for non-causal functions. Instead of saying that non-causal policies were non-causal, participants tended to say that non-causal policies that they preferred were better than non-causal policies that were non-preferred. Note: Y-axis is the percentage of judgments. The dotted line represents the correct judgment. Study 1 data were collapsed into three bins for visual congruence with subsequent studies but not for analysis. Policies for which a participant held a neutral preference prior to the study were omitted. Error bars represent 95% confidence interval.

(which was also incorrect). In Fig. 10, instead of using the 11-point scale, we plot the three groups <5, 5, and >5 for consistency with subsequent studies. Participants very rarely concluded correctly that there was no difference and usually concluded that their preferred policy option was better.

To determine if participants were more likely to assess their preferred policy as being better, we took the average of the scores for the two non-causal functions. A one-sample Wilcoxon signed-rank test against 5 confirmed that the judgments were biased toward the preferred policy (median $= 4$), $W = 42.50$, $p < .001$, $r = .65$.

3.2.3. Function identification

At the end of the study, participants were asked to match each of the six policies to a figure that represented different policy functions. Responses were scored as correct or incorrect to test if participants were able to accurately identify the mathematical function for each policy.

3.2.3.1. Causal functions (Table 3). A mixed effects logistic regression9 analysis was conducted to test for differences in the ability to correctly choose the graph that represented functions by preference-congruence, ambiguity, and their interaction. The model used a by-subject random intercept and random slopes for all three predictors.

Participants' accuracy at function identification did not differ based upon congruence ($\beta = 0.02$, $SE = 0.45$, $p = .973$). However, participants were more likely to correctly identify a function if it was less ambiguous ($\beta = 1.48$, $SE = 0.46$, $p = .001$). No interaction between congruence and ambiguity was found ($\beta = 1.15$, $SE = 0.95$, $p = .229$). In general, the accuracy was fairly low, which is expected given that it was a surprise task, participants did not know the set of possible functions in advance, and furthermore, it is an unusual task; people rarely have to interpret graphs of abstract functional forms in other settings.

Table 3
Accuracy of function identification for causal policies. Participants were better at correctly identifying the functions for low ambiguity policies than high ambiguity policies

Study	*Ambiguity*	*Function Exposure*	*Preference*			*Total*
			Congruent	*Neutral*	*Incongruent*	
1	High	No	0.08	–	0.13	0.11
	Low	No	0.39	–	0.27	0.33
2A	High	No	0.24	0.13	0.10	0.15
	Low	No	0.33	0.27	0.27	0.28
2B	High	No	0.21	0.15	0.14	0.16
	Low	No	0.34	0.21	0.22	0.25
3	High	No	0.20	–	0.05	0.13
	Low	No	0.27	–	0.18	0.22
	High	Yes	0.14	–	0.14	0.14
	Low	Yes	0.44	–	0.29	0.37

Note. Study 1 chance = 12.50%. Study 2A, 2B, and 3 chance = 14.29%.

Table 4
Accuracy of function identification for non-causal policies. Participants rarely identified the non-causal functions as being non-causal; they typically thought they had some sort of causal influence

Study	Exposed to Mechanism	Has Preference		Total
		Yes	No	
1	No	0.02	–	0.02
2A	No	0.03	0.11	0.07
2B	No	0.06	0.18	0.12
3	No	0.11	–	
3	Yes	0.15	–	

Note. Study 1 chance = 12.50%. Study 2A, 2B, and 3 chance = 14.29%. "Exposed to mechanism" refers to the function exposure manipulation used in Study 3.

3.2.3.2. Non-causal functions (Table 4). Table 4 shows the mean accuracy of correctly identifying that the non-causal policies were non-causal. Participants were very rarely accurate, only 2% of the time; chance performance given the eight graphs were 12.50%.

3.2.4. Relations between choices during the learning task and judgments of policy efficacy

We sought to examine relations between choices during learning and judgments afterward. Though there are some relations, they are not especially reliable across studies and also are not directly related to questions around motivated reasoning. Thus, we report these findings in Supplement C.

3.3. Discussion

Here, we revisit the questions posed in the introduction. With regards to characterizing participants' testing habits and learning curves, we found the following. First, participants tended to make many changes early on, especially on the first trial, and then fewer over time. Second, participants often did not hold the system stable for very long after making a change to a policy. Third, participants learned to exploit the low ambiguity policies but had a harder time exploiting the high ambiguity policies.

With regards to the second set of questions, in Study 1, we found that participants' testing behavior was greatly influenced by their policy preferences. During learning, participants exhibited four patterns of motivated reasoning. First, at the very beginning of testing, participants tended to switch multiple policies from the non-preferred state to the preferred state, which means that these changes tended to be confounded. Second, in instances in which participants did not test a policy at all, the policy tended to already be set to the preferred policy. Third, participants tended to test the preferred option of the policy more overall than the non-preferred option. Fourth, participants more frequently tested the optimal policy if it was also their preferred policy.

With regards to the third question about the impact of political preferences on judgments of the policies after learning, we found the following. First, participants were more likely to correctly assess the policies (to correctly determine which version of the policy is better) when they were preference-congruent (when the participants' preferred the option that happened to be better). Second, participants' ability to identify the underlying function was not influenced by their preferences; however, in general, this ability was low especially for the highly ambiguous functions.

Given the converging evidence that strong preferences can alter behavior and lead to biased conclusions, the next study investigated whether it is better to be open-minded (have neutral beliefs) than strong beliefs when learning cause-effect relationships.

4. Study 2: Strong versus neutral preferences

Study 2 extended Study 1 by comparing policies for which participants did not have prior preferences versus policies for which participants had prior preferences. We hypothesized that participants may be more accurate at learning about policies when they do not have strong preferences about them as opposed to when they do have strong preferences.

One reason that causal learning might be worse for policies for which they have strong preferences is that their preferences may bias their ability to learn about the policy if they just assume that one version of the policy is better and fail to sufficiently test it. Study 1 showed that participants tended to choose their preferred policy options more often than their non-preferred policy options. This tendency could impede causal learning for both preference-congruent and preference-incongruent policies because participants tended to mainly select their preferred policy option rather than switch between the two options; switching is neces-

sary to test which option is better. In contrast, if a participant has no preferences, the lack of a bias could lead to more accurate learning.

At the same time, there are other reasons that learning could be better for policies that participants have preferences about. For example, participants may care more about these policies and pay more attention to them while testing.

There are also reasons to hypothesize that having preferences versus not having preferences could lead to the same learning and or judgments on average. Consider making a final judgment about which policy is better, Option A or B. Suppose that a participant prefers A, and they bias their final judgment toward A, to some degree. For policies that are preference-congruent (Option A really is better than B), this bias would lead to a more accurate judgment than they might otherwise have made. However, for policies that are preference-incongruent (Option B really is better), this bias would lead to a less accurate judgment. Potentially the benefit from preference-congruence and the cost from preference-incongruence could wash out, compared to a judgment about a policy for which a participant does not have a preference.

In sum, there are many potential reasons for better, worse, or no difference in performance about policies for which participants do versus do not have a preference.

4.1. Method

Study 2 was very similar to Study 1 except for the following changes. First, instead of only selecting policies for which participants had strong preferences, three policies with strong preferences and three policies with neutral preferences were selected for each participant. Policies with neutral preferences were defined as having ratings (the average of the preference and belief ratings) between 3 and 5 on the 7-point scale. We first selected policies with ratings of exactly 4 (the middle of the scale), but if a participant did not have enough policies of exactly 4, then policies with ratings of 3.5 and 4.5 were chosen next, followed by policies with ratings of 3 and 5. Additionally, since participants made two ratings for each policy, we selected policies with the smallest difference between the two ratings first (i.e., an average rating of 4 could be due to two ratings of 4 and 4 or ratings of 3 and 5; in these cases, we chose 4 and 4). In cases where there were not enough ratings that fell into the "neutral preference" or "strong preference" bins, it was possible for a participant to have more neutral policies than strong preference policies (or vice versa). Most participants had a perfect balance between strong and neutral policies (MTurk sample: 99%; Introduction to Psychology (Intro. Psych sample): 96%).

