Cheater detection in politics: Evolution and citizens' capacity to hold political leaders accountable

Troels Bøggild

Department of Political Science

Aarhus University, Denmark

tboeggild@ps.au.dk

Abstract

The average citizen is often unable to distinguish and choose between political leaders according to their ideological profiles. Research using evolutionary theory shows that citizens instead turn to perceptions of procedural fairness concerning whether leaders follow basic decision-making rules such as passing policies without personal interest and displaying responsiveness to citizens' opinions. Some argue that this helps citizens "weed out" questionable leaders; others question citizens' ability to distinguish those following the rules from those who do not. To address this question, I build on psychological research showing that the mind possesses a natural ability, a cheater-detection system, enabling the detection of self-interested others who violate social rules. Introducing an experimental protocol from psychology embedded in cross-national surveys, I show that this system also operates when citizens evaluate political leaders, facilitating identification of leaders who violate basic decision-making rules. The findings advance our understanding of citizens' democratic competences and followers' cognitive abilities generally.

Keywords: politics, democratic accountability, procedural justice, evolutionary psychology, the Wason selection task.

An extensive literature questions the democratic capabilities of citizens. The average citizen cannot distinguish between political leaders according to their ideological positions on, for example, taxation or the welfare state (Converse, 1964; Delli Carpini & Keeter, 1996). Scholars argue that this leaves citizens ill-suited to hold leaders accountable for the policies they introduce and, in turn, vulnerable to exploitation from self-serving elites (Shenkman, 2009). However, research based on evolutionary theory has argued that citizens evaluate leaders and their policies through other means: they rely on perceptions of procedural fairness concerning whether politicians pass policies without personal interest and allow societal actors to voice their opinions in the decision making process (Bøggild & Petersen, 2015; Hibbing & Alford, 2004; von Rueden & van Vugt, 2015).

While it is well established that citizens rely on perceptions of politicians' adherence to procedural fairness, the current debate concerns whether such perceptions are in fact reliable and meaningful to act upon. An optimistic group of scholars holds that such perceptions are useful and constitute a "second-best" option for citizens without an understanding of politicians' ideological orientations. They expect citizens to be able to meaningfully distinguish between politicians according to their adherence to basic democratic rules and principles, which helps ensure that questionable politicians are "weeded out" and that "the process of representation is well-served" (McCurley & Mondak, 1995, 865, 884; Bøggild, 2016b; Smith et al., 2007). Skeptics, however, argue that citizens are merely substituting one inaccurate evaluation (of politicians' ideological orientations) with another (of politicians' adherence to procedural fairness) and consider the typical citizen "an unreliable defender of democratic values" (Sniderman, et al., 1993, 15; Klosko, 2004; Shenkman, 2009). Thus, the unresolved question concerns the *ability* of citizens to *form* procedural fairness evaluations by reliably distinguishing those who play by the rules from those who do not.

Bringing this debate forward requires taking a step back in the causal chain. This article does not provide further evidence that citizens rely on procedural fairness when evaluating

leaders. Taking a step back, it instead asks: Do citizens possess a natural cognitive ability that allows them to distinguish between political leaders according to their adherence to basic rules and principles in group decision making (i.e., procedural fairness)? And does such an ability exceed their reasoning capacities on other logically equivalent cognitive tasks?

To address these questions, the article applies theoretical insights and an experimental protocol from the field of evolutionary psychology. Theoretically, it builds on Cosmides' (1989) social contract theory holding that evolution has equipped humans with a specialized cognitive ability—a cheater-detection system—that allows them to identify cheaters who violate social rules in cooperative relationships. This insight is integrated with Price and van Vugt's (2014) service-for-prestige theory, holding that followers also think of their relationship with leaders as a cooperative relationship involving the risk of cheaters. Against this backdrop, it is argued that citizens are able to recruit their evolved cheater-detection system to identify political leaders who violate basic rules and expectations in group decision making beyond their ability to solve other logically equivalent cognitive tasks.

Methodologically, this prediction is tested using an experimental protocol from psychology, *the Wason selection task (WST)*. For the first time, the WST is embedded in cross-national and nationally representative surveys and applied to a leader-follower context. The WST provides an objective measure of people's cognitive performance when identifying cases or individuals that violate different types of rules. Results from two experiments conducted in culturally diverse countries, Denmark and the United States, confirm that citizens' cheater-detection system assists them in identifying political leaders who violate basic rules in group decision making—an ability that exceeds their ability to reason about other logically equivalent, political tasks.

The presented findings have three important implications. First, they demonstrate that citizens can rely on their evolved cheater-detection system for making distinctions between political leaders according to their adherence to basic democratic rules and principles. Second, the results show that this is a particularly useful fallback option for the least politically sophisticated citizens who are least capable of reasoning about politicians' ideological orientations. Third, while existing work has pointed to several shortcomings when it comes to citizens' abilities to reason about politics, the findings imply that specialized abilities are present for solving tasks that fit the logic of an ancestral adaptive problem such as identifying rule-violating cheaters. This implies that citizens' cognitive abilities for reasoning about politics are mismatched (see Li, van Vugt, & Colarelli, 2018) as they vary substantially according to the evolutionary or adaptive relevance of the given task. As such, the article casts light on the democratic competences of citizens as well as on why citizens often fall short of the ideals laid out in democratic theory.

Evolution, cognitive abilities, and leader evaluations

Evolutionary psychological theory provides a unique framework for understanding the cognitive abilities of the human mind. This framework holds that our ancestors have faced recurring threats to their survival (i.e., adaptive problems) throughout evolutionary history, the implication being that natural selection has equipped the human mind with a set of psychological systems designed to deal with specific adaptive problems (Cosmides & Tooby, 1992; Petersen, 2012). Each of these psychological systems entails two subsystems: 1) A representational system that monitors and scans the environment for cues that the given adaptive problem is present, creating an internal representation or perception to act upon. 2) A motivational system that drives or motivates the individual to respond in specific ways conducive to survival when the representational system detects that the adaptive problem is present (Cosmides, 1985; Delton & Sell, 2014; Petersen, 2015;

Cosmides & Tooby, 2015).[1] A simple example comes from research on pathogen avoidance. Because humans have recurrently faced the adaptive problem of being infected by pathogens, evolution has selected for a sophisticated psychological system to deal with this threat. This system entails a motivational system that makes people strongly averse to objects holding potentially harmful pathogens (e.g., rotten food, open wounds, or sick individuals) and a representational system that assists them in detecting cues signaling that such threats are present (e.g., through smell, taste, or visual cues) (Faulkner et al., 2004; Oaten, Stevenson, and Case, 2009).

This framework should also have important implications for understanding which cognitive abilities are (not) available to followers when they evaluate and choose between leaders in modern society. From an evolutionary perspective, it is hardly surprising that citizens exhibit limited cognitive abilities for evaluating political leaders according to their ideological positions. Mass politics, in which politicians formulate ideological packages for voters to choose from, arose with the recent transition from small-scale to large-scale societies, which in most places has occurred within the last 500 years (Diamond, 1999; Petersen, 2015). Placing leaders on a liberal-conservative ideological continuum according to their opinions on, for example, taxation, the size of the welfare state, or climate change is therefore a relatively novel challenge in human evolutionary history and, by implication, one that should not have been selected for by evolution.[2] However, this does not imply that followers have no cognitive abilities assisting them in evaluating and

---

[1] There is some conceptual overlap concerning the distinctions between representational vs. motivational systems and cognition vs. affect. Affect and emotions are key drivers for motivational systems to propel the individual toward certain adaptive behaviors (e.g., disgust to avoid pathogens or anger to punish cheaters, see Petersen, 2015). Representational systems, in contrast, are akin to the concept of cognition according to its traditional definition as a process of acquiring knowledge through the senses of the organism (i.e., forming representations).

[2] Rather than reasoning about politicians according to their ideological positions, existing work finds that voters turn to stereotypical physical traits (e.g., facial features) to make inferences about politicians' partisan and ideological affiliations (Olivola, Dustin, & Todorov, forthcoming; see also Laustsen & Petersen, 2015)

distinguishing between leaders in and outside the domain of politics. Contrary to much existing work, the evolutionary psychological perspective implies that we cannot simply extrapolate from empirical findings about people's limited abilities for reasoning about leaders' ideological positions to their reasoning about, for example, leaders' adherence to procedural fairness.[3]

Indeed, a growing literature holds that evolution has selected for a sophisticated followership psychology. This psychology motivates followers to rely on information that signals leaders' (1) competence in solving ancestral group problems and (2) pro-social intentions in terms of advancing the interests of the group and its followers instead of self-interested goals (King, Johnson, & Van Vugt, 2009; von Rueden et al., 2014; von Rueden & van Vugt, 2015). For example, followers evaluate and choose between leaders according to physical features (e.g., age, height, vocal pitch, facial masculinity) that signal competence in solving ancestral problems, such as inter-group conflict, maintaining social order, and joint production of public goods (Bøggild & Laustsen, 2016; Klofstad & Anderson, 2018; Laustsen & Petersen, 2015; Spisak et al., 2012; Spisak et al., 2014). Followers also attend to information that signals pro-social rather than exploitative, dominant behavior, particularly leaders' adherence to basic procedural fairness criteria. Procedural fairness criteria concern whether group leaders make impartial decisions without considering personal interests, whether they display responsiveness by allowing followers a voice, and whether they rely on sufficient and balanced information in the decision-making process (Tyler, 1994). Exhibiting adherence to procedural fairness through voice and impartiality, for example, serves as "validation" to followers and as a "diagnostic tool" to ensure that leaders attend to the interests of the group at large (Bøggild & Petersen, 2015; Hibbing & Theiss-Morse, 2008; von Rueden & van

---

[3]Specifically, evolutionary psychology builds on the notion of domain specificity, holding that the brain comprises distinct, specialized systems each designed to solve a specific class of adaptive problems. This implies that humans' cognitive abilities must be evaluated on a case-to-case basis according to their relevance for survival in ancestral environments.

