CENTRE FOR DECISION RESEARCH & EXPERIMENTAL ECONOMICS

University of Nottingham
UK | CHINA | MALAYSIA

Discussion Paper No. 2026-01

Dominik Suri, Simon Gächter,
Sebastian Kube and
Johannes Schultz

January 2026

The resilience of rule compliance in a polarized society

CeDEx Discussion Paper Series
ISSN 1749 - 3293

CENTRE FOR DECISION RESEARCH & EXPERIMENTAL ECONOMICS

The Centre for Decision Research and Experimental Economics was founded in 2000, and is based in the School of Economics at the University of Nottingham.

The focus for the Centre is research into individual and strategic decision-making using a combination of theoretical and experimental methods. On the theory side, members of the Centre investigate individual choice under uncertainty, cooperative and non-cooperative game theory, as well as theories of psychology, bounded rationality and evolutionary game theory. Members of the Centre have applied experimental methods in the fields of public economics, individual choice under risk and uncertainty, strategic interaction, and the performance of auctions, markets and other economic institutions. Much of the Centre's research involves collaborative projects with researchers from other departments in the UK and overseas.

Please visit http://www.nottingham.ac.uk/cedex for more information about the Centre or contact

Samantha Stapleford-Allen
Centre for Decision Research and Experimental Economics
School of Economics
University of Nottingham
University Park
Nottingham
NG7 2RD
Tel: +44 (0)115 74 86214
Samantha.Stapleford-Allen@nottingham.ac.uk

The full list of CeDEx Discussion Papers is available at

http://www.nottingham.ac.uk/cedex/publications/discussion-papers/index.aspx

The resilience of rule compliance in a polarized society

Dominik Suri Simon Gächter Sebastian Kube Johannes Schultz

January 2026

Abstract: Democratic societies depend on citizens following rules even when those rules are set by political opponents. Rising polarization may threaten this behavior. We test the impact of polarization on rule compliance in the United States across three pre-registered waves (May and November 2024; April 2025; n = 8,340) using the "coins task", which is a non-political, generic rule-following task, where breaking the rule increases payoffs. Participants were randomly assigned to follow rules set by the experimenter, a political co-partisan, a political opponent, or a non-partisan US citizen. Rule compliance ranged from 52.3% to 57.8%, and equivalence testing indicates no meaningful differences across waves or partisan rule-setter identities. However, greater affective distance from partisan rule setters is associated with lower compliance and weaker descriptive and normative beliefs about rule-following. These findings suggest that rule compliance is resilient to the rule-setter's identity. While affective polarization may erode this behavior somewhat, substantial compliance remains: the human tendency to follow rules, even when incentivized to break them, survives the "stress test" of partisan rule-setting in highly polarized times.

JEL classification codes: C91, D72, D91, Z13.

Keywords: Political polarization, affective polarization, rule-following, coins task, norms, online experiments, political identity, equivalence testing, replication.

Contact: Suri (corresponding author): Institute for Food and Resource Economics & Center for Economics and Neuroscience, University of Bonn, dsuri@uni-bonn.de. Gächter: Centre for Decision Research and Experimental Economics, University of Nottingham & Institute of Labour Economics, Bonn & CESifo, Munich, Simon.Gaechter@nottingham.ac.uk. Kube: Institute for Applied Microeconomics & Center for Economics and Neuroscience, University of Bonn, kube@uni-bonn.de. Schultz: Center for Economics and Neuroscience & Institute for Experimental Epileptology and Cognition Research, University of Bonn, JohannesSchultz@uni-bonn.de. **Acknowledgments:** We thank Anto Marcinkovic and Emir Kavukcu for the excellent research assistance. We are grateful for comments from Gwen-Jiro Clochard, Eugen Dimant, Lata Gangadharan, Ana Gantman, Holger Gerhardt, Angela Jiang-Wang, Erik O. Kimbrough, Julia Seither, Chris Starmer, Marie Claire Villeval, and seminar audiences at CeDEx (University of Nottingham) and CENs (University of Bonn), as well as from participants at the 2025 MIT Polarization Workshop, the Second Annual Chicago School in Experimental Economics, and the 2025 Workshop on Behavioral and Applied Economics at Technical University of Munich. **Funding:** Our research was funded by the European Research Council [grant number ERC-AdG 101020453 PRINCIPLES] and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy [EXC 2126/1-390838866]. **Ethics approval:** Our research received ethical approval from the School of Economics Research Ethics Committee at the University of Nottingham (protocols ERCP-2024-029, ERCP-2024-037 and ERCP-2025-040). All participants provided informed consent. **Research transparency:** The main research hypotheses, survey design and sampling approach were pre-registered on AsPredicted: https://aspredicted.org/2wvt-r3by.pdf, https://aspredicted.org/tsmz-wyq7.pdf, https://aspredicted.org/r64m-ksdk.pdf. Data and analysis code will be made available on OSF.