Second, the policy assessment judgments that participants made during the learning task and right after Trial 140 were changed to a 3-point scale ("Policy A is better," "No Effect/Uncertain," "Policy B is better") instead of the 11-point scale. This was because, in Study 1, participants mainly used the extremes of the scale resulting in a non-normal distribution.

Third, during the function identification task, we removed the "oscillating" lure plot because so many participants chose it, and success rates were very low. We were worried that participants chose it because it looked like the noise function we were using rather than any of the causal functions.

Fourth, we collected two samples for Study 2; MTurk (Study 2A) and undergraduate Intro. Psych. students (Study 2B). Because of the similarity of results, we report the results side-by-side.

4.1.1. Participants

In the MTurk sample, there were 102 participants. Participants were paid $6.50 for participation (which amounted to approximately $8–10/h) with an opportunity to be awarded up to $3.00 in bonuses contingent upon performance. We removed 12 participants for making fewer than two policy changes throughout the entire learning task. One participant was removed who both had the shortest completion time and selected the middle answer for every item in the individual difference measures, which we viewed as evidence for not meaningfully participating. Additionally, another was removed because they only partially completed the full study. In all, 88 participants were included in our analysis.

In the Intro. Psych. participant pool, there were 385 participants. Participants received course credit for participation. We removed 101 participants for making fewer than two policy changes throughout the entire learning task. The higher rate of disengagement, compared to the MTurk sample, could be due to the lack of payment and bonus. In all, 283 participants were included in our analysis.

4.2. Results

The organization of the results is similar to Study 1, focusing first on choices during the learning task, then judgments of policy efficacy after the learning task, and finally the function identification task. Within each section, we first do the same analysis as in Study 1 (e.g., comparing preferred vs. non-preferred or preference-congruent vs. incongruent, etc.). Then, when possible, we followed up the analysis with a comparison between policies for which participants had neutral preferences versus policies for which they had strong preferences. Unless specified, all analyses were conducted the same way as in Study 1.

4.2.1. Choices in the learning task

4.2.1.1. Switching multiple policies to the preferred option at the beginning of learning (Table 1, Figs. 4 and 5). Similar to Study 1, on the first trial, participants tended to make many confounded changes—they switched multiple policies from the non-preferred option to the preferred option (Table 1). After the first trial, the majority of changes to policies were controlled, not confounded (Table 1 and Fig. 4).

To test whether participants switched non-preferred policies to preferred earlier than the reverse (Fig. 5), we first replicated our finding from Study 1, excluding neutral policies. Indeed, participants switched non-preferred policies to preferred earlier than they switched preferred to non-preferred (MTurk: $\beta = -0.31$, $SE = 0.05$, $p < .001$; Intro. Psych.: $\beta = -0.21$, $SE = 0.03$, $p < .001$).

We then compared policies for which participants had neutral preferences versus policies for which they had strong preferences and tested whether they tested neutral policies (switching from one neutral option at the start to the other) earlier or later than policies for which

they had strong preferences (which includes both switching from preferred at the start to non-preferred or non-preferred at the start to preferred). We again used a generalized linear model with a gamma distribution and an inverse link function to predict when a policy was first switched. The model included a by-subject random intercept with a random slope of preference strength (strong vs. neutral). We did not find an overall difference between strong and neutral preferences (MTurk: $\beta = –0.01$, $SE = 0.04$, $p = .845$; Intro. Psych.: $\beta = 0.01$, $SE = 0.02$, $p = .607$). As can be seen in Fig. 5, participants' first switch of a neutral policy tended to be after their first switch of a non-preferred to preferred and before their first switch of a preferred to non-preferred policy.

4.2.1.2. Changing a policy and holding others stable for periods of time (Table 2). The findings for Study 2 are similar to those for Study 1. After making a change to a policy, participants tended to hold the system stable for only one or two trials before making a subsequent change. This meant that they had relatively good evidence about the short-term impacts of the policies, but that they did not produce good evidence about the long-term impacts of the policies; this is especially problematic for the high ambiguity policies for which the short and long-term impacts are in conflict.

4.2.1.3. Never testing bias by preference (Fig. 6). Though most participants tested both versions of each policy, on average across all participants and all policies, 4.55% of policies in the MTurk sample and 5.77% in the Intro. Psych. sample were never changed. Participants were more likely to have not tested a policy at all if the initial testing required switching a preferred policy to a non-preferred policy (18% for MTurk, 14% for Intro. Psych) versus if the initial testing required switching a non-preferred policy to a preferred policy (0% for MTurk, 1.40% for Intro. Psych), and these proportions were significantly different (MTurk: McNemar's $\chi^2(1) = 8.10$, $p = .004$; Intro. Psych.: McNemar's $\chi^2(1) = 23.31$, $p < .001$).

Next, we compared policies for which participants had neutral preferences versus policies for which they had strong preferences on whether the policies differed in never being tested. To allow for a within-subjects comparison, we omitted participants that had all three of their "strong preference" policies set to either preferred or non-preferred; the analysis only included those who had at least one strong preference that was preferred and one that was non-preferred at the start (MTurk: $N = 57$; Intro. Psych. $N = 214$).

We found that participants were equally likely to have not tested a policy at all, if the initial testing required switching a neutral policy to a competing neutral policy (MTurk: 12%; Intro. Psych: 10%) versus if the initial testing required switching a strong-preference policy to a competing strong-preference policy (MTurk: 18%; Intro. Psych: 14%; MTurk: McNemar's $\chi 2(1) = .57$, $p = .450$; Intro. Psych.: McNemar's $\chi 2(1) = 2.70$, $p = .100$). In Fig. 6, it can be seen that the rates of non-testing neutral-to-neutral switches are in-between the rates of the other two groups.

In sum, participants' preferences did make a difference as to whether they tested versus never tested a policy; however, this difference is driven by whether the policy was initially set to the preferred versus non-preferred option, not by having preferences versus not having preferences.

4.2.1.4. Percentage of trials during which the preferred option was selected (Fig. 7). For the causal functions, participants tended to test their preferred policies more often than their non-preferred policies for both the MTurk ($M = 66\%$; $SD = 29\%$, $t(87) = 5.40$, $p < .001$, $d = 0.58$) and Intro. Psych. samples ($M = 64\%$; $SD = 30\%$, $t(279) = 7.99$, $p < .001$, $d = 0.48$).

For the non-causal functions, participants were much more likely to test the preferred policy than the non-preferred policy for both the MTurk ($M = 74\%$; $SD = 27\%$, $t(70) = 7.53$, $p < .001$, $d = 0.89$) and Intro. Psych. samples ($M = 68\%$; $SD = 30\%$, $t(225) = 9.25$, $p < .001$, $d = 0.62$) .

4.2.1.5. Percentage of trials the optimal policy was selected by preference (Fig. 8). The same regression from Study 1 produced similar findings in Study 2. Participants were more likely to test the optimal policy when it was preference-congruent as opposed to preference-incongruent (MTurk: $\beta = 0.34$, $SE = 0.05$, $p < .001$; Intro Psych.: $\beta = 0.26$, $SE = 0.03$, $p < .001$). Participants were also more likely to test the optimal policy if it was less ambiguous (MTurk: $\beta = 0.38$, $SE = 0.05$, $p < .001$; Intro Psych.: $\beta = 0.44$, $SE = 0.03$, $p < .001$). There was not a significant interaction (MTurk: $\beta = –0.05$, $SE = 0.07$, $p = .477$; Intro Psych.: $\beta = 0.02$, $SE = 0.04$, $p = .707$).

We then tested whether participants were more likely to test the optimal version of a policy if they had neutral preferences about the policy as opposed to having strong preferences (including both preference-congruent and incongruent). We used a mixed effects model predicting the percentage of optimal choices by preference strength (strong vs. neutral), ambiguity, and their interaction. The model included a by-subject random intercept with random slopes for the two main predictors but not the interaction (due to convergence difficulties). Participants were more likely to select the optimal policy if it was less ambiguous (MTurk: $\beta = 0.41$, $SE = 0.05$, $p < .001$; Intro Psych.: $\beta = 0.45$, $SE = 0.02$, $p < .001$). However, there was not a significant difference between strong versus neutral preferences (MTurk: $\beta = 0.00$, $SE = 0.04$, $p = .890$; Intro Psych.: $\beta = 0.03$, $SE = 0.02$, $p = .136$) nor an interaction (MTurk: $\beta = 0.06$, $SE = 0.06$, $p = .364$; Intro Psych.: $\beta = 0.01$, $SE = 0.03$, $p = .820$).