Vugt, 2015, 986; see also Tyler, 1994; van den Bos, Wilke, & Lind 1998 on fairness heuristic theory). That followers rely on perceptions of procedural fairness when evaluating group leaders and their decisions has been labeled one of the most replicated findings in social psychology (Bos, Wilke, and Lind 1998, 1449) and has been demonstrated in multiple leader–follower contexts, including legal, managerial, and educational settings across widely difference cultures (for reviews see (Bøggild & Petersen, 2015; Lind & Earley, 1992). The attention of followers to procedural fairness has also been demonstrated in the political domain when citizens evaluate politicians and has been shown to affect evaluations across different voter segments and partisan divides (Bøggild, 2016a; Hibbing & Alford, 2004; Ramirez, 2008; Tyler, 1994; Ulbig, 2008).

This article adds to the existing literature by testing another observable implication of the evolutionary framework: that the evolved followership psychology not only motivates followers to react to such information but also relies on representational systems for forming reliable evaluations upon which to act. Prior work on competence evaluations shows that followers —even little children—can predict real-world elections through competence evaluations of candidates' faces (Antonakis & Dalgas, 2009; Todorov et al., 2005; see also Bor, 2017), and that followers flexibly turn to different physical cues of competence (e.g., masculinity, age) depending on the given context and problem facing the group (Bøggild & Laustsen, 2016; Laustsen & Petersen, 2015; Spisak et al., 2012; Spisak et al., 2014). This provides at least suggestive evidence that followers make meaningful, systematic distinctions between leaders according to physical cues of competence (but see Antonakis & Dalgas, 2009; Klofstad & Anderson, 2018 for critical discussions). In contrast, no previous studies have addressed the ability of followers to form reliable perceptions of leaders' adherence to procedural fairness. As laid out above, this leaves open whether such perceptions are in fact meaningful to act upon and, in particular, if they constitute a useful alternative for citizens to evaluate politicians and their policies.

The next section builds on insights from social contract theory (Cosmides 1989) and service-for-prestige theory (Price & van Vugt, 2014), arguing that humans hold a specialized cognitive ability—a cheater detection system—that followers can utilize to distinguish between group leaders according to their adherence to procedural fairness—an ability that should exceed their capacity to solve other logically equivalent tasks.

Cheater detection in leader-follower relations

Throughout human evolutionary history, the survival of the individual has depended critically on cooperation with others in social groups (e.g., exchanging food or other resources). Alongside cooperation, however, the risk of being exploited by *cheaters* benefitting from one's cooperative efforts without reciprocating also arose. For cooperation to be a successful strategy, humans would therefore have to engage in cooperation *conditional* on the other party keeping up their end of the bargain by reciprocating cooperative behavior (Axelrod 1984; Trivers 1971).

Cosmides' (1989) social contract theory holds that evolution in response to the recurring problem of cheaters has equipped humans with a psychological system that motivates them to seek out cooperative relationships with reciprocators and to avoid and punish cheaters (see also Cosmides & Tooby, 2015; Trivers, 1971) . As outlined above, this psychological system would only have evolved if it entailed a representational system capable of forming reliable internal representations for the individual to act upon (am I dealing with a cheater or a reciprocator?). Social contract theory therefore also predicts that this system entails a representational system—a cheater-detection system—that becomes active and structures reasoning whenever a situation of conditional cooperation presents itself (Cosmides & Tooby, 2015). This cheater-detection system has two main properties. First, it organizes conditional cooperation in terms of social contracts operating on simple conditional rules that specify an agreement between two agents, which takes the following form: "*If you receive benefit X from me, then you must fulfill my requirement Y.*". Second, this

system effectively directs attention toward information about whether someone cheats by knowingly taking the benefit without fulfilling the requirement.[4] For example, if your neighbor asks to borrow your car and you agree on the condition that the neighbor fills the tank afterwards, the cheater-detection system structures this as an instance of conditional cooperation based on the conditional rule "*If you borrow my car, then you must fill the tank afterwards*" and directs attention toward information on whether the neighbor cheats by borrowing the car (i.e., the benefit) and refuses to fill the tank (i.e., the requirement).

       Price and van Vugt's (2014) service-for-prestige theory argues that our evolved followership psychology also structures leader–follower relationships as an instance of conditional cooperation involving a social contract. Specifically, followers grant leaders who make decisions for the group with a benefit, referred to as prestige, including status, resources, and reproductive opportunities. In exchange, followers make the requirement that the leader uses this position of authority to make group decisions that advance their interests, referred to as service, in terms of coordinating group behavior, maintaining social order, etc. rather than advancing the personal interests of the leader. Importantly, as followers exchange prestige for service they face the risk of cheaters in terms of someone who assumes a leadership position and introduces decisions upon the group (i.e., takes the benefit) but exploits this authority to advance personal rather than group interests (i.e., refuses to fulfill the requirement). As also noted by Tooby, Cosmides, and Price (2006, 110), "[c]heating in leadership involves directing the group in excessively self-interested, group-injurious ways" (see also Hibbing & Alford, 2004; Smith et al., 2007). Because the mind also structures leader-follower relations as instances of conditional cooperation involving a social

---

[4] The point is not that this cheater-detection system is perfect and always identifies cheaters but, rather, that it yields higher performance relative to equivalent cognitive tasks in which no specialized systems are available and people must rely on their "baseline rationality" (Cosmides & Tooby, 2015).

contract, we should expect followers to be able to recruit the cheater-detection system laid out in social contract theory when evaluating the pro-social or "cheating" dispositions of leaders.[5] That is, we should expect the cheater-detection system to structure followers' reasoning about leaders in terms of social contracts operating according to conditional rules, such as "*If a group leader introduces decision X, then he/she must fulfill requirement Y*" and effectively direct attention toward "cheaters" who enforce decisions without attending to the interests of followers.

What information, specifically, do followers attend to when evaluating if leaders are inclined to cheat on this social contract? Identifying cheaters is relatively straightforward in simple, dyadic relationships, as the requirement often constitutes handing over a specific resource (e.g., gas in exchange for borrowing the car). In contrast, evaluating if leaders adhere to the requirement of making group decisions that benefit followers is a complex evaluation; group decisions often have multiple simultaneous consequences, and decisions imposing short-term costs on followers can yield long-term benefits (and vice versa). Research building on the service-for-prestige model has argued that this problem is the key to understanding followers' preoccupation with leaders' adherence to procedural fairness. Information concerning, for example, the responsiveness of leaders to the opinions of followers in the decision-making process (i.e., voice) should serve as a relevant proxy for, and, on average, be correlated with the leaders' pro-social dispositions in group decision making (for a full account of this argument, see Bøggild & Petersen, 2015 see also Hibbing & Alford, 2004; Smith et al., 2007; van Vugt & von Rueden, 2015).

---

[5] The argument following service-for-prestige theory is that followers should recruit the same system for detecting cheaters among leaders as identified in past work on dyadic cooperative relationships (e.g., Cosmides & Tooby, 2015). The evolved followership psychology should constitute one of several subsystems within the broader domain of social exchange, and each subsystem should be able to recruit the cheater detection system as long as the subsystem regulates interactions that fit the logic laid out in social contract theory (i.e., situations involving risks of cheaters who intentionally take benefits without fulfilling requirements).

In sum, followers' cheater-detection system should process their relationships with leaders as conditional cooperation, operating according to conditional rules, such as "*If a group leader introduces a decision, then he/she must first allow followers to voice their opinions*," and assist them in identifying leaders who fail to live up to such conditional rules, beyond their ability to solve other logically equivalent tasks.

Research design

Testing if followers hold a specialized ability to detect leaders who "cheat" by violating procedural fairness criteria in group decision making, requires a research design beyond what is currently available in the leadership or political science literature. Fortunately, the psychological literature offers an experimental paradigm geared specifically for testing people's cognitive abilities in detecting individuals or cases that violate conditional rules of different kinds. Conditional rules can either represent social contracts, which include risks of cheaters (e.g., "*If you borrow my car, then you must fill the tank afterwards*"), or they can represent conditional rules where cheating is not relevant (e.g., "*If someone lives in New York, then he/she is a Yankees fan*"). To test if citizens hold a specialized cognitive ability for detecting leaders who violate social contract rules compared to other conditional rules without risks of cheating, the Wason selection task (WST) is particularly useful. The strength of the WST is that it allows for subtly manipulating the content of the conditional rule presented to subjects. This provides strong internal validity in terms of pinpointing the exact types of tasks or social situations in which citizens exhibit the strongest abilities for detecting rule violations. However, the external validity of this approach—for example, in terms of analyzing how cheater detection abilities affect citizens' vote choice in elections—is low (a point further addressed in the discussion below).

The WST provides an objective performance measure of people's cognitive abilities across different types of conditional rules. The WST is a logic puzzle asking subjects to identify cases that violate a conditional rule with the structure "if P, then Q" (Wason, 1968). The conditional rule states that if an antecedent (P) is true, then a consequent (Q) must also be true. The task then provides the respondent with four cases, each represented by one card. One side of the card states whether the antecedent is true (i.e., whether P or Not-P is the case) and the other side states whether the consequent is true (i.e., whether Q or Not-Q is the case). The subject, who can only see one side of the card, is asked to indicate which card(s) must be turned over to see if any of the cases violate the conditional rule. The four cards represent the logical categories P, Not-P, Q, and Not-Q.