1 Introduction

Rules in form of laws, norms, regulations, and guidelines steer social life in myriad of ways by placing demands on people's behavior—"Do *x*!", "Don't do *y*!" (Posner, 2000; Gelfand, 2019; Daston, 2022). They are "the grammar of society" (Bicchieri, 2006) and the foundation of social order (e.g., Hobbes, 1651/1996; Weber, 1922/1978; Elster, 1989; Kliemt, 2020). In this paper, we ask whether compliance with rules depends not only on the substance of a rule and the incentives to comply with it, but also on the identity of the authority imposing it.

Polarized contexts provide a stark and natural test case for the role of rule-setter identity for rule compliance. Democratic societies rely on citizens following rules, even when those rules are set by political opponents. Citizens may respond differently to rules issued by political opponents than to those set by co-partisans, even when the rules themselves are substantively identical. Beyond ideological disagreement, affective considerations—such as feelings of social closeness to co-partisans and aversion toward political opponents—may further shape rule-following behavior because ideological and affective polarization are increasingly salient and rising in Western democracies (see, e.g., Guriev & Papaioannou, 2022; Puryear et al., 2024).

In this paper, we study the role of rule-setter identity for voluntary rule compliance in a highly polarized society: the United States of America. To keep things simple, and to isolate the role of rule-setter identity, we employ an experimental approach that allows us to focus directly on the role of the political identity of the rule setter for voluntary compliance with a non-political, decontextualized rule. We study compliance with an abstract rule because it avoids the confound of rule-setter identity and rule content of which there are infinitely many possibilities given the large number of rules that regulate social and political life.

In our setting, compliance with the rule is costly, thereby militating for breaking the rule. Rule compliance is also anonymous and not enforced—making it voluntary—, and has no consequences for the rule setter or anyone else. This setting implies that there are no conventional reasons (ideological conformance, incentives, social preferences) to comply with the rule. However, previous research in this setup (e.g., Kimbrough & Vostroknutov, 2016, 2018; Molleman et al., 2023; Gächter et al., 2025a; Suri et al., 2025) suggests that between 58% and 65% followed an abstract rule anyway. In these experiments, the rules were set by the experimenter. Yet, as argued above, in many naturally-occurring situations, people are asked to follow rules that are set by people who may or may not share their identity with the rule follower. Ideological and affective polarization provide a stark case for a potential identity conflict. Hence our question: Does the origin of an otherwise identical rule—whether set by a co-partisan or a political opponent—affect rule compliance?

In our rule-following task—the "coins task" (Suri et al., 2025, described in detail in Section 3.1)—participants are given 20 coins. Every second, one coin disappears, but participants can stop this disappearance at any time. A non-enforced rule instructs them to wait for a corresponding signal before stopping. Control questions ensure that participants understand that neither following nor breaking the rule produces external consequences for others; it only affects their own payoff. Conceptually, the coins task builds on the traffic light task introduced by Kimbrough and Vostroknutov (2016) and the simplified version by Gächter et al. (2025a). These papers demonstrate that generic, non-political rule-following tasks like our

coins task have desirable behavioral properties that make it an useful tool to investigate rule compliance. Unlike in previous literature, where the experimenter sets the rule, in our experiments the rule is set by a co-partisan, a political opponent, or a stranger where partisanship is not disclosed. For reasons of comparability, we also have settings where the experimenter sets the rule.