4.2.1.6. Summary of choices during the learning task. The previous analyses replicated the results from Study 1: Participants tested their preferred policy options more frequently, earlier, and were less likely to never test their preferred policy options, compared to their non-preferred policy options, and participants were less successful at using the optimal policy when the optimal policy was preference-incongruent.

However, while we had speculated that perhaps participants would be better at testing policies for which they had neutral preferences, compared to policies for which they had strong preferences (an average of congruent and incongruent), we found few differences.

4.2.2. Judgments of policy efficacy after the learning task

4.2.2.1. Causal functions (Fig. 9). Given that participants rarely selected "no effect" as their final judgment of a policy, we collapsed the responses from three levels into two (correct vs. incorrect) for ease of analysis. We first replicated our finding from Study 1, excluding neutral policies. A mixed effects logistic regression analysis was conducted to test for differences

in the ability to correctly identify which policy option was better for economic output. The main effects and interaction between ambiguity and preference-congruence were included. In the MTurk sample, there was a by-subject random intercept and random slopes for all three predictors. For the Intro. Psych. sample, the random slope for the interaction was dropped due to non-convergence.

Replicating Study 1, MTurk participants were less likely to correctly assess preference-incongruent policies than preference-congruent, ($\beta = –1.68$, $SE = 0.55$, $p = .002$). A similar trend was found for the Intro Psych. sample ($\beta = –2.31$, $SE = 1.42$, $p = .104$), though the effect was not statistically significant and the standard error was considerably larger. Participants were significantly worse at assessing policies with high ambiguity, compared to low ambiguity (MTurk: $\beta = –2.85$, $SE = 0.74$, $p < .001$; Intro. Psych.: $\beta = –21.28$, $SE = 1.89$, $p < .001$). There was no interaction (MTurk: $\beta = 0.72$, $SE = 0.91$, $p = .430$; Intro. Psych.: $\beta = –0.33$, $SE = 2.33$, $p = .888$).

Next, we tested whether participants were better at assessing policies for which they had strong preferences (preference-congruent or incongruent) versus no preferences. A mixed effects logistic regression analysis was conducted with preference-strength (strong vs. weak), ambiguity, and the interaction as predictors, and a by-subject random intercept with random slopes for all three predictors.

There was no significant difference in correctly assessing policies when participants did versus did not have a preference (MTurk: $\beta = 0.44$, $SE = 0.32$, $p = .168$; Intro. Psych.: $\beta = –0.09$, $SE = 0.19$, $p = .621$). Participants were significantly worse at assessing policies with high ambiguity, compared to low ambiguity (MTurk: $\beta = –3.24$, $SE = 0.55$, $p < .001$; Intro. Psych.: $\beta = –3.80$, $SE = 0.39$, $p < .001$). There was no interaction between preference-strength and ambiguity (MTurk: $\beta = –0.70$, $SE = 0.62$, $p = .259$; Intro. Psych.: $\beta = 0.05$, $SE = 0.37$, $p = .898$).

4.2.2.2. Non-causal function (Fig. 10). We first replicated our results from Study 1 demonstrating that participants were more likely to assess their preferred policy as being better, despite no actual difference. To do this, we used the subset of policies for which participants had a preference and for which they failed to correctly assess the policy as non-causal (which was relatively rare). A logistic mixed effects model was run predicting judgment bias (1 = assessing preferred policy as being better; 0 = assessing non-preferred policy as being better) with only a by-subject random intercept to account for repeated measures (each participant had between 0–2 observations). When participants had an initial preference for one policy version over another, after testing it, they were still more likely to view the preferred option as the better policy (MTurk: $M = 0.79$; 95% $CI = 0.68$–0.87; $\beta = 1.34$, $SE = 0.29$, $p < .001$; Intro. Psych.: $M = 0.73$; 95% $CI = 0.65$–0.80; $\beta = 0.99$, $SE = 0.20$, $p < .001$).

We also tested whether participants would be more likely to make accurate judgments of policy efficacy for the non-causal functions if they held neutral preferences versus strong preferences. We conducted a logistic mixed effects regression with preference strength (strong vs. weak) predicting accuracy (correct vs. incorrect) with a by-subject random intercept and a random slope for preference strength. Though participants were a bit more accurate when they held neutral preferences (MTurk: 25.84%; Intro. Psych: 27.50%) than strong preferences

(MTurk: 17.24%; Intro. Psych.: 15.03%), the difference was not significant (MTurk: $\beta = -0.93$, $SE = 2.61$, $p = .721$; Intro. Psych.: $\beta = -1.00$, $SE = 0.86$, $p = .244$).

4.2.3. Function identification

4.2.3.1. Causal functions (Table 3). We first replicated our finding from Study 1, excluding policies for which participants had no preferences. For the MTurk sample, random slopes were included for all three predictors, but for the Intro. Psych. sample, the random slope for the interaction was dropped due to non-convergence. Participants were significantly better at function identification with low ambiguity than high ambiguity in the Intro. Psych sample ($\beta = 0.66$, $SE = 0.23$, $p = .004$); this finding was marginal for the MTurk sample ($\beta = 0.89$, $SE = 0.46$, $p = .051$). Participants were better at function identification for policies that were preference-congruent than incongruent for Intro. Psych. ($\beta = 0.61$, $SE = 0.23$, $p < .001$), though this was marginal for MTurk ($\beta = 0.80$, $SE = 0.44$, $p = .071$). No interaction was found (MTurk: $\beta = -0.81$, $SE = 0.86$, $p = .347$; Intro. Psych: $\beta = 0.14$, $SE = 0.45$, $p = .760$).

We also tested whether participants were more likely to correctly identify functions if they held neutral preferences versus strong preferences. We used a mixed effects logistic regression with the predictors preference strength (strong vs. neutral), ambiguity, and their interaction. The model included a by-subject random intercept with random slopes for preference-congruence and ambiguity but not the interaction due to non-convergence. Participants were significantly better at function identification for low ambiguity functions (MTurk: $\beta = 0.82$, $SE = 0.30$, $p = .006$; Intro. Psych: $\beta = 0.57$, $SE = 0.17$, $p < .001$). There was not a significant effect of preference strength (MTurk: $\beta = -0.26$, $SE = 0.28$, $p = .351$; Intro. Psych: $\beta = -0.28$, $SE = 0.17$, $p = .109$). There was no interaction (MTurk: $\beta = 0.18$, $SE = 0.55$, $p = .742$; Intro. Psych: $\beta = -0.25$, $SE = 0.32$, $p = .436$).

*4.2.3.2. Non-causal functions (*Table 4*).* We conducted a mixed effects logistic regression to test for differences in non-causal function identification by preference-strength (strong vs. weak). A by-subject random intercept with a random slope was used. Though the accuracy of function identification was a bit higher with neutral preferences than strong preferences, the difference was not significant in the MTurk sample ($\beta = 3.59$, $SE = 6.76$, $p = .595$). However, participants in the Intro. Psych sample were more likely to correctly identify the non-causal functions if they did not have preferences ($\beta = 9.63$, $SE = 1.67$, $p < .001$).

4.3. Study 2A and 2B discussion

Study 2 largely replicated the findings in Study 1. In addition, Study 2 found that when participants had neutral preferences, their performance was in the middle between preference-congruence and preference-incongruence such that there was no difference in performance between having strong and weak preferences in most all cases. Stated another way, the benefits of preference congruence (when the participant's preference happens to be right) and the costs of preference incongruence (when the participant's preference happens to be wrong) roughly cancel out. This means that likely the "bias" of preference or prior beliefs is

working in two directions—for preferred policies (relative to neutral) and against non-preferred policies (relative to neutral).

There were some hints that the neutral condition might not be right in the middle, or might actually flip for low versus high ambiguity functions, but these were not statistically significant. In sum, this study reconfirms that preferences have a strong influence on causal learning and judgments; however, it does not provide evidence that having preferences, on the whole, leads to better or worse learning and judgments, compared to not having preferences.