FIGURE 1 ABOUT HERE

Figure 1 presents an example of a classic WST: "the student document task." Subjects are asked to imagine that they have been hired as a secretary at a local high school to ensure that student documents have been processed in accordance with the following conditional rule: "*If a student has the grade 'D,' then his/her document must be marked code '3'*" (if P, then Q). This constitutes a prescriptive conditional rule, as it prescribes how something should or ought to be done. Subjects are then asked to identify student documents that violate this rule and are presented with four cases, each represented by one card. One side of the card contains the student's grade, the other side the code on this student's document. Subjects can only see one side of the card and are asked to indicate which card(s) they would need to turn over to see if these student documents violate the conditional rule. The four cards represent the logical categories P (grade D), Not-P (grade not D), Q (code 3), and Not-Q (code not 3). The correct answer is always to turn over the cards representing the logical categories P (grade D) and Not-Q (code not 3). A student with the grade D must have a document marked code 3 in order to live up to the conditional rule. And a document with the code 7 cannot belong to a student with the grade D to comply with the

conditional rule. The cards representing the logical categories Not-P and Q should not be turned over, as these cases can never violate the rule regardless what is on the other side of these cards.

Early work on the WST has demonstrated that performance on tasks such as the student document task is surprisingly low; typically 0-10 % solve this task correctly (answering P, Not-Q). The major fallacy that most people commit on such tasks is to seek out cases that confirm the rule, even though the WST specifically asks them to seek out disconfirming, rule-violating cases (Dawson, Gilovich, & Regan, 2002; Wason, 1968); that is, conditional reasoning is subjected to an extensive confirmation bias. In the task presented in Figure 1, for example, people look for student documents with a "D" on one side and a "3" on the other (P, Q responses), even though this is logically incorrect.

Later work by evolutionary psychologists has forcefully demonstrated that this confirmation bias and the frequency of logically incorrect answers are dramatically reduced when subjects are presented with a task including a conditional rule that represents a social contract - that is, when there is a risk of someone intentionally taking a benefit without fulfilling an associated requirement. Moreover, performance on tasks including social contracts increases when the risk of cheaters is accentuated; for example, by informing subjects that others might have an incentive to violate the social contract (Cosmides & Tooby, 2015). These findings indicate that there is something specialized in the brain—a cheater-detection system—for identifying individuals with a disposition to "cheat" on social contracts but not for detecting violations of conditional rules in general, which would have had no survival value over the course of evolutionary history (Cosmides & Tooby, 2015).[6] Hence, the WST is useful when testing if people are particularly skilled at

---

[6] Scholarly debate continues as to whether the ability to detect cheaters on the WST is in fact facilitated by an evolved, domain-specific psychological system or reflects the operation of a more general-purpose system (Liberman & Klar, 1996; see Fiddick, Cosmides & Tooby, 2000; Cosmides & Tooby, 2015 for responses). This article does not aim to advance this debate on the exact nature

detecting rule-violating instances in situations involving an individual with a disposition to cheat *relative* to situations in which searching for violations will not reveal individuals disposed to cheat. Importantly, however, it does not provide an *absolute* measure of cheater-detection abilities, because performance in absolute terms will always depend on how clearly the possibility that a search for violations will reveal a cheater is described in the given task. In short, the WST allows for a test of whether followers possess a superior ability to identify group leaders who introduce decisions without adhering to procedural fairness criteria in group decision making compared to other logically equivalent tasks without risks of cheaters.

FIGURE 2 ABOUT HERE

Figure 2 presents two other WSTs, "the voice task" and "the ideology task." In the voice task, the subjects are asked to imagine that they have been hired as journalists to investigate if the mayors in the country comply with the conditional rule: "If a mayor decides to raise taxes, then he/she must first hold public hearings." This is a prescriptive conditional rule like the one presented in the student document task as it prescribes how something should be done. However, the voice task also includes an element of permission (i.e., a permission rule) because it implies that a leader must meet some precondition (holding public hearings) before being allowed to take a given action (raising taxes). As outlined above, followers may be prompted to think of this as a social contract rule because they should identify cheaters in group decision making by looking for individuals who assume a leadership position and exert influence on the group (i.e., take the benefit) and refuse to display consideration and responsiveness to the interests of followers (i.e., fulfill the requirement).

In the ideology task, subjects are presented with an almost identical task but including a slightly different conditional rule: "If a mayor decides to raise taxes, then he/she is a liberal." As

---

of the psychological mechanisms facilitating cheater detection but rather to demonstrate that cheater detection abilities also extend to leader-follower relationships.

in the other tasks, the ideology task asks subjects to identify violations of this rule (i.e., mayors who raise taxes and are conservatives). The crucial difference from the voice task is that this task contains a descriptive conditional rule, as it describes a typical pattern and logic of an ideological belief system (politicians who raise taxes are often liberal). Hence, while the conditional rule is similar to that in the voice task in the sense that it should be familiar to many subjects, it does not include an apparent risk of cheaters since there is no benefit or requirement involved in the rule. Importantly, the three tasks are logically equivalent: Each task includes a conditional rule of the structure "If P, then Q," the same instructions for solving the task, and four cards representing the same four logical categories. According to first order logic, choosing the cards that confirm the antecedent (P) and disconfirm the consequent (Not-Q) is always the correct answer to the WST.

If followers can utilize their evolved cheater-detection system to identify leaders who make decisions without allowing them a voice in the process, it follows that performance should be relatively high on the voice task (i.e., relatively more correct P, Not-Q responses and fewer incorrect P, Q responses). In contrast, performance on the student document task and the ideology task should be markedly lower since there is no cognitive system specialized for detecting violations of such conditional rules, where cheaters are not of relevance. This would lead subjects to rely on their general intelligence or "baseline rationality." This gives rise to *the specialized cheater-detection hypothesis* holding that violation detection will be higher on the voice task than on the student document task and ideology task.

An alternative hypothesis is that citizens cannot rely on their cheater detection system for identifying leaders who refuse followers a voice in group decision making and that subjects must instead rely on their general-purpose cognitive system of general intelligence or "baseline rationality" when asked to solve the voice task. This perspective represents the null hypothesis suggesting that people draw on the same cognitive system to solve this task as they do for reasoning

about all other types of cognitive tasks. This gives rise to *the general rationality hypothesis*, which

holds that performance should be roughly similar across all three tasks.

Another alternative hypothesis is that people perform well on the voice task not

because they hold a specialized ability to detect group leaders who cheat but because they have

become familiar with the rule through past experience. This perspective also implies that there is no

specialized cheater detection system that helps citizens detect leaders who cheat but that the brain

instead possesses a general-purpose learning ability that allows the individual to obtain information

from the environment and recognize and memorize patterns across any type of situation. According

to this perspective, performance on the WST is determined by whether the researcher "create[s] a

sense of reality about the task" by presenting a problem that subjects recognize from their everyday

lives (Johnson-Laird, Legrenzi, & Legrenzi, 1972, 385). This gives rise to *the familiarity hypothesis*

holding that people should exhibit a superior ability to reason about any conditional rule that is

familiar to them through past experience; since the voice task and the ideology task both represent

familiar conditional rules, we should expect performance to be roughly equal on these tasks and

relatively lower on the unfamiliar student document task, independently of whether or not these

rules involve a risk of being cheated.

Overview of studies

To test the competing hypotheses outlined above, two experimental studies were conducted in

different cultural settings: Denmark and the US. While both countries are modern democracies,

Denmark has higher levels of political knowledge, trust, and participation, and Danish media and

citizens focus comparatively less on the political process and scrutinizing politicians' self-serving

motivations (Binderkrantz & Green-Pedersen, 2009; Curran et al., 2009; Norris, 1999). Replicating

the results across these countries would thus strengthen the generalizability of the findings and the

theoretical account stressing an evolved, natural human ability to detect political leaders who break the basic rules of the game.

Study 1 was conducted in Denmark and randomly assigned subjects to one of the three tasks presented above (except for the bottom line specifying the four logical categories). This allows for a simple test of whether performance varies across the three tasks and, in particular, if the subtle difference between the voice task and the ideology task yields a significant difference in performance. Study 2 was conducted in the US. Aside from replicating the findings of Study 1 in a different context, this study aimed to move closer to the proposed mechanism: that the ability of followers to detect leaders who violate procedural fairness criteria in group decision making is in fact facilitated by a psychological system sensitive to the possibility that detecting violations will reveal a cheater—an individual with a disposition to cheat. To this end, the study included two different versions of the voice task either accentuating or downplaying the risk of politicians deliberately cheating on the rule. If followers' abilities to identify political leaders who violate procedural fairness criteria in group decision making are in fact facilitated by a cheater-detection system, we should expect performance on the voice task to increase when possibilities that violations will reveal a cheater is made salient.

Both studies were conducted using online surveys. Compared to earlier work on the WST conducted in laboratory settings with university students, this approach yields higher external validity by allowing for larger, more diverse samples that more closely resemble the populations and for the respondents to answer in the context of their everyday lives (e.g., at home, on the bus). The possible downside of this approach is that subjects are less motivated to spend time reading and solving the task, which would result in a set of random answers at the expense of measurement validity. This should systematically lower the estimates of the absolute levels of correct responses. However, lowering the absolute levels should only bias against the theoretical expectation

expressed in the specialized cheater-detection hypothesis by levelling out the hypothesized differences in correct responses between the voice task and the other tasks. Further analyses from Study 1 using response time measures demonstrate that these differences do indeed increase with respondents' time spent on the page with the task (for full logit model, see the online appendix).