Our deliberately simplified setting allows us to isolate the importance of rule-setter identity for rule compliance in the context of intense ideological and affective polarization. Beyond studying the behavioral consequences of ideological and affective polarization, we believe our study also advances the emerging literature on rule-following, which suggests that, in settings with no incentives to follow the rule and no negative consequences of rule breaking for others, people follow rules out of respect for the rule and conformity with social expectations (Gächter et al., 2025a). How does the respect for rules and conformity with social expectations change when people don't share the rule-setter's political identity as compared to when they do? What is the role of perceived social closeness to the rule setter for rule compliance?

We ran our pre-registered rule-following study with 8,340 US Americans from across all regions of the United States (average age 40.3 years, 49.2% males, between-subject design) on the online platform Prolific Academic during a time when polarization was particularly salient: the 2024 US presidential elections which took place on November 5, 2024. We conducted our study in three main phases with different participants in each phase: the Pre-election Phase (May 2024, n=1,727), the Election Phase (one week before, n=2,291, and one week after the election, n=2,249), and the Post-Inauguration Phase (April 2025, n=2,073).

We find that political polarization in each of the phases is pronounced: As expected, Democrats overwhelmingly identify with liberal policy positions, while Republicans align with conservative positions. At the same time, participants report substantial affective polarization as measured by people's subjective distance from their political opponents, alongside relative closeness to their co-partisans. These gaps remain consistent across all three phases, and the difference in subjective closeness actually widens around the election.[1]

This polarized environment provides a useful setting for stress-testing how rule compliance is influenced by the political identity of the rule setter for two reasons. First, individuals' ideological commitments, voting behavior, and moral values are highly structured along partisan lines (see, e.g., Enke, 2020). Second, political opponents are perceived as a morally and socially distinct out-group (Puryear et al., 2024), generating affective distance and, in some cases, intra-group tensions. Together, these features may create a form of "moral wiggle room" (Dana et al., 2007), allowing individuals to justify violating a costly rule when it originates from a political opponent, even if the rule itself is identical to one set by a co-partisan.[2]

The manipulation of the rule-setter's identity serves as our main treatment variable across all three phases. In the Co-partisan condition, the rule setter shares the participant's party preference (self-identified Democrat or Republican). In the Political Opponent condition, the rule setter belongs to the opposing party. As a benchmark, we also include the standard treatment employed in prior rule compliance studies (e.g., Kimbrough & Vostroknutov, 2016, 2018; Karakostas & Zizzo, 2016; Desmet & Engel, 2021; Molleman et al., 2023; Bicchieri et al.,

[1] These findings are consistent with a large and growing literature on the polarization of US society (see, e.g., Huddy et al., 2015; Iyengar et al., 2019; Mason, 2023; Fasching et al., 2024).

[2] See Bicchieri et al. (2022) for the role of social proximity in norm erosion.

2025; Gächter et al., 2025a; Hoyer et al., 2025; Suri et al., 2025), in which the EXPERIMENTER sets the rule.[3]

Our main results on rule-following are as follows: First, rule-following rates when the rule is set by the experimenter are 57.8% across the three phases and similar in each of them. These rule-following rates are also comparable to those observed in related literature. For instance, Kimbrough and Vostroknutov found rule-following of 62.5% (Kimbrough & Vostroknutov, 2016) and 58.5% (Kimbrough & Vostroknutov, 2018). Molleman et al. (2023), Gächter et al. (2025a) and Suri et al. (2025) report average rule-following rates of 58% to 65% in their most comparable experiments. Kimbrough et al. (2024) conducted a rule-following experiment in two waves one month apart with the same individuals and found individual rule-following rates of 60.8% and 62.2%. Our result that rule-following rates are stable across phases is consistent with their finding. Thus, our baseline, with the experimenter as rule setter, reveals results similar to those reported previously.