5. Study 3: Whether knowledge of potential causal functions reduces motivated reasoning

One of the central challenges participants faced in Studies 1 and 2 is that they did not know in advance about the possible functions for how the policies worked. For example, if a participant assumed that the policies worked immediately, they might make a change to one policy on one trial, and then make a change to another policy on the subsequent trial, and because it actually takes a number of trials for the policies to work, their causal attributions could be wrong. For another example, a participant might not even consider the possibility that a policy could have short-term costs but long-term benefits, and upon noticing a short-term cost, they might switch away from that policy without investigating whether there are long-term benefits.

On the one hand, in many real-world situations, decision-makers do not know the possible functions, or might only have rough guesses about the length of time it might take for a policy to produce its full impact, or whether it is possible for a policy to have different short versus long-term influences. On the other hand, in some situations, more informed decision-makers might have hypotheses about possible functional forms (e.g., see the quotes at the beginning of the introduction).

The goal of Study 3 was to investigate whether being more informed about the potential types of influences (function exposure) would improve learning, which would appear as the main effect of function exposure. Furthermore, we hypothesized that if participants are exposed to the possible functional forms in advance, it might reduce the biases seen due to preference, which would appear as an interaction between preference (congruent vs. incongruent) and function exposure.

Previous studies using the "melioration" paradigm have tested a couple of ways to improve performance on the task, with various successes. It has been found that giving participants a perceptual cue that corresponds with the underlying state of the payoff function (how many times the optimal choice has been chosen in the past 10 trials) can improve learning (Gureckis & Love, 2009; Herrnstein, Loewenstein, Prelec, & Vaughan, 1993; Otto, Gureckis, Markman, & Love, 2009; Stillwell & Tunney, 2009). However, this approach would have been very confusing with six causes instead of just one, and furthermore, we wanted to test whether a more explicit form of knowledge of the possible functions could matter. Unlike the previous studies on melioration which focused on the percentage of optimal choices, we also studied

participants' explicit beliefs about which policy option was better and their beliefs about the functional form of the payoff. Herrnstein et al. (1993) found that giving participants explicit instructions about how to maximize earnings improved performance. However, these instructions did not clearly state that the different options could have different short-term versus long-term consequences. In the current study, we explicitly told some participants about the possibility of such temporal tradeoffs.

5.1. Method

5.1.1. Participants

One-hundred participants were recruited via MTurk. They were paid $5.50 for participation (which amounted to approximately $8–10/h) with an opportunity to be awarded up to $3.00 in bonuses contingent upon performance.

5.1.2. Design

Study 3 was very similar to Study 1 with the following changes. First, half of the participants were exposed to the possible functional forms of the policies before starting the task, and the other half were not (like in Studies 1 and 2). Second, similar to Study 1, Study 3 focused on learning in the context of strong preferences, so only policies with strong prior preferences were selected. However, if a participant did not have six policies with strong prior preferences, the computer would choose the "next most extreme" to be included in the task. In these cases, the policies that did not meet our criteria to be categorized as "strong prior preferences" would be omitted from analysis (but not the participant altogether). Third, as an improvement to Study 1, we counterbalanced the causal functions such that one of the low ambiguity functions was preference-congruent and one was preference-incongruent and the same for the two high ambiguity functions.

5.1.3. Function exposure task

In the function exposure task, participants read the following statement:

"In the following task, you will pretend to be the elected leader of a large industrialized country, and you will be responsible for making important decisions about economic policies. But before doing so, **we want you to reflect on the possible ways that your changes to economic policies might influence the economy**.
A change to a policy might:

- Have no influence on the economy.
- Have a positive or negative influence, but it might also take some amount of time for these positive or negative influences to appear.
- Initially have a positive influence but eventually have a negative influence or vice versa.
- Have a temporary positive or negative influence, but no long-term influence.

Thinking about the possible ways that your policy changes might influence the economy will help you to determine which policies are best in order to maximize the economy's output."

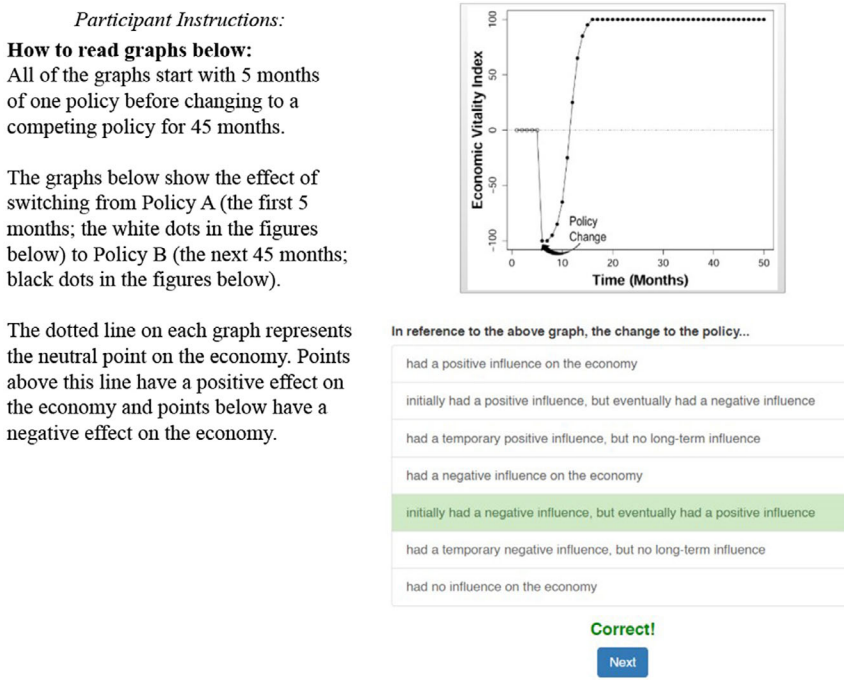

Fig 11. Function exposure comprehension test. Participants had to correctly choose the text that best described the function to verify that they could understand the graphs in the function exposure condition.

Then participants were shown graphs of the seven functions (five that were present in the learning task, and two lures; see Fig. 3), and for each graph, they had to match the function to the text describing the function to verify that they understood the different types of functions before moving on (Fig. 11).

5.2. Results

In Study 3, we only investigated learning in the presence of strong prior preferences. However, as explained in the methods, it was possible that for some of the policies, participants would hold moderate views. In the few cases in which participants did not have strong prior preferences for certain functions, these were omitted from the analysis. Sixty-eight participants had strong preferences for all six policies. Three participants had five preferred policies and one neutral policy. Seven participants had four preferred policies and two neutral policies. Seven participants had three or fewer policies with strong preferences, and these participants were dropped entirely from the study.

In addition, 14 participants were removed from analyses for making fewer than two policy changes throughout the entire learning task, and one was removed for not following directions. In all, 78 participants were included in the analyses.

5.2.1. Choices in the learning task

5.2.1.1. Switching multiple policies to the preferred option at the beginning of learning (Table 1, Figs. 4 and 5). Similar to the prior studies, on the first trial, participants tended to switch multiple policies from the non-preferred option to the preferred option (Table 1). After the first trial, the majority of changes to policies were controlled, not confounded (Table 1 and Fig. 4).

To test whether participants switched non-preferred policies to preferred earlier than the reverse (Fig. 5), a gamma mixed effects regression was conducted predicting time until testing by the interaction of function exposure condition and policy preference at the start, with a by-subject random intercept and a random slope of preference at the start. Replicating the prior studies, participants switched non-preferred policies to preferred earlier than they switched preferred policies to non-preferred ($\beta = –0.30$, $SE = 0.05$, $p < .001$). There was no effect of function exposure ($\beta = 0.03$, $SE = 0.05$, $p = .584$), nor an interaction ($\beta = 0.08$, $SE = 0.09$, $p = .339$).

5.2.1.2. Changing a policy and holding others stable for periods of time (Table 2). The findings for Study 3 are similar to the prior studies. After making a change to a policy, participants tended to hold the system stable for only one or two trials before making a subsequent change. This meant that they had relatively good evidence about the short-term impacts of the policies, but that they did not produce good evidence about the long-term impacts of the policies. Patterns of testing were similar across function exposure conditions, so these results were collapsed within Table 2.