Study 1

Research design and data

Study 1 was an online, between-subject survey experiment conducted on a nationally representative sample of Danish citizens recruited through YouGov, a survey agency, matching the adult voting population on age, sex, education, and geographical location (n = 285). Due to the limited sample size, the analyses were also conducted with a set of control variables (age, gender, education, party affiliation[7]), which did not change the reported findings below in any systematic or substantial way. To test the three competing hypotheses, subjects were randomly assigned to either the student document task, the voice task, or the ideology task, as displayed in the figures above. In addition to obtaining the objective performance of subjects across the three tasks, the survey measured perceived, subjective performance by asking respondents: "How certain are you that you have turned over the right cards?" on a 0-10 scale, ranging from "very uncertain" (0) to "very certain" (10).

Results

Panel a, Figure 3 reports the percentage of correct responses across the three tasks. The pattern that immediately presents itself is that there is substantial variation in performance across the three tasks. Of the subjects assigned to the voice task, 27.37 % provided the correct P, Not-Q response,

---

[7] Subjects answering "Don't know" to any of these items were dropped from the analyses.

which is significantly and substantially higher than the 3.85 % answering correctly on the ideology task (p < 0.001). Subjects assigned to the voice task also performed substantially better than subjects on the student document task in which 0% provided the correct response (p < 0.001). In line with the specialized cheater-detection hypothesis, these results show that when followers reason about group leaders' adherence to procedural fairness criteria in group decision making, their cheater-detection system directs attention toward relevant information on potential rule violators or "cheaters." This ability extends beyond their capacity to single out rule-violating instances on other logically equivalent tasks where detecting cheaters is not of relevance.

FIGURE 3 ABOUT HERE

In contrast, the findings do not support the alternative hypotheses. The relatively high performance on the voice task compared to the student document task and the ideology task demonstrates that the capacity of people to identify leaders who "cheat" extends beyond their general-purpose "baseline rationality" (contrary to the general rationality hypothesis). Moreover, the relatively high performance on the voice task compared to the ideology task suggests that this capacity also extends beyond their abilities to detect violations of other logically equivalent, political tasks where cheater detection is not relevant (contrary to the familiarity hypothesis). In fact, subjects assigned to the ideology task did not perform better than they would by chance[8] if all subjects had provided random answers (3.85 % vs. 6.25 %, p = 0.207). The ideology task asks subjects to identify politicians who make decisions that are inconsistent with their ideological platform (i.e., mayors who raise taxes and are conservatives), and the results indicate that people are

---

[8] Since subjects can provide a total of 16 different possible answers to a WST (i.e., 15 possible combinations of the four cards plus refraining from answering), the proportion of correct answers obtained by chance if all subjects provided random answers should be 6.25% (1/16 = 6.25%).

not equipped with a specialized cognitive ability to evaluate the conduct of politicians according to such descriptive, conditional rules.

Panel b, Figure 3 shows that the pattern is reversed when we consider the extent of confirmation bias (P, Q responses) across the three tasks. On the voice task, only 6.19 % of the subjects made the common mistake of seeking out instances that confirm the rule, which is significantly lower than the 19.63 % on the ideology task and 31.18 % on the student document task. The percentage of P, Q responses on the voice task is not significantly different from chance (6.19 % vs. 6.25 %, $p = 0.979$) while this is the case on both the ideology task (19.63% vs. 6.25%, $p < 0.001$) and the student document task (31.18 % vs. 6.25 %, $p < 0.001$). Again, in line with the specialized cheater-detection hypothesis, the findings indicate that followers draw on a specialized psychological system that directs their attention toward leaders who "cheat" by violating procedural fairness criteria in group decision making while reasoning is more biased toward seeking confirmation of expectations when no risks of cheaters are present.

Panel c, Figure 3 reports how certain respondents subjectively felt about providing the correct answer across the three tasks. In line with the specialized cheater-detection hypothesis, subjects were on average more certain about having turned over the right cards in the voice task (mean = 5.90, SD = 3.02) compared to the ideology task (mean = 4.95, SD = 2.72). Noticeably, however, there is no significant difference between the response certainty of subjects in the student document task (mean = 5.89, SD = 3.42) and the voice task. Why do subjects feel as certain about their response in the student document task, in which 0% provided the correct answer, as in the voice task, in which 27.37 % did? Further analyses show that the answer lies in the extensive confirmation bias making subjects on the student document task *think* they answered correctly thus providing them with a false sense of certainty; subjects providing P, Q responses tended to feel more certain about their answer compared to subjects providing other answers within both the

student document task ((mean = 6.76, SD = 2.92) vs. (mean = 5.48, SD = 3.58), this difference was marginally significant (p = 0.098)) and the ideology task ((mean = 6.10, SD = 2.23) vs. (mean = 4.67, SD = 2.76), this difference was significant (p = 0.031)) but not for subjects in the voice task.[9] These patterns further tentatively illustrate how cognitive processing is qualitatively different—and less biased toward confirmation—in the voice task than in the other tasks.

In sum, Study 1 demonstrates that followers are not generally equipped with a strong psychological ability to reason about conditional rules, even those expressing familiar relationships according to ideological belief systems, but that they do possess a natural cognitive ability for singling out leaders who violate basic rules and expectations in group decision making. This ability exceeds beyond their "baseline rationality" and abilities to solve other equivalent, political tasks.

Study 2

Research design and data

Study 2 was conducted with a US sample recruited through Amazon Mechanical Turk (N = 818). Aside from replicating the results from Study 1 in a different context, Study 2 serves two purposes. First, and most importantly, the study aims to move closer to the proposed mechanism—that the ability to identify group leaders who make decisions without allowing followers a voice in the process does in fact reflect an adaptation for identifying *cheaters* and is not a product of cultural exposure to the rule itself (as suggested by the familiarity hypothesis). Existing work in evolutionary psychology has demonstrated that the extent to which the cheater-detection system is activated depends on a number of situational factors. Did the individual have an incentive to cheat?

---

[9] In fact, the opposite seemed to be the case in the voice task; subjects providing P, Q responses scored *lower* on response certainty compared to the remaining subjects, although this difference does not reach statistical significance ((mean = 4.83, SD = 2.32) vs. (mean = 5.97, SD = 3.06), p = 0.376).

Was cheating accidental or intentional? Such factors each increase the likelihood of activating the cheater-detection system and yield additive effects on performance on the WST (Cosmides, Barrett, & Tooby, 2010; Cosmides & Tooby, 2015). In other words, the cheater-detection system is more likely to be activated and facilitate rule violation detection when the given social situation entails cues that a cheater may be involved. Accordingly, citizens' ability to identify leaders who refuse citizens a voice should vary considerable according to whether the social situation entails a risk of them being taken advantage of by self-interested leaders. Is the ability to identify leaders who refuse followers a voice sensitive to the possibility of actually identifying a cheater?

To test this, the experimental setup from Study 1 was altered in two ways (for full experimental stimuli, see the online appendix). First, the risk of cheating was accentuated in the voice task by informing subjects that they were hired to monitor if mayors "make decisions that benefit themselves rather than ordinary citizens" and that the mayors might cheat on the rule because of the risk of "drawing public attention to unpopular decisions." Compared to the voice task of Study 1, this task makes the social contract element more explicit to subjects by fleshing out that mayors face an incentive to take the benefit without living up to the associated requirement. Second, Study 2 included a fourth task, "the no-cheater voice task," which is identical to the voice task in the sense that subjects are faced with the same rule, the same instructions for solving the task, and the same four cards. The crucial difference is a change in the social situation or scenario surrounding the conditional rule (see also Gigerenzer & Hug, 1992). Here, the subjects are asked to imagine that they have been hired as a secretary at the local mayor's office and are in charge of planning the public hearings. Because nobody in the mayor's office knows when mayors should hold public hearings, the subject is asked to investigate the practice in other municipalities, that is, what mayors customarily do. Hence, the rule in the no-cheater voice task is still a prescriptive rule as subjects are asked to find out when mayors are supposed to hold public hearings (how something

should be done). However, the crucial difference from the voice task is that the risk of mayors intentionally cheating on the rule is downplayed: the subjects are instructed that the purpose of the task is to investigate an administrative practice, and the benefit and requirement involved in this task are not made explicit. According to the specialized cheater detection hypothesis, this change in scenario should be important and decrease the propensity for the cheater detection system to be activated and facilitate rule violation detection. In contrast, if the ability to solve the voice task reflects general intelligence or familiarity with the rule, this scenario change should be unimportant.

Second, Study 2 looks into individual differences in abilities to solve the different tasks. As mentioned above, scholars who are optimistic about citizens' use of procedural fairness perceptions highlight that this should especially be a useful second-best option for those with low political sophistication (i.e., low interest in and abilities to follow politics; see Zaller, 1992). The specialized cheater-detection hypothesis supports this notion: Although the ability to reason about political ideology should be concentrated among the most politically sophisticated (Delli Carpini & Keeter, 1996; Zaller, 1992), the ability to reason about and identify politicians who violate procedural fairness criteria in group decision making should be a species-typical feature independent of prior interest in and familiarity with politics.[10] Hence, we should expect that political sophistication should be a significant predictor of performance on the ideology task but not on the voice task, and the effects should be significantly larger in the former.[11]

Results

---

[10] This prediction is contrary to what we should expect from the familiarity hypothesis, holding that the ability to detect leaders who cheat increases with past exposure to the political rule and, by implication, with political sophistication.

[11] Effects on performance in the student document task and the no-cheater voice task are not reported since no immediate theoretical expectations relate to these (but see the online appendix for regression models across all tasks).

Panel a, Figure 4 shows the percentage of correct responses across the four tasks. Subjects assigned to the voice task provided substantially and significantly more correct P, Not-Q responses (50.90 %) than subjects on the student document task (2.74 %). Notably, performance on the voice task is considerably higher in this study than in Study 1 (p < 0.001). This is consistent with the theoretical expectations, as the risk of cheaters was accentuated in the voice task of Study 2, which should increase performance according to the specialized cheater-detection hypothesis (but not according to the general rationality or familiarity hypothesis).[12] Performance on the voice task was also significantly higher than on the ideology task (16.95%). As in Study 1, in line with the specialized cheater-detection hypothesis, these findings indicate that followers hold a specialized cognitive ability that allows them to identify leaders who violate procedural fairness criteria that extends beyond their reasoning abilities in other situations where cheaters is not of relevance.