Rule-following rates are also stable over the three phases when the rule is set by a copartisan (average rule-following rate is 52.6%) and when the rule is set by a political opponent (average is 52.3%). Using equivalence tests that assess the size of a null result (see, e.g., Lakens et al., 2018), we show that these treatment differences are not only statistically insignificant but also behaviorally small— suggesting that rule-following behavior is strikingly robust with respect to both time and rule-setter identity.

A somewhat different picture emerges when we consider the *subjective social closeness* to the rule setter (measured using the "Inclusion of the Other in the Self" task from social psychology (Aron et al., 1992; Baader et al., 2024)). Here we find significantly lower rule-following rates when participants feel distant to the rule setter than when they feel close (50.3% vs. 55.7%); this holds irrespective of the political identity of the rule setter. Moreover, perceived social distance is associated with lower general rule-following attitudes, and descriptive as well as normative beliefs. These findings are consistent with research showing that social closeness predicts behavioral outcomes in games of coordination (e.g., Gächter et al., 2022, 2025b). Furthermore, using random assignment to naturally-occurring groups, Tufano et al. (mimeo) show that social closeness, measured the same way as we do here, also has a causal impact on cooperation, which indicates that perceived social closeness likely has a causal influence on rule compliance as well.

Taken together, our results suggest that rule-following is negatively related to the social distance to the rule setter. Nevertheless, overall compliance remains high in our stress-testing environment of high polarization: even when participants interact with a political opponent to whom they feel socially distant, rule compliance persists at around 48%. This suggests that voluntary rule compliance is a remarkably resilient behavioral phenomenon.

The remainder of the paper is structured as follows. Section 2 discusses the related literature and our contributions. This is followed by Section 3 which outlines the experimental design and sampling procedures. We then present the results in Section 4. Section 5 provides a concluding discussion.

[3] Note that we implemented a fourth treatment condition in the Election and Post-inauguration Phases—called STRANGER—where we reveal the rule setter to be a fellow US Prolific user but do not disclose her political identity.

2 Related literature and our contributions

Our paper contributes to four strands of literature. The first concerns experimental studies of rule-following behavior, which typically use simple, abstract tasks to assess whether individuals comply with prescribed rules (e.g., Kimbrough & Vostroknutov, 2016, 2018; Karakostas & Zizzo, 2016; Desmet & Engel, 2021; Molleman et al., 2023; Kimbrough et al., 2024; Bicchieri et al., 2025; Gächter et al., 2025a; Hoyer et al., 2025; Suri et al., 2025). All previous studies use rules set by the experimenter. Our main contribution is to examine the role of the rule setter's identity in a context where it is highly salient: partisan differences during the 2024 US election cycle. Because, as we will show in Section 4.1, ideological and affective polarization is very strong, having highly polarized partisan rule setters creates a strong "stress-testing" environment (e.g., Smith, 1994; Croson & Gächter, 2010) for the robustness of rule-compliant behavior.

Our experiments do not only observe whether rule-setter identity matters for rule compliance but we also investigate how rule-setter identity is related to social expectations (normative and descriptive beliefs about rule compliance). In doing so, we build on Gächter et al. (2025a) who developed the interdisciplinary CRISP framework that captures four motives involved in rule conformity, all of which are supported by experimental evidence reported in Gächter et al. (2025a): rule compliance (C) generally depends on intrinsic respect for rules (R), extrinsic incentives (I), conformity with social expectations (S), and social preferences (P). Using CRISP language, we study rule compliance under stark conditions in which incentives favor violating the rule ($I < 0$) and social preferences cannot play a role ($P = 0$), so that only intrinsic respect for rules (R) and conformity with social expectations (S) can influence behavior. We therefore elicit measures of R and S in our experiments.