5.2.1.3. Never testing bias by preference (Fig. 6). On average, across all participants and all policies, 10% were never changed (function exposure condition: 8%; no exposure condition: 12%). We used a logistic mixed effects model predicting the likelihood that a policy was tested by preference at the start (preferred vs. non-preferred policy), function exposure, and their interaction. The model included a by-subject random intercept with a random slope for preference at the start. No differences in the likelihood of not testing a policy were found for preference at start ($\beta = 3.86$, $SE = 3.51$, $p = .271$), function exposure ($\beta = –0.15$, $SE = 1.75$, $p = .930$), nor their interaction ($\beta = 0.94$, $SE = 3.19$, $p = .768$). When examining the figure, the preferred at the start policies are trending in the predicted direction (same with function exposure), but these differences are not significant.

5.2.1.4. Percentage of trials during which the preferred option was selected (Fig. 7). For the causal functions, participants tested their preferred version of the policies more than their non-preferred version, $M = 65\%$; $SD = 18\%$, $t(77) = 7.36$, $p < .001$, $d = 0.83$. There was no difference between those who received the function exposure ($M = 64\%$; $SD = 16\%$) and those who did not ($M = 65\%$; $SD = 19\%$), $t(75.84) = 0.18$, $p = .86$, $d = 0.04$.

For the non-causal functions, participants also tested their preferred version more frequently than their non-preferred version (M = 68%; SD = 24%), $t(76) = 6.63$, $p < .001$, $d = 0.76$. There were no differences between the participants who received the function exposure ($M = 66\%$; $SD = 24\%$) or not ($M = 71\%$; $SD = 24\%$), $t(72.25) = 0.90$, $p = .37$, $d = 0.20$.

5.2.1.5. Percentage of trials the optimal policy was selected by preference (Fig. 8). Similar to the prior studies, we used a random effects regression to predict the percentage of trials during which the policy was set to the optimal choice by congruence, ambiguity, condition, and their interactions. The model had a by-subject random intercept with random slopes for preference-congruence and ambiguity—it did not have a random slope for the interaction between these two as the model could not converge given that there was only one observation per cell.

Confirming findings from Studies 1 and 2, participants were more likely to select the optimal policy for preference-congruent as opposed to preference-incongruent policies ($\beta = 0.30$, $SE = 0.04$, $p < .001$) and for less ambiguous policies ($\beta = 0.42$, $SE = 0.04$, $p < .001$). We did not find a significant effect of function exposure ($\beta = –0.01$, $SE = 0.03$, $p = .708$). None of the interactions were significant.

5.2.2. Judgments of policy efficacy after the learning task

5.2.2.1. Causal functions (Fig. 9). A near-identical approach was taken here as that of Study 2A and 2B, the only exception being that function exposure condition (between-subjects), and its interactions with the other predictors were included as predictors. The model included a by-subject random intercept but no random slopes.10

First, participants were less likely to correctly assess policies if they were preference-incongruent than congruent ($\beta = –1.66$, $SE = 0.34$, $p < .001$). Second, participants were significantly worse at assessing policies with high ambiguity, compared to low ambiguity ($\beta = –3.01$, $SE = 0.34$, $p < .001$). Third, and most relevant to Study 3, there was no effect of function exposure ($\beta = –0.30$, $SE = 0.34$, $p = .383$); participants were about equally accurate in the function exposure condition ($M = 45.77\%$) as in the no-exposure condition ($M = 50.63\%$). There were also no significant two or three-way interactions.

5.2.2.2. Non-causal functions (Fig. 10). We first replicated the finding that participants were more likely to assess their preferred policy as being better, despite there being no difference. We used the same approach as in Study 2, and Study 3 only used the no-function-exposure group for comparability. Replicating prior results, we found that when participants had an a priori preference, after testing it, they were still more likely to view it as the better policy ($M = 0.73$; 95% $CI = 0.61$–0.83; $\beta = 0.99$, $SE = 0.28$, $p < .001$).

We next tested whether participants who were in the function exposure condition performed better on this task, compared to those who were not. To test for this difference, we conducted a mixed effects logistic regression with function exposure condition predicting accuracy (correct vs. incorrect) with a by-subject random intercept. The mean accuracy in the function exposure condition (22.06%) and the no exposure condition (24.10%) were similar, and the effect of condition was not significant, $\beta = 0.15$, $SE = 0.51$, $p = .775$.

5.2.3. Function identification

5.2.3.1. Causal functions (Table 3). A mixed effects logistic regression analysis was used to predict the ability to correctly choose the graph that represented the function of each policy from preference-congruence, ambiguity, and function exposure condition. A by-subject

random intercept was used with no random slopes. There were positive effects of preference-congruence ($\beta = 0.68$, $SE = 0.20$, $p = .040$) and lower ambiguity ($\beta = 1.09$, $SE = 0.33$, $p = .001$). However, there was no main effect of function exposure ($\beta = 0.52$, $SE = 0.34$, $p = .120$), nor were there any significant two or three-way interactions.

5.2.3.2. Non-causal functions (Table 4). A mixed effects logistic regression analysis was used to predict the ability to correctly identify that the two non-causal policies per participant were non-causal based on condition. The model included a by-subject random intercept. Being exposed to the functions prior to learning did not improve function identification ($\beta = 0.40$, $SE = 0.61$, $p = .513$).

5.3. Study 3 discussion

Study 3 replicated many of the findings from the prior studies. The added intervention of being exposed to the possible functional forms for the policies was largely ineffective at improving performance

6. General discussion

In three studies, we evaluated how successfully participants tested and utilized policies in an explore-exploit task and the impact of having political preferences on this testing process. At a general level, we found a number of specific ways in which having preferences impact how people go about testing and utilizing policies. Here, we first revisit the questions posed in the introduction.

6.1. Summary of results

With regards to characterizing participants' testing habits and learning curves, we found three general patterns. First, participants make many changes early on and then fewer over time, which makes sense from an explore-exploit perspective. Second, participants often did not hold the system stable for very long after making a change to a policy, which would make it hard to reveal the long-term implications of the high ambiguity functions. Third, participants learned to exploit the low ambiguity policies but had a harder time exploiting the high ambiguity policies.

With regards to the second set of questions about the influence of preferences, we found evidence that people's testing behavior and learning outcomes were greatly influenced by their a priori preferences for some policy options over others (e.g., increasing border security funding vs. decreasing it), which could be viewed as a type of motivated reasoning. We identified four specific biased habits during testing; the first has to do with confounded versus controlled testing, and the remaining three can all be viewed as different manifestations of positive testing (e.g., Klayman & Ha, 1987). First, at the very beginning of testing, participants tended to switch multiple policies from the non-preferred state to the preferred state, which means that these changes tended to be confounded rather than controlled. Second, in instances in which participants did not test a policy at all (did not make a change to the policy), the policy tended to already be set to the preferred policy. Third, participants tended to

test the preferred option of the policy more overall than the non-preferred option. Fourth, all of these habits led participants to use the optimal policy more when the optimal policy was congruent with their preferences and less when it was incongruent.

With regards to the third question about the impact of political preferences on judgments of the policies after learning, we found the following. First, participants were more likely to correctly assess the policies (to correctly determine which version of the policy is better) when they were preference-congruent (when the participant preferred the option that happened to be better). Second, participants' ability to identify the underlying function was sometimes but not always influenced by their preferences. Furthermore, in general, the ability to identify the functional form was low especially for the highly ambiguous functions.

Though the above findings of the role of preferences can be viewed as types of positive testing and confounded testing, we believe that an important contribution of this research is to reveal the specific ways that biased causal testing can play out. It is easy to imagine that similar testing patterns could play out in real-world situations. For example, when a politician, business executive, or decision-maker more broadly initially assumes office, they may try to make multiple changes as quickly as possible based on their prior preferences or beliefs about various policies; however, making confounded changes would make it harder to assess the influence of each individual change. If a policy is already set in a preferred state before assuming office, they may be less willing to test the non-preferred state, or across their time in office, they may only rarely test less preferred policies and more frequently test preferred policies. Collectively, these habits would make it harder to learn which policy options are actually better.

In the introduction, we raised three other general questions about the ambiguity of the functions, the impact of having strong versus neutral preferences, and having prior knowledge about potential functional forms. These questions are addressed in the following sections.