Moreover, a comparison of performance on the no-cheater voice task (33.67%) shows that the change in scenario causes a significant 17.23 percentage point drop in correct responses. Hence, although the rule is identical in the two tasks, subjects exhibit a superior ability to detect leaders who violate rules and expectations for group decision making when this might reveal that the mayors are cheaters, than when that would simply reveal what their common practice is. This effect further underlines that the ability of subjects to solve the voice task is facilitated by a specialized psychological system sensitive to detecting cheaters, and is not a product of cultural exposure to the rule itself. The effect of the scenario change also illustrates how the WST does not provide a meaningful measure of followers' cheater-detection abilities in absolute terms, since levels of violation detection on the WST always depends on the extent to which the possibility of

---

[12] Moreover, the difference in performance between Studies 1 and 2 is much less pronounced for respondents assigned to the student document task and the ideology task. This means that the substantial difference in performance on the voice task between the two studies cannot simply be attributed to general country differences in reasoning abilities.

detecting cheaters is made salient in the task description (see also Cosmides & Tooby, 2015).

Instead, the results show how the cheater-detection system is flexibly designed to direct attention

toward rule violators whenever they face a situation that entails the possibility of a group leader

choosing to cheat on a social contract with citizens.

Panel b, Figure 4 reports the extent of confirmation bias across the four tasks. As in

Study 1, subjects were significantly less likely to make the mistake of seeking out instances that

confirm the rule on the voice task (14.8%) than on the student document task (35.91%). The same

trend is present when we compare the voice task to the ideology task (20.90%) although this

difference does not reach statistical significance at conventional levels. Finally, there is a significant

difference in confirmation bias between the voice task and the no-cheater voice task (26.26%),

ruling out the possibility that enhanced violation detection for the voice task is due simply to

familiarity with the rule itself.

Panel c, Figure 4 displays the perceived, subjective certainty of subjects in providing

the correct response across the four tasks. In line with the specialized cheater-detection hypothesis,

the results show that subjects felt significantly more certain about their choice on the voice task

(mean = 7.41, SD = 2.46) than on the student document task (mean = 6.30, SD = 3.04), the ideology

task (mean = 5.92, SD = 2.75), and the no-cheater voice task (mean = 6.85, SD = 2.44).

FIGURE 4 ABOUT HERE

Turning to the analysis of individual differences in performance, Figure 5 graphically

displays the predicted probabilities of correct responses in the voice task and the ideology task

across levels of political sophistication,[13] controlling for age, gender, education, and ideological

---

[13]An index for political sophistication was created consisting of two components (see, e.g., Zaller 1992). First, subjects were asked about their political interest on a 1–4 scale. Second, subjects were asked four factual questions about politics to create a sub-index of the number of correct responses (0–4) to tap individual differences in respondents' cognitive abilities for understanding politics. The

placement (for full logit models, see the online appendix).[14] As expected, panel a shows that the

likelihood of subjects providing correct responses does not significantly increase with political

sophistication in the voice task ($p = 0.908$), while this is the case in the ideology task ($p = 0.006$). In

substantial terms, the predicted probability of a correct response in the ideology task increases from

1.34 % among the least politically sophisticated to 29.67 % among the most sophisticated; in the

voice task, however, the predicted probability of a correct response is similar with 52.39 % and

50.41 % among the least and most politically sophisticated, respectively. This difference in effects

across the two tasks is significant ($p = 0.019$ for interaction effect). In line with the specialized

cheater-detection hypothesis, these findings suggest that the ability to detect political leaders who

violate procedural fairness criteria is a species-typical feature and independent of past knowledge of

and familiarity with such rules. The results also suggest that while the ability to reason about and

evaluate political leaders and their conduct according to political concepts such as ideology might

be reserved for the politically sophisticated (Delli Carpini & Keeter, 1996; Zaller, 1992) the

capacity to single out group leaders who "cheat" on rules in group decision making is not.

Evaluating leaders according to their adherence to procedural fairness is therefore particularly

useful for the least politically sophisticated whose cheater detection abilities are no less accurate

than those of the most politically sophisticated.

FIGURE 5 ABOUT HERE

        In sum, Study 2 corroborates the findings of Study 1 in a different cultural setting by

demonstrating that followers exhibit a specialized ability to identify leaders who violate procedural

---

two components were summarized (each weighting 50%) into an additive index that was rescaled
0–1, reflecting minimum and maximum political sophistication, respectively ($r = 0.265$, $M = 0.737$,
$SD = 0.215$).
[14]For operationalizations of all variables, see the online appendix. The reported findings remain
significant without control variables.

fairness criteria in group decision making compared to logically equivalent cognitive tasks involving no possibility of detecting a leader with a disposition to cheat. Moreover, this ability is not contingent on the individual possessing high levels of political sophistication but reflects a basic psychological system sensitive to the risk of cheating by potentially self-interested group leaders.

Discussion

As many citizens struggle to understand the main ideological differences between political leaders and their policies and make informed decisions on this basis, they often turn to other information. Such information includes whether leaders pass policies in line with procedural fairness criteria: did the politician allow citizens a voice in the decision-making process? Did the politician have a personal interest in the policy? Was the policy based on sufficient and balanced information? (Bøggild, 2016a; Hibbing & Alford, 2004; Tyler, 1994; Ulbig, 2008). This raises two important questions that remain unprobed in existing work: Do citizens possess a natural cognitive ability that allows them to distinguish between political leaders according to their adherence to basic rules and principles in group decision making (i.e., procedural fairness)? And does such an ability exceed their reasoning capacities on other logically equivalent cognitive tasks? This article draws on theoretical insights and an experimental protocol from evolutionary psychology (Cosmides 1989; Price & van Vugt, 2014) and demonstrates that citizens can utilize a specialized cheater-detection system that assists them in identifying and distinguishing political leaders according to their adherence to democratic rules and expectations, such as allowing societal actors a voice in group decision making.

These findings have important implications for understanding the democratic capabilities of citizens and the relationship between leaders and followers more generally. Specifically, the findings address an unresolved issue in the literature by demonstrating that citizens can rely on their evolved cheater-detection system for "weeding out" politicians who display anti-

social, rule-violating behavior. The fact that this cognitive ability also manifests itself among the least politically sophisticated citizens implies that relying on perceptions of procedural fairness should be a particularly viable fallback option among those least capable of reasoning about ideology. This cognitive ability should also be important outside the political domain, for example, in corporate settings where leaders' adherence to procedural fairness is important for employee motivation, job satisfaction, and commitment to the firm and its goals (Lavelle et al., 2009; Pillai, Schriesheim, & Williams, 1999).

More generally, the article points to evolutionary theory as a useful framework for understanding and mapping the cognitive abilities available to followers when evaluating and choosing between group leaders. While existing work has provided compelling evidence that humans hold a sophisticated followership psychology that motivates them to evaluate and choose between group leaders according to specific physical and behavioral traits (e.g., Laustsen & Petersen, 2015; Spisak et al., 2012; von Rueden & van Vugt, 2015), this article suggests that this psychology should also be accompanied by subsystems for forming reliable internal representations or perceptions for followers to act upon. In this respect, it is important to stress that "evolved" means neither "optimal" nor "perfect." Indeed, this article suggests that citizens' cognitive architecture is mismatched in the sense that they are not equipped with specialized abilities for solving other—arguably important—political tasks, for example, reasoning about the relationship between politicians' ideological labels and their actual policies. Hence, the evolved followership psychology is ecologically rational, meaning that it is designed only to categorize leaders according to information that has been conducive to survival over evolutionary history (see e.g., von Rueden & van Vugt, 2015; Li, van Vugt, & Colarelli, 2018). Moreover, as demonstrated in this article, the specialized cheater detection system substantially reduces incorrect attempts at identifying rule violators but does not eliminate them entirely (see also Cosmides & Tooby, 2015). Citizens who are

unable to distinguish between politicians according to their ideological profile can turn to

perceptions of procedural fairness as a more reliable—albeit imperfect—substitute. Given that the

debate remains open as to whether and to what extent followers' use of evolved psychological

systems for leader choice is cause for comfort or alarm (e.g., Dalgas & Antonakis, 2009; Klofstad &

Anderson, 2018; Shenkman, 2009), it is important to further theorize about and test which tasks and

problems the followership psychology is (not) configured to solve.

      At the methodological level, the article underlines that the cognitive abilities available

to followers are domain-specific and must be evaluated across different types of tasks and

problems. As the findings demonstrate, cognitive performance varies dramatically across logically

equivalent tasks according to their ancestral, evolutionary relevance. This implies that generic

measures of how well equipped followers are to reason about and choose between group leaders

(e.g., questions tapping citizens' political sophistication; e.g., Zaller, 1992) will likely overlook

variation in performance across different domains and, possibly, underestimate the natural,

cognitive abilities of those scoring low on such measures. This underlines the need for developing

innovative research designs and measures to better understand the cognitive abilities of followers in

and outside the political domain.

      Two points concerning the external validity of the findings warrant mention. First,

while the strength of the WST is that it allows us to isolate and compare cognitive performance

across different tasks and domains, it also puts subjects in a stylized and artificial information-

processing context. In the real world, citizens primarily receive political information from the media

or their social network, and it remains to be seen exactly how the cheater-detection system directs

attention and affects opinions in such settings and in competition with other types of information

and motivations. Second, the WST provides no meaningful measure of followers' cheater detection

ability in absolute terms, demonstrating only that this capacity extends beyond their "baseline

rationality" and ability to identify rule-violating instances in other logically equivalent, political tasks. Both points underline the need for future work to investigate how and to what extent the cheater-detection psychology operates in real, everyday life settings.