A second contribution is to the literature on the replicability of experimental results. Replications are important to establish robust knowledge (e.g., Maniadis et al., 2014; Camerer et al., 2016; Dreber & Johannesson, 2025). Our experiments replicate elements of the CRISP framework as reported in Gächter et al. (2025a)—most importantly the rate of rule compliance (C) and social expectations (S). Our replications (i) use a slightly different population than Gächter et al. (2025a) (who ran experiments on MTurk with US Americans) whereas our participants are registered on the Prolific platform (but are US Americans as well); (ii) use the abstract coins task, whereas Kimbrough and Vostroknutov (2016) and Gächter et al. (2025a) used a more naturalistic traffic light task; (iii) are conducted in three separate waves around politically turbulent time points in the 2024 US election cycle, thereby offering evidence on the temporal replicability of our results; and (iv) most importantly for our research purposes, use rule setters that are not the experimenter, unlike all previous research. We argue, therefore, that our "conceptual replication" (Dreber & Johannesson, 2025) makes a useful—and to our knowledge a first—contribution about the robustness of basic rule compliance to the fledgling literature on rule compliance. The fact that the partisan identity of the rule setter does not matter for rule compliance (and for the associated social expectations) is an important piece of evidence for the fundamental human tendency to comply with rules even in the absence of incentives and of social-preference-based motives.

The third strand of literature to which we contribute is on the behavioral consequences of polarization (see, e.g., Dimant & Kimbrough, 2024). Here, we add evidence on rule-following

behavior. Most closely related to our paper are the studies by Iyengar and Westwood (2015), Dimant (2024), Feldhaus et al. (2024), and Freundt and Herz (2024). They find that political polarization affects delegation decisions, re-claiming decision rights, fairness preferences in a dictator game, or scholarship selection. We complement these findings by showing that the political identity of the rule setter has no effect on behavior in our context: compliance rates do not differ significantly whether the rule originates from a co-partisan or a political opponent.

While all these studies primarily focus on ideological polarization, either using a Trump lover/Trump hater or Democrats/Republican framing, we additionally explore to what extent affective relations toward other members within a polarized society matters in our context. Here we find that subjective distance to the rule setter actually decreases rule compliance. Moreover, in these studies the participant's behavior immediately affects co-partisans or opposing partisans. We, however, focus on behavior that only affects the participant's own payoff under a given rule, varying the partisanship of the rule setter.

The fourth and final strand of literature to which we contribute focuses on measures of affective polarization. The standard measure of affective polarization is the "feeling thermometer", which was institutionalized by the American National Election Studies (ANES). This instrument asks respondents to rate their feelings toward Democrats and Republicans on a scale from 0 (cold) to 100 (warm). Using this measure, numerous studies have documented that affective polarization in the US peaks around elections (see, e.g., Sood & Iyengar, 2016; Gidron et al., 2020; Garzia & da Silva, 2022; Martin & Nai, 2024). However, this measure primarily captures partisan dislike and does not directly target one of the core problems of affective polarization: the perceived social distance between partisans. Because social distance is considered "a less obtrusive measure of partisan affect" (Iyengar et al., 2019, p. 132), we employ the "Inclusion of Other in the Self" (IOS) scale from social psychology (Aron et al., 1992; Baader et al., 2024). The IOS scale measures perceived social closeness toward both co-partisans and political opponents, allowing us to capture heterogeneity in social distance that goes beyond ideological differences. In this way, the IOS scale complements existing measures and provides a more nuanced proxy for affective polarization. Using this scale, we find that affective polarization is most pronounced around elections, highlighting not only general partisan animosity but also variation in the social distance people feel toward others.

3 Experimental design and procedures

All experiments received approval from the Research Ethics Committee in the School of Economics at the University of Nottingham (protocols ERCP-2024-029, ERCP-2024-037 and ERCP-2025-040). Our hypotheses (see Section 4), experimental design, and procedures were pre-registered on AsPredicted.org[4]. Detailed screenshots of our experiments are displayed in the Supplementary Information (SI).