6.2. Ambiguity

As expected, participants were much worse at learning the high ambiguity policies than the low ambiguity ones. We had hypothesized that, in addition, the motivated reasoning effect would be magnified for the high ambiguity policies because the ambiguity could license interpreting these policies in the ways that participants preferred; however, we did not find evidence for this hypothesis. We have a couple of speculations about why.

One possibility is that the low ambiguity function was fairly ambiguous. Indeed, the low ambiguity functions themselves took multiple trials to reach their full influence and there was also noise that made all the functions harder to detect. Overall, the challenges involved in learning the "low ambiguity" functions, such as those already mentioned as well as the fact that six policies need to be learned about simultaneously, could have left an opportunity for considerable bias due to preference.

Another possibility (not mutually exclusive with the first) is that when learning about the high ambiguity policies, participants did not notice the ambiguity (the opposing short-term and long-term influences) at all, and instead, only noticed the short-term influence. In Fig. 1, the "high ambiguity" Functions 3 and 4 produce strong influences on the very first trial that

they are implemented. In fact, the immediate influence for the "high ambiguity" functions is stronger than the immediate influence for the "low ambiguity" Functions 1 and 2, which take a couple of trials to reach their full strength. It is possible that most participants therefore viewed Function 3 as fairly strong unambiguous evidence for a negative effect and Function 4 as fairly strong unambiguous evidence for a positive effect, when in reality, their long-term influences are the opposite. In fact, we found that participants did not hold the system stable for very long, which would have made it much harder to learn the long-term effects of the policies than the short-term effects. Likewise, Sims et al. (2013) argued that when learning about policies with different short versus long-term influences, the data that participants experience is not sufficient for them to learn the true functional form, and learning the short-term relation is rational. Stated another way, even if these policies are ambiguous from the perspective of the experimenter, perhaps they were not ambiguous from the perspective of the participant.

Under this possibility, the participants' subjective experiences and interpretations would have been quite similar for the low and high ambiguity policies. This fits with the very poor performance for the high ambiguity policies (Figs. 8 and 9) because the poor performance of learning the long-term influence can be reinterpreted as very *good* performance for learning the short-term influence, just like the good performance of learning the low ambiguity functions.

What is clear is that people have considerable difficulty learning about the high ambiguity policies for which the short-term and long-term influences contradict each other, which is consistent with the prior findings using similar payoff functions (Gureckis & Love, 2009, and citations therein). This is especially problematic given that many economic policies (e.g., President Trump's justification for a trade war with China, providing universal early education, free college tuition) are believed to involve a trade-off between the short- versus long-term.

Even though in this paper, we did not find support for biased reasoning increasing in response to greater ambiguity, we suspect that such a pattern might be found in other situations. For example, it might be found when comparing the current policies to a policy that is truly unambiguous (it has an immediate and strong influence). Alternatively, it might be found when comparing a learning situation that involves very little noise (low ambiguity) to a learning situation with considerable noise. Or, if the long-term benefit of the ambiguous policies came earlier, perhaps participants would become more aware of the temporal trade-off and the ambiguity therein. In sum, there are many different ways in which ambiguity can arise, and other sorts of ambiguity could potentially moderate the motivated reasoning effect.

6.3. Open-mindedness and neutral preferences

In Study 2, we had speculated that perhaps having neutral preferences would lead to better learning and more accurate causal judgments, compared to having strong preferences. The hypothesis was that when a learner has neutral preferences, they might be less biased, which could lead to more accurate learning. In contrast, when a learner has strong preferences, sometimes those preferences would be "congruent" (their preferred policy option happens to

be better), but sometimes they would be "incongruent" (their preferred policy option happens to be worse). We speculated that perhaps the costs of preference-incongruence, compared to neutral preferences, would be larger than the benefits of preference-congruence, compared to neutral preferences. The reason was that if participants avoid testing their non-preferred options, they would learn little about them, potentially leading to very poor learning. In fact, avoiding testing non-preferred options could hurt both preference-incongruent as well as preference-congruent policies because if a preferred option is repeatedly utilized, a learner does not get to test the comparison between the preferred versus the non-preferred option, which is critical for determining which policy option is better.11

Despite some apparent asymmetries in the means between preference-congruent, neutral, and preference-incongruent policies, no asymmetries were significant. On the one hand, this could be thought of as a fortunate finding; even if people are biased, being biased on average in this task did not lead to worse causal learning. On the other hand, in the current study, randomization was used such that on average there was the same number of preference-congruent and preference-incongruent policies. However, in the real world, it is entirely possible that a population in general, or one sub-population due to polarization, might, in general, have more preference-incongruent views than congruent (i.e., they might tend to prefer policies that are actually worse for the economy). If so, holding more neutral views initially could still be beneficial.

6.4. Prior knowledge and expertise

In Study 3, we tested whether participants would perform better at learning and when making causal assessments if they were given initial instructions about possible types of functional forms of the policies. Most importantly, we wanted them to consider the possibility that a policy might have no influence on the economy at all or that a policy might have a short-term benefit and a long-term consequence or vice versa since participants had so much difficulty learning about all of these policies. In a sense, having some more knowledge about potential functional forms could be viewed as a very light manipulation of expertise; true experts would presumably have more specific views about the timeframes within which a policy could play out.

Despite this hypothesis, there was little evidence that this manipulation made a difference. It did not seem to help them identify when a policy was non-causal (Table 4). It also did not improve the accuracy of assessing high ambiguity functions (Table 3). Participants were about 15% more accurate in the function identification for the low ambiguity functions (Table 3); we did not test whether this particular difference was significant only for low ambiguity functions, but it was not significant for both low and high ambiguity functions.

There are a couple of potential explanations for the failure of the intervention. First, perhaps the task is just so hard for the neutral policies and the high-ambiguous policies that the instructions were insufficient to make a difference. Second, it is possible that upon starting the task participants did not think back to the instructions. Third, though we think that this is fairly unlikely, perhaps even though participants passed the questions requiring some amount of understanding the instructions, they did not really understand all the functions.

Other research has found that even though people can use prior knowledge about aspects like delay and carryover effects to adapt their causal testing strategies, they have difficulty using other knowledge such as wave-like changes over time (Rottman, 2016). Thus, it appears that adapting testing strategies based on prior knowledge of functional forms can be very challenging. Furthermore, other studies on the melioration paradigm have found that giving explicit instructions can help, but the largest benefit came when essentially telling participants which option is better in the long run (Herrnstein et al., 1993), a very heavy-handed approach, and one not available to real-world conditions in which the truth is unknown. The current research suggests that even with some forewarning, people still have considerable difficulty learning about policies that have different short- and long-term influences, but perhaps other forms of instruction or expertise could help.

6.5. Incentives and taking the task seriously

One important question is the extent to which participants thought that their preferences and beliefs prior to the learning task could actually help them perform well during the learning task. For example, consider a participant who fervently believed that certain policies help the economy and others hurt, and imagine that they believed that the study was programmed to reflect how the actual economy works. In this case, it would be entirely rational to use the prior preferences and beliefs to guide learning. For a participant like this, the current study would be a good simulation of how motivated reasoning could play out in more real-world high-stakes situations.

Alternatively, consider another participant who believed that the study was just a game and that their real-world beliefs and preferences were irrelevant to performing well on the study. If so, then presumably they would be able to hold their preferences at bay and try to learn in the most rational way possible in order to maximize their bonus rewards for the task; accuracy was incentivized with bonuses in all studies except 2B. In fact, accuracy incentives have been found to reduce and sometimes eliminate the partisan bias effect when assessing the current state of the economy (Bullock et al., 2013; Prior, Sood, & Khanna, 2015). Yet, in our study, we still observed strong and reliable effects of prior preferences when learning about economic policies, which suggests that in some cases, people do not just ignore their preferences even when financially motivated to do so and even when dealing with an artificial study about a hypothetical society in the future. This evidence speaks to the powerful biasing effect of motivated reasoning.

It is entirely possible that the participants in these studies included a mixture of both of these sorts of beliefs or primarily one type more than the other. However, we believe that the results of the current study are important regardless of the composition of the participants. In the first case, the study is a fairly good simulation of more real-world learning. In the second, it shows the power of preferences even when participants believe them to be irrelevant and are incentivized not to use them. Furthermore, this research revealed not just that preferences bias learning and judgment but specific ways in which they bias learning and judgment.