In closing, the article points to new avenues for future research to explore. First, whereas the results focus on followers' abilities to identify cheaters among leaders, it also follows from service-for-prestige theory that this system should help leaders and followers to monitor and identify potential followers who refuse to deliver on their end of the social contract by not granting prestige to group leaders who provide service for group members. Identifying and counteracting such cheaters among followers may be equally important, since leaders who experience that followers fail to uphold the social contract should lose their incentives to provide benefits for the group and turn to anti-social and self-aggrandizing behavior. Hence, the ability to identify and counteract followers who cheat could have important implications for preventing organizational failure or even democratic backsliding.

Second, it remains unknown how well the cheater-detection system functions to identify cheaters among leaders (the focus in this study) relative to identifying rule violators in simple, dyadic relationships between group members (the focus in existing work, e.g., Cosmides, 1989). It is possible that anti-social, exploitative behavior on the part of leaders with the authority to impose decisions on the group has constituted a more severe adaptive problem and, in turn, that the cheater-detection system is particularly sensitive to leaders who display signs of exploitative behavior.

Third, it remains an open question how the cheater detection system categorizes leaders when relevant information on cooperative or cheating dispositions is unavailable or such information is ambiguous. On the one hand, existing research on cooperation in dyadic relationships suggests that the cheater detection system should be biased in favor of categorizing others as

reciprocators rather than cheaters because losing out on a potential long-term cooperative relationship with a reciprocator is a bigger loss than incurring a one-time cost from interacting with a cheater (Cosmides, Barret & Tooby, 2010; Delton, Krasnow, Cosmides & Tooby, 2011). This would imply that followers generally put trust in the motives of leaders until they are proven wrong and could help explain why performance in identifying cheaters on the WSTs in this study is not high in absolute terms. On the other hand, the cheater detection system could be biased in favor of categorizing leaders as cheaters (as opposed to in dyadic cooperation) because ill-intentioned leaders can potentially inflict substantial costs on the individual due to their authority to introduce decisions upon others. This may help explain why public trust in political leaders is low across most Western countries (Dalton, 2017; Norris, 2011) and why high levels of political cynicism and mistrust remain relatively fixed and difficult to change. Ultimately, it may be part of the human condition to be skeptical and wary of the intentions of those who hold positions of power.

ACKNOWLEDGMENTS

REFERENCES

Antonakis, J., & O. Dalgas. 2009. "Predicting Elections: Child's Play!" *Science* 323 (5918): 1183–1183. https://doi.org/10.1126/science.1167748.

Axelrod, R. 1984. *The Evolution of Cooperation*. New York: Basic Books.

Binderkrantz, A., and C. Green-Pedersen. 2009. "Policy or Processes in Focus?" *The International Journal of Press/Politics* 14 (2): 166–85. https://doi.org/10.1177/1940161209333088.

Bøggild, T., and M. Petersen. 2015. "The Evolved Functions of Procedural Fairness: An Adaptation for Politics." In *The Evolution of Morality*, T. Shackelford & R. Hansen, 247–76. New York: Springer.

Bøggild, T. 2016a. "How Politicians' Reelection Efforts Can Reduce Public Trust, Electoral Support, and Policy Approval." *Political Psychology* 37 (6): 901–19.

———. 2016b. *Procedural Fairness and Public Opinion*. Aarhus: Politica.

Bøggild, T., and L. Laustsen. 2016. "An Intra-Group Perspective on Leader Preferences: Different Risks of Exploitation Shape Preferences for Leader Facial Dominance." *The Leadership Quarterly* 27 (October). https://doi.org/10.1016/j.leaqua.2016.09.003.

Bor, A. 2017. "Spontaneous Categorization along Competence in Partner and Leader Evaluations." *Evolution and Human Behavior* 38 (4): 468–73. https://doi.org/10.1016/j.evolhumbehav.2017.03.006.

Bos, K. van den, H. Wilke, and E. Lind. 1998. "When Do We Need Procedural Fairness? The Role of Trust in Authority." *Journal of Personality and Social Psychology* 75 (6): 1449–58. https://doi.org/10.1037/0022-3514.75.6.1449.

Converse, P. 1964. "The Nature of Belief Systems in Mass Publics." In *Ideology and Discontent*, D. Apter, 206–61. New York: Free Press.

Cosmides, L. 1985. *Deduction or Darwinian Algorithms? An Explanation of the "elusive" Content Effect on the Wason Selection Task (Doctoral Dissertation, Harvard University)*. (University Microfilms No. 86-02, 206).

———. 1989. "The Logic of Social Exchange: Has Natural Selection Shaped How Humans Reason? Studies with the Wason Selection Task." *Cognition* 31 (3): 187–276.

Cosmides, L., H. Barrett, & J. Tooby. 2010. "Adaptive Specializations, Social Exchange, and the Evolution of Human Intelligence." *Proceedings of the National Academy of Sciences* 107 (Supplement 2): 9007–14. https://doi.org/10.1073/pnas.0914623107.

Cosmides, L., & John T. 2015. "Adaptations for Reasoning about Social Exchange." In Buss, D. (Ed.) *The Handbook of Evolutionary Psychology*, 2nd ed., 625–68. Hoboken, NJ: John Wiley & Sons.

Curran, J., S. Iyengar, A. Lund, & I. Salovaara-Moring. 2009. "Media System, Public Knowledge and Democracy." *European Journal of Communication* 24 (1): 5–26. https://doi.org/10.1177/0267323108098943.

Dalton, R. 2017. "Political Trust in North America." In *Handbook on Political Trust*, 20. London: Edward Elgar.

Dawson, E., T. Gilovich, & D. Regan. 2002. "Motivated Reasoning and Performance on the Was on Selection Task." *Personality and Social Psychology Bulletin* 28 (10): 1379–87. https://doi.org/10.1177/014616702236869.

Delli Carpini, M., & S. Keeter. 1996. *What Americans Know about Politics and Why It Matters*. New Haven: Yale University Press.

Delton, A., & A. Sell. 2014. "The Co-Evolution of Concepts and Motivation." *Current Directions in Psychological Science* 23 (2): 115–20. https://doi.org/10.1177/0963721414521631.

Delton, A., M. Krasnow, L. Cosmides, & J. Tooby. 2011. Evolution of direct reciprocity under uncertainty can explain human generosity in one-shot encounters. *PNAS* 108 (32), 13335-40.

Diamond, J. 1999. *Guns, Germs and Steel: The Fates of Human Societies*. New York: Norton.

Faulkner, J., M. Schaller, J. Park, & L. Duncan. 2004. "Evolved Disease-Avoidance Mechanisms and Contemporary Xenophobic Attitudes." *Group Processes & Intergroup Relations* 7 (4): 333–53. https://doi.org/10.1177/1368430204046142.

Gigerenzer, G., & K. Hug. 1992. "Domain-Specific Reasoning: Social Contracts, Cheating, and Perspective Change." *Cognition* 43 (2): 127–71.

Hibbing, J., & J. Alford. 2004. "Accepting Authoritative Decisions: Humans as Wary Cooperators." *American Journal of Political Science* 48 (1): 62–76. https://doi.org/10.1111/j.0092-5853.2004.00056.x.

Hibbing, J., & E. Theiss-Morse. 2008. "Voice, Validation, and Legitimacy." In *Cooperation: The Political Psychology of Effective Human Interaction*, Sullivan & Sullivan (Eds.) 123–42. Blackwell: Malden, MA.

Johnson-Laird, P. , P. Legrenzi, & M. Sonino Legrenzi. 1972. "Reasoning and a Sense of Reality." *British Journal of Psychology* 63 (3): 395–400. https://doi.org/10.1111/j.2044-8295.1972.tb01287.x.

King, A., D. Johnson, & M. van Vugt. 2009. "The Origins and Evolution of Leadership." *Current Biology* 19 (19): R911–16. https://doi.org/10.1016/j.cub.2009.07.027.

Klofstad, C., & R. Anderson. 2018. "Voice Pitch Predicts Electability, but Does Not Signal

Leadership Ability." *Evolution and Human Behavior* 0 (0).

https://doi.org/10.1016/j.evolhumbehav.2018.02.007.

Klosko, G.. 2004. *Democratic Procedures and Liberal Consensus*. Oxford; New York: Oxford

University Press.

Laustsen, L., & M. Petersen. 2015b. "Does a Competent Leader Make a Good Friend? Conflict,

Ideology and the Psychologies of Friendship and Followership." *Evolution and

Human Behavior* 36 (4): 286–93.

https://doi.org/10.1016/j.evolhumbehav.2015.01.001.

Li, N., M. van Vugt, & S. Colarelli. 2018. "The Evolutionary Mismatch Hypothesis: Implications

for Psychological Science." *Current Directions in Psychological Science* 27 (1): 38–

44. https://doi.org/10.1177/0963721417731378.

Liberman, N. & Y. Klar. 1996. Hypothesis testing in Wason 's selection task: social exchange

cheating detection or task understanding. *Cognition* (58), 127-156

Lind, E., & P. C. Earley. 1992. "Procedural Justice and Culture." *International Journal of

Psychology* 27 (2): 227–42. https://doi.org/10.1080/00207599208246877.

McCurley, C., & J. Mondak. 1995. "Inspected by #1184063113: The Influence of Incumbents'

Competence and Integrity in U.S. House Elections." *American Journal of Political

Science* 39 (4): 864–85. https://doi.org/10.2307/2111660.

Norris, P. 1999. *Critical Citizens: Global Support for Democratic Government*. New York: Oxford

University Press.