6.6. Preferences versus beliefs and motivated reasoning

In this paper, we have extensively used the term "preference," and at the beginning clarified that we would use "preference" to also include "beliefs." We did not try to distinguish beliefs from preferences because we felt that they would often be correlated and would likely be hard to distinguish empirically. Thus, it is possible that some of the motivated reasoning could be due to participants importing their actual beliefs about economic policies and thinking that using such beliefs would help them perform better on this task if this task is an accurate simulation of the actual economy.

Though this changes the nature of the motivation, we still think that it is important, perhaps even more important, to understand how prior beliefs affect learning about policies. Future research could try to study how people learn about and test policies for which they prefer one option even if they believe it to be harmful to the economy (e.g., perhaps it has other benefits such as fairness).

6.7. Conclusion

The current research integrates paradigms from motivated reasoning and causal reasoning/reinforcement learning to understand how prior preferences affect how people go about testing the causal impact of policies and how people draw conclusions about policies. We found strong impacts of participants' prior preferences, even in this artificial task and even despite accuracy incentives. Similar processes may occur in real-world situations when one's preferences are even more likely to determine one's willingness to implement certain policies over others.

Acknowledgment

All data and analysis scripts are posted at https://github.com/caddickzac/Motivated-Reasoning-in-an-Explore-Exploit-Task.

Open Research Badges

This article has earned Open Data badge. Data are available at https://github.com/caddickzac/Motivated-Reasoning-in-an-Explore-Exploit-Task.

Notes

1 The terms "preference," "belief," and "attitude" are often used interchangeably in the literature. For simplicity, we use the word "preference" and discuss this in more detail in the section on Relations Between Motivated Reasoning and Learning from Experience.
2 Studies in the melioration literature typically use a flat payoff distribution over the prior 10 trials, which results in straight lines instead of curved lines in Fig. 1. We chose a

slightly different payoff function in order to make the returns curved, similar to Functions 1 and 2. However, the general shape of the function is quite similar.

3 During Trials 2 through 140, unlike on Trial 1, there was more of a balance between switching policies toward preferred vs. non-preferred options (Table 1). In fact, among the controlled changes there are somewhat more changes toward the non-preferred option; this likely represents an attempt to learn about each policy individually, and since so many participants switched policies to the preferred option on Trial 1, during the remainder of trials, as they accurately learn about which option is better they would necessarily need to switch more towards the non-preferred option since the optimal option was randomly assigned as preferred or non-preferred.

4 The dependent variable for this analysis (and the analogous gamma distribution analyses in Studies 2 and 3) was transformed to z-scores with a minimum value of 1, to improve model convergence. If a participant never tested a policy at any point during the learning task, that particular policy for that participant was omitted from analysis.

5 We initially conducted a mixed effects logistic regression at the policy level for each of the six policies for each participant, but we ran into convergence issues. This is likely due that fact that we are attempting to detect differences in rare events where large individual differences were present. In response, we simplified the approach and analyzed the data at a higher level.

6 Note, these means are higher than the overall average (7.72%) and the averages in Fig. 6 because the inferential statistics analyze whether a participant failed to test any of the policies initially set to preferred or non-preferred, whereas in Figure 6 we report the likelihood that an individual policy was not tested.

7 Throughout the results, when we ran into difficulties with convergence, we followed Barr et al.'s (2013) advice for the order of simplifying the model from the maximal model. Dropping the correlation between random slopes is one of the first recommended steps. A number of other models in this manuscript also drop the correlation parameter and are not specifically identified for concision.

8 Participants could have up to four preference-congruent policies or as few as zero. This means there were repeated measures for some users (e.g., two preference-congruent with the high ambiguity functions), only between group measures for some, and an absence of measurement for other participants (e.g., no preference-congruence with the high ambiguity functions).

9 We classified answers as correct versus incorrect and conducted logistic regressions instead of multinomial logistic regressions because the number of possible hypotheses is very large for multinomial logistic regressions, and we cared most about whether participants got the answer correct or not.

10 Though preference congruence and ambiguity were within-subjects, there was only one observation per person per cell, and this was the maximal model that would converge here and for other similar models in Study 3.

11 In theory, these factors could play out differently for different measures. For example, if a participant blindly uses a preferred option during learning and rarely, if ever, tests the non-preferred option, then they would do very well at selecting the optimal choice

during learning, but when identifying the functional form, they could do very poorly if they barely learned anything about the policy. For neutral policies, the performance on both tasks presumably depends largely on the task difficulty, which could affect the relative performance, compared to preference-congruent and incongruent policies.

References

Alloy, L. B., & Tabachnik, N. (1984). Assessment of covariation by humans and animals: The joint influence of prior expectations and current situational information. *Psychological Review*, *91*(1), 112–149.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), 255–278.

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (2005). The Iowa Gambling Task and the somatic marker hypothesis: Some questions and answers. *Trends in Cognitive Sciences*, *9*, 159–162. https://doi.org/10.1016/j.tics.2005.02.002

Brehmer, B. (1971). Subjects' ability to use functional rules. *Psychonomic Science*, *24*, 259–260.

Brehmer, B. (1974). Hypotheses about relations between scaled variables in the learning of probabilistic inference tasks. *Organizational Behavior and Human Decision Processes*, *11*, 1–27.

Buehner, M. J., & McGregor, S. (2006). Temporal delays can facilitate causal attribution: Towards a general timeframe bias in causal induction. *Thinking & Reasoning*, *12*(4), 353–378.

Bullock, J. G., Gerber, A. S., Hill, S. J., & Huber, G. A. (2013). Partisan bias in factual beliefs about politics (No. w19080). *National Bureau of Economic Research*.

Bullock, J. G., Gerber, A. S., Hill, S. J., & Huber, G. A. (2015). Partisan bias in factual beliefs about politics. *Quarterly Journal of Political Science*, *10*(4), 519–578.

Busemeyer, J. R., Byun, E., DeLosh, E. L., & McDaniel, M. A. (1997). Learning functional relations based on experience with input- output pairs by humans and artificial neural networks. In K. Lam-berts & D. Shanks (Eds.), *Concepts and categories* (pp. 405–437). Cambridge: MIT Press.

Campbell, T. H., & Kay, A. C. (2014). Solution aversion: On the relation between ideology and motivated disbelief. *Journal of Personality and Social Psychology*, *107*(5), 809–824.

Coenen, A., Rehder, B., & Gureckis, T. M. (2015). Strategies to intervene on causal systems are adaptively selected. *Cognitive Psychology*, *79*, 102–133.

Coenen, A., Ruggeri, A., Bramley, N. R., & Gureckis, T. M. (2019). Testing one or multiple: How beliefs about sparsity affect causal experimentation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(11), 1923.

Derringer, C., & Rottman, B. M. (2018). How people learn about causal influence when there are many possible causes: A model based on informative transitions. *Cognitive Psychology*, *102*, 41–71.

Ditto, P. H., Scepansky, J. A., Munro, G. D., Apanovitch, A. M., & Lockhart, L. K. (1998). Motivated sensitivity to preference-inconsistent information. *Journal of Personality and Social Psychology*, *75*(1), 53.

Factbase (2018). Interview: WABC Radio's Bernie McGuirk and Sid Rosenberg interview Donald Trump on Bernie & Sid–April 6, 2018. Available at: https://factba.se/transcript/donald-trump-interview-bernie-sid-show-wabc-april-6-2018. Accessed January 14, 2020.

Fugelsang, J. A., & Thompson, V. (2000). Strategy selection in causal reasoning: When beliefs and covariation collide. *Canadian Journal of Experimental Psychology*, *54*(1), 15–32.

Fugelsang, J. A., & Thompson, V. A. (2003). A dual-process model of belief and evidence interactions in causal reasoning. *Memory & Cognition*, *31*(5), 800–815.

Gershman, S. J. (2018). Deconstructing the human algorithms for exploration. *Cognition*, *173*, 34–42.

Gershman, S. J. (2019). How to never be wrong. *Psychonomic Bulletin & Review*, *26*(1), 13–28.