———. 2011. *Democratic Deficit: Critical Citizens Revisited*. Cambridge: Cambridge University

    Press.

Oaten, M., R. Stevenson, & T. Case. 2009. "Disgust as a Disease-Avoidance Mechanism."

    Psychological Bulletin 135 (2): 303–21. https://doi.org/10.1037/a0014823.

Olivola C., T. Dustin, & T. Alexander. forthcoming. "Republican Voters Prefer Candidates Who

    Have Conservative-Looking Faces: New Evidence From Exit Polls." *Political*

    Psychology (Early view). https://doi.org/10.1111/pops.12489.

Petersen, M. 2012. "Social Welfare as Small-Scale Help: Evolutionary Psychology and the

    Deservingness Heuristic." *American Journal of Political Science* 56 (1): 1–16.

———. 2015. "Evolutionary Political Psychology." In *Handbook of Evolutionary Psychology*,

    Buss, D. Hoboken: John Wiley & Sons Ltd.

———. 2015. "Evolutionary Political Psychology: On the Origin and Structure of Heuristics and

    Biases in Politics." *Political Psychology* 36 (February): 45–78.

    https://doi.org/10.1111/pops.12237.

Price, M., & M. van Vugt. 2014. "The Evolution of Leader-Follower Reciprocity: The Theory of

    Service-for-Prestige." *Frontiers in Human Neuroscience* 8: 363.

    https://doi.org/10.3389/fnhum.2014.00363.

Ramirez, M. 2008. "Procedural Perceptions and Support for the U.S. Supreme Court." *Political*

    Psychology 29 (5): 675–698. https://doi.org/10.1111/j.1467-9221.2008.00660.x.

Shenkman, Rick. 2009. *Just How Stupid Are We?: Facing the Truth About the American Voter*.

    New York: Basic Books.

Smith, K., C. Larimer, L. Littvay, & J. Hibbing. 2007. "Evolutionary Theory and Political

  Leadership: Why Certain People Do Not Trust Decision Makers." *The Journal of*

  Politics 69 (02): 285–99. https://doi.org/10.1111/j.1468-2508.2007.00532.x.

Sniderman, P., R. Brody, & P. Tetlock. 1993. *Reasoning and Choice: Explorations in Political*

  Psychology. Cambridge: Cambridge University Press.

Spisak, B., A. Grabo, R. Arvey, & M. van Vugt. 2014. "The Age of Exploration and Exploitation:

  Younger-Looking Leaders Endorsed for Change and Older-Looking Leaders

  Endorsed for Stability." *The Leadership Quarterly* 25 (5): 805–16.

  https://doi.org/10.1016/j.leaqua.2014.06.001.

Spisak, B., A. Homan, A. Grabo, & M. van Vugt. 2012. "Facing the Situation: Testing a Biosocial

  Contingency Model of Leadership in Intergroup Relations Using Masculine and

  Feminine Faces." *The Leadership Quarterly* 23 (2): 273–80.

  https://doi.org/10.1016/j.leaqua.2011.08.006.

Todorov, A., A. Mandisodza, A. Goren, and C. Hall. 2005. "Inferences of Competence from Faces

  Predict Election Outcomes." *Science* 308 (5728): 1623–26.

  https://doi.org/10.1126/science.1110589.

Tooby, J., L. Cosmides, & H. Barrett. 2005. "Resolving the Debate on Innate Ideas." In *The Innate*

  Mind: Structure and Content, P. Carruthers. S. Laurence, S. Stich, 305–37. New

  York: Oxford University Press.

Tooby, J., L. Cosmides, & M. Price. 2006. "Cognitive Adaptations for N-Person Exchange: The

  Evolutionary Roots of Organizational Behavior." *Managerial and Decision*

  Economics 27 (2-3), 103-129

Trivers, R. 1971. "The Evolution of Reciprocal Altruism." *The Quarterly Review of Biology* 46 (1): 35–57.

Tyler, T. 1994. "Governing amid Diversity: The Effect of Fair Decisionmaking Procedures on the Legitimacy of Government." *Law & Society Review* 28 (4): 809. https://doi.org/10.2307/3053998.

Ulbig, S. 2008. "Voice Is Not Enough The Importance of Influence in Political Trust and Policy Assessments." *Public Opinion Quarterly* 72 (3): 523–39. https://doi.org/10.1093/poq/nfn030.

von Rueden, C. , & M. van Vugt. 2015. "Leadership in Small-Scale Societies: Some Implications for Theory, Research, and Practice." *The Leadership Quarterly* 26: 978–90.

von Rueden, C. , M. Gurven, H. Kaplan, & J. Stieglitz. 2014. "Leadership in an Egalitarian Society." *Human Nature* 25 (4): 538–66. https://doi.org/10.1007/s12110-014-9213-4.

Wason, C. 1968. "Reasoning about a Rule." *The Quarterly Journal of Experimental Psychology* 20 (3): 273–81. https://doi.org/10.1080/14640746808400161.

Zaller, J. 1992. *The Nature and Origins of Mass Opinion*. Cambridge University Press.

APPENDIX A: Correlation matrices (Studies 1 and 2)

Table I: Correlation matrix (Study 1)

	Student document task vs. Voice task	Student document task vs. Ideology task	Voice task vs. Ideology task	Correct response	Confirmation Bias	Response certainty
Student document task vs. Voice task	1					
Student document task vs. Ideology task	.	1				
Voice task vs. Ideology task	.	.	1			
Correct response	0.3897	0.1334	-0.3284	1		
Confirmation Bias	-0.3223	-0.1331	-0.1981	-0.1696	1	
Response certainty	0.0011	-0.1516	-0.1631	0.1937	0.1185	1

Table II: Correlation matrix (Study 2)

	Student document task vs. Voice task	Student document task vs. Ideology task	Voice task vs. Ideology task	Voice task vs. No Cheater	Ideology task vs. No Cheater	Correct response	Confirmation Bias	Response Certainty	Political Sophistication	Female	Age	Education	Ideology
Student document task vs. Voice task	1												
Student document task vs. Ideology task	.	1											
Voice task vs. Ideology task	.	.	1										
Voice task vs. No Cheater	.	.	.	1									
Ideology task vs. No Cheater	.	.	.	.	1								
Correct response	0.5425	0.2457	-0.3518	-0.1737	0.191	1							
Confirmation Bias	-0.2429	-0.1655	0.0784	0.1425	0.0643	-0.3431	1						
Response Certainty	0.1963	-0.0664	-0.2759	-0.1135	0.1774	0.1745	0.0259	1					
Political Sophistication	0.0234	0.0026	-0.0217	-0.0105	0.0109	0.0791	-0.0164	0.0777	1				
Female	0.0567	0.0341	-0.0223	-0.0213	0.0011	0.0134	0.0351	-0.0504	-0.0746	1			
Age	0.0454	0.029	-0.0163	0.006	0.0229	0.0261	-0.0339	0.0574	0.2332	0.0594	1		
Education	0.0022	-0.0073	-0.0096	-0.0803	-0.0683	0.0747	-0.0001	0.0068	0.1875	-0.0121	0.0839	1	
Ideology	0.0549	0.0292	-0.0252	-0.0027	0.0226	-0.0241	-0.0274	0.0469	-0.0707	-0.0596	0.1212	-0.0509	1

The student document task

Imagine that you have been hired as a secretary at a local high school. Part of your new job is to make sure that the student documents have been processed correctly.

You have been asked to make sure that the documents have been marked with different codes according to the grades that the students got.

In general, if a student has the grade "D," then his/her student document must be marked with the code "3." However, your colleague thinks that some documents do not actually comply with this rule.

Therefore, you choose to investigate if the student documents of the school comply with the following rule:

If a student has the grade "D," then his/her document must be marked code "3."

The cards below have information about documents of four different students. Each card represents one student. One side of the card tells the student's grade, and the other side tells the code on the student's document.

It is up to you to investigate if one or more of the documents violate the rule above. Indicate which card(s) you would need to turn over in order to see if the student documents violate this rule. Do not turn over any more cards than absolutely necessary.

Grade D	Grade F	Code 3	Code 7
☐	☐	☐	☐
P	*Not-P*	*Q*	*Not-Q*

Figure 1: The student document task (Study 1 and 2)

The voice task	The ideology task
Imagine that you have been hired as a journalist at a local news station. Part of your new job is to monitor whether mayors across the country act and behave according to certain expectations.	Imagine that you have been hired as a journalist at a local news station. Part of your new job is to monitor whether mayors across the country act and behave according to certain expectations.
It is clear that mayors in most municipalities have raised taxes for ordinary citizens. Since the decision to raise taxes is often unpopular among citizens, you have been asked to investigate this trend in detail.	It is clear that mayors in most municipalities have raised taxes for ordinary citizens. Since the decision to raise taxes is often unpopular among citizens, you have been asked to investigate this trend in detail.
In general, if a mayor decides to raise taxes, then he/she must first hold public hearings where people have the opportunity to voice their opinions on the decision. However, your colleague thinks that some mayors do not actually comply with this rule.	In general, if a mayor decides to raise taxes, then he/she is a liberal rather than a conservative. However, your colleague thinks that some mayors do not actually comply with this rule.
Therefore, you choose to investigate if mayors across the country comply with the following rule:	Therefore, you choose to investigate if mayors across the country comply with the following rule:
If a mayor decides to raise taxes, then he/she must first hold public hearings.	**If a mayor decides to raise taxes, then he/she is a liberal.**
The cards below have information about four different mayors. Each card represents one mayor. One side of the card tells whether or not the mayor raised taxes, and the other side tells whether or not he/she held public hearings.	The cards below have information about four different mayors. Each card represents one mayor. One side of the card tells whether or not the mayor raised taxes, and the other side tells whether or not he/she is a liberal.
It is up to you to investigate if one or more of the mayors violate the rule above. Indicate which card(s) you would need to turn over in order to see if the mayors violate this rule. Do not turn over any more cards than absolutely necessary.	It is up to you to investigate if one or more of the mayors violate the rule above. Indicate which card(s) you would need to turn over in order to see if the mayors violate this rule. Do not turn over any more cards than absolutely necessary.
Raised taxes / Did not raise taxes / Held public hearings / Did not hold public hearings	Raised taxes / Did not raise taxes / Liberal / Not liberal
P Not-P Q Not-Q	P Not-P Q Not-Q

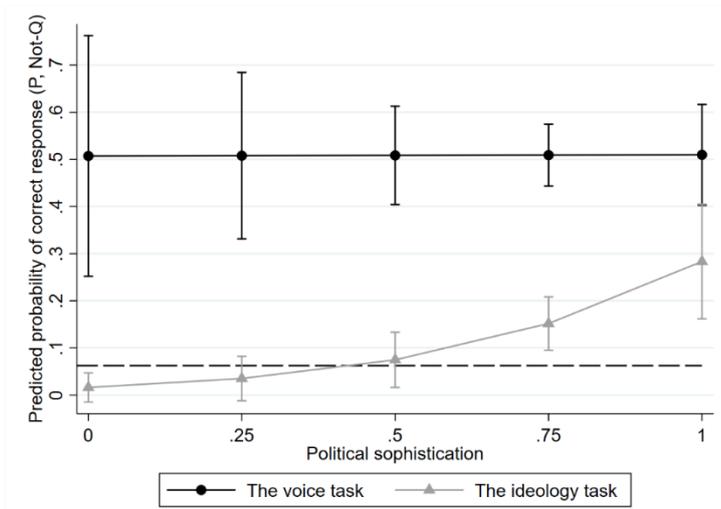**Figure 2: The voice task and the ideology task (Study 1).**
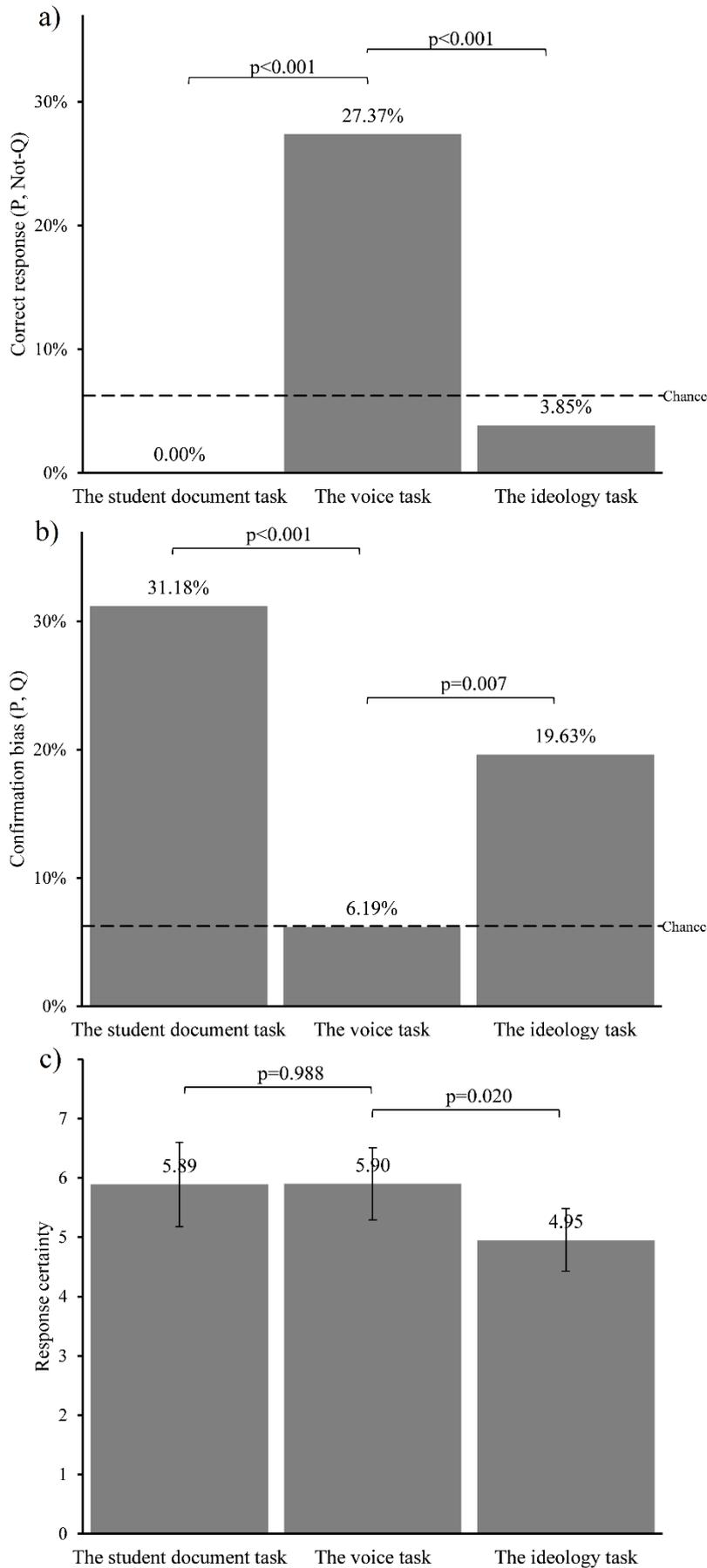
Note: T-tests are two-sided. Error bars around means represent 95% confidence intervals. n = 285.

Figure 3: Percentage correct P, Not-Q responses (panel a) and incorrect P, Q responses (panel b), and response certainty (panel c) across the three Wason selection tasks (Study 1).
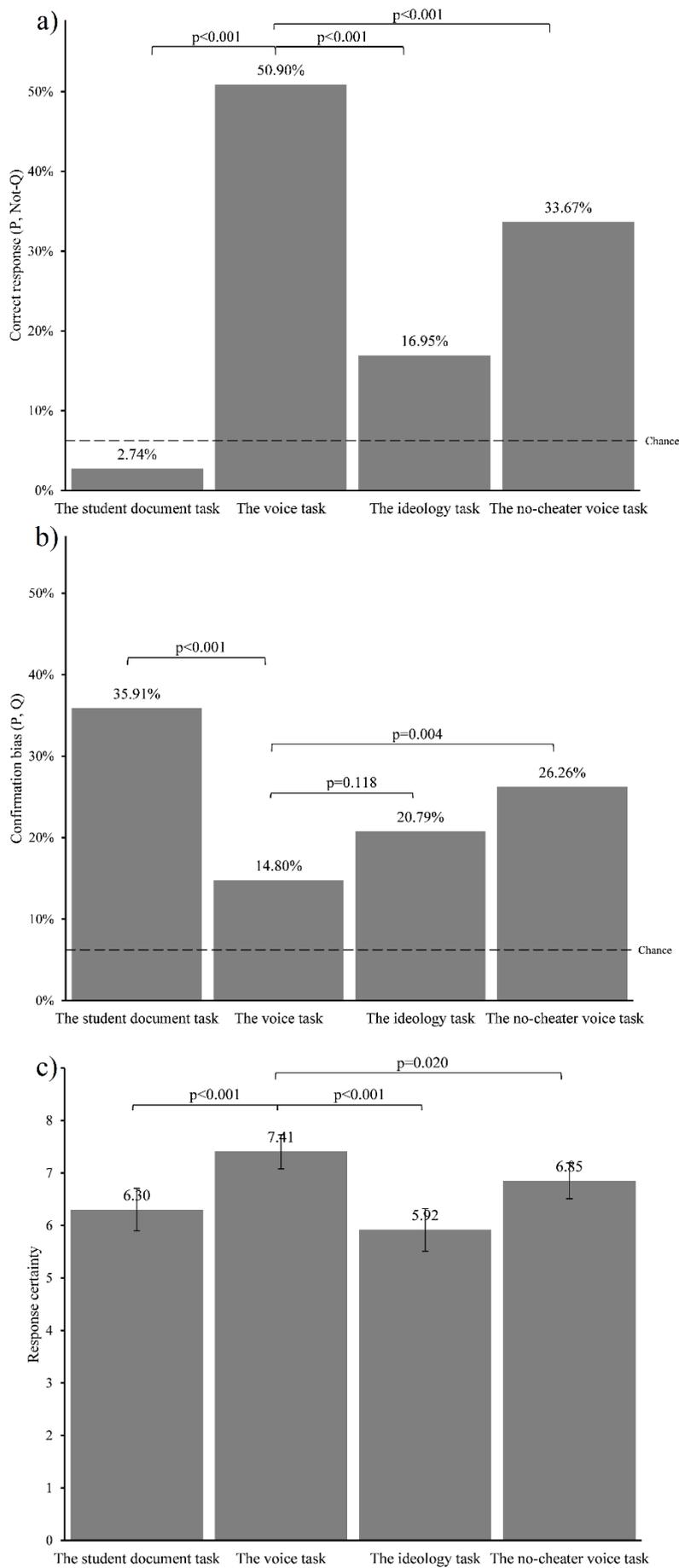
Note: T-tests are two-sided. Error bars around means represent 95% confidence intervals. n = 818.

Figure 4: Percentage correct P, Not-Q responses (panel a) and incorrect P, Q responses (panel b), and response certainty (panel c) across the four Wason selection tasks (Study 2).

Note: Predicted probabilities are extracted from Models I and II, Table III in the online appendix. 95% confidence intervals. n = 393.

Figure 5: Predicted probabilities of correct response (P, Not-Q) on the voice task and the ideology task by levels of political sophistication.