Goedert, K. M., Ellefson, M. R., & Rehder, B. (2014). Differences in the weighting and choice of evidence for plausible versus implausible causes. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *40*(3), 683–702.

Gureckis, T. M., & Love, B. C. (2009). Short-term gains, long-term pains: How cues about state aid learning in dynamic environments. *Cognition*, *113*(3), 293–313.

Hagmayer, Y., & Waldmann, M. R. (2002). How temporal assumptions influence causal judgments. *Memory & Cognition*, *30*(7), 1128–1137.

Hart, P. S., & Nisbet, E. C. (2011). Boomerang effects in science communication: How motivated reasoning and identity cues amplify opinion polarization about climate mitigation policies. *Communication Research*, *39*(6), 701–723.

Herrnstein, R. J., Loewenstein, G. F., Prelec, D., & Vaughan Jr., W. (1993). Utility maximization and melioration: Internalities in individual choice. *Journal of Behavioral Decision Making*, *6*, 149–185.

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

Koh, K., & Meyer, D. (1991). Function learning: Induction of continuous stimulus-response relations. *Journal of Experimental Psychology: Learning Memory, and Cognition*, *17*(5), 811–836.

Klaczynski, P. A. (1997). Bias in adolescents' everyday reasoning and its relationship with intellectual ability, personal theories, and self-serving motivation. *Developmental Psychology*, *33*(2), 273–283.

Klayman, J., & Ha, Y. W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, *94*(2), 211.

Kahan, D. M., Braman, D., Cohen, G. L., Gastil, J., & Slovic, P. (2010). Who fears the HPV vaccine, who doesn't, and why? An experimental study of the mechanisms of cultural cognition. *Law and Human Behavior*, *34*(6), 501–516.

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

Klayman, J. (1995). Varieties of confirmation bias. *Psychology of learning and motivation*, *32*, 385–418.

Kruglanski, A. W. (2004). *The psychology of closed-mindedness*. Hove: Psychology Press.

Jern, A., Chang, K. K., & Kemp, C. (2014). Belief polarization is not always irrational. *Psychological Review*, *121*(2), 206–224.

Kruglanski, A. W., Jasko, K., & Friston, K. (2020). All thinking is 'wishful'thinking. *Trends in Cognitive Sciences*, *24*(6), 413–424.

Kunda, Z. (1987). Motivated inference: Self-serving generation and evaluation of causal theories. *Journal of Personality and Social Psychology*, *53*(4), 636–647.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*, 480.

Lucas, C. G., Bridgers, S., Griffiths, T. L., & Gopnik, A. (2014). When children are better (or at least more open-minded) learners than adults: Developmental differences in learning the forms of causal relationships. *Cognition*, *131*(2), 284–299.

Lucas, C. G., Griffiths, T. L., Williams, J. J., & Kalish, M. L. (2015). A rational model of function learning. *Psychonomic Bulletin & Review*, *22*, 1193–1215.

Luhmann, C. C., & Ahn, W. (2007). BUCKLE: A model of unobserved cause learning. *Psychological Review*, *114*(3), 657–677.

Marks, J., Copland, E., Loh, E., Sunstein, C. R., & Sharot, T. (2018). *Epistemic spillovers: Learning others' political views reduces the ability to assess and use their expertise in nonpolitical domains* (Working Paper No. 18–22). Harvard Public Law.

Marsh, J. K., & Ahn, W. (2009). Spontaneous assimilation of continuous values and temporal information in causal induction. *Journal of Experimental Psychology: Learning Memory and Cognition*, *35*(2), 334–352.

National Academy of Sciences (2013). *Next generation science standards: For states, by states*. Washington, DC: National Academies Press.

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

Nisbett, R. E., & Ross, L. (1980). Human inference: Strategies and shortcomings of social judgment.

Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, *32*(2), 303–330.

Otto, A. R., Gureckis, T. M., Markman, A. B., & Love, B. C. (2009). Navigating through abstract decision spaces: Evaluating the role of state generalization in a dynamic decision-making task. *Psychonomic Bulletin & Review*, *16*, 957–963.

Paharia, N., Vohs, K. D., & Deshpandé, R. (2013). Sweatshop labor is wrong unless the shoes are cute: Cognition can both help and hurt moral motivated reasoning. *Organizational Behavior and Human Decision Processes*, *121*(1), 81–88.

Prior, M., Sood, G., & Khanna, K. (2015). You cannot be serious: The impact of accuracy incentives on partisan bias in reports of economic perceptions. *Quarterly Journal of Political Science*, *10*(4), 489–518.

Rottman, B. M. (2016). Searching for the best cause: Roles of mechanism preferences, autocorrelation, and exploitation. *Journal of Experimental Psychology: Learning Memory and Cognition*, *42*(8), 1233–1256.

Schulz, E., Konstantinidis, E., & Speekenbrink, M. (2017). Putting bandits into context: How function learning supports decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(6), 927–943. https://doi.org/10.1037/xlm0000463

Schulz, E., Tenenbaum, J. B., Duvenaud, D., Speekenbrink, M., & Gershman, S. J. (2017). Compositional inductive biases in function learning. *Cognitive Psychology*, *99*, 44–79.

Schulz, E., Konstantinidis, E., & Speekenbrink, M. (2018). Putting bandits into context: How function learning supports decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(6), 927.

Shepardson, D. (2018). Senator McConnell says tariffs hurt, urges trade progress. Available at: https://www.reuters.com/article/us-usa-senate-mcconnell-trade/senator-mcconnell-says-tariffs-hurt-urges-trade-progress-idUSKCN1MR2UZ. Accessed January 28, 2020.

Sims, C. R., Neth, H., Jacobs, R. A., & Gray, W. D. (2013). Melioration as rational choice: Sequential decision making in uncertain environments. *Psychological Review*, *120*(1), 139.

Speekenbrink, M., & Konstantinidis, E. (2015). Uncertainty and exploration in a restless bandit problem. *Topics in Cognitive Science*, *7*(2), 351–367.

Speekenbrink, M., & Shanks, D. R. (2010). Learning in a changing environment. *Journal of Experimental Psychology: General*, *139*(2), 266.

Spellman, B. A. (1996). Conditionalizing causality. *Psychology of Learning and Motivation*, *34*, 167–206.

Steyvers, M., Lee, M. D., & Wagenmakers, E. -J. (2009). A Bayesian analysis of human decision-making on bandit problems. *Journal of Mathematical Psychology*, *53*, 168–179. http://doi.org/10.1016/j.jmp.2008.11.002

Stillwell, D. J., & Tunney, R. J. (2009). Melioration behavior in the Harvard game is reduced by simplifying decision outcomes. *Quarterly Journal of Experimental Psychology*, *62*, 2252–2261.

Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political preferences. *American Journal of Political Science*, *50*(3), 755–769.

Tappin, B. M., Pennycook, G., & Rand, D. G. (2020). Thinking clearly about causal inferences of politically motivated reasoning: why paradigmatic study designs often undermine causal inference. *Current Opinion in Behavioral Sciences*, *34*, 81–87.

Torry, H. (2019). Trump Tariffs are short-term pain without long-term gain, economists say. Available at: https://www.wsj.com/articles/trump-tariffs-are-short-term-pain-without-long-term-gain-economists-say-11560440436. Accessed January 28, 2020.

Yi, M. S., Steyvers, M., & Lee, M. (2009). Modeling human performance in restless bandits with particle filters. *The Journal of Problem Solving*, *2*(2), 81–101.

Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review*, *27*, 172–223.

APPENDIX A: List of policies

Public transportation safety standards	Internet infrastructure	Taxes on imported goods
Maternity/paternity leave	Flood risk management	Military spending
Workplace discriminatory policies	Drainage and sewerage	Counterterrorism spending
Equal pay for equal work	Carbon tax	Drug treatment
Social security	Affordable housing	Police spending
Childcare subsidies	Financial regulations	K-12 education spending
Road maintenance	Taxes for the rich	University spending
Public transportation	Taxes for the poor	Border security
Large-scale "green" tech.	Monopolies	Immigration
Subsidize public transit	Reduce drug prices	Marijuana legalization
Air travel infrastructure	Corporate tax rate	Small business tax rate
Gender equality and sexual harassment training		