[4]As we explain in Section 3.5, we sampled our data in three phases and pre-registered each one of them: Pre-election Phase (May 2024, https://aspredicted.org/2wvt-r3by.pdf), Election Phase (November 2024, https://aspredicted.org/tsmz-wyq7.pdf) and Post-inauguration Phase (April 2025, https://aspredicted.org/r64m-ksdk.pdf).

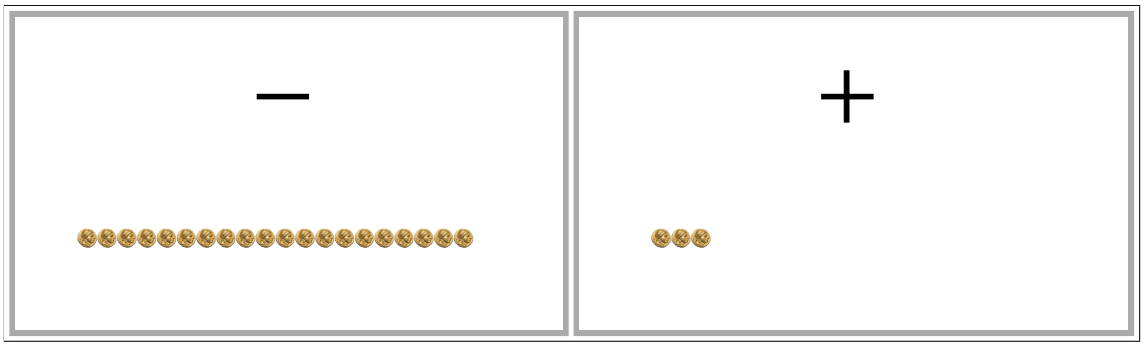


Figure 1: The coins task to measure rule compliance. The figure shows a screenshot of the coins task at two different time points: The left image shows the starting position. The sign is "minus" and there are 20 coins available. The right image shows the situation after 17 seconds have passed. The sign is "plus" and there are 3 coins left. Within the coins task, the following *non-enforced rule* is implemented: *"Press the space bar after the sign has changed from 'minus' to 'plus'."* The time change occurs after 12 seconds, which is not disclosed to the participants.

3.1 The coins task to measure rule compliance

A main goal of our experiment is to have a clean behavioral measure of rule compliance. To do so, we use the "coins task" (Suri et al., 2025). The coins task is inspired by Kimbrough and Vostroknutov (2016) and is a more abstract task than the traffic light tasks, which they and Gächter et al. (2025a) used, with the advantage that the costliness of complying with the rule is very salient. The coins task works as follows: The participant is presented with 20 coins that are displayed on her screen (see left panel of Figure 1). Every second, one coin disappears. The participant can press the space bar on her keyboard at any time—i.e., when 20, 19, …, 1, or 0 coins remain—to stop the coins from disappearing. Once the space bar is pressed, the participant receives the monetary equivalent of £0.1 per remaining coin and must wait until the remaining seconds have passed before she can continue. The participant is explicitly told that she can press the space bar at any time, which means that she can get a maximum of 20 coins and a minimum of 0 coins.

Additionally, a sign is displayed on the decision screen: a "minus" symbol that changes to a "plus" after 12 seconds (i.e., after 12 coins have disappeared). The participant only knows that the sign will change after some time, but not the exact number of seconds. Furthermore, following Kimbrough and Vostroknutov (2016), Gächter et al. (2025a), and Suri et al. (2025), a rule regarding the sign change is implemented as follows: *"There is a rule for this. The rule is: Press the space bar after the sign has changed from 'minus' to 'plus'."* Participants therefore have a clear choice between stopping the disappearance of the coins immediately (breaking the rule) or waiting for the sign change before stopping the disappearance (following the rule). Importantly, following or breaking the rule has no monetary consequences beyond the number of coins that the participant will have left on her screen.

Prior to starting the task, participants see a graphical representation of the process (see Figure 1) and answer two comprehension questions, which must be answered correctly to proceed.[5] For each of the two situations presented in Figure 1, we ask how many coins the par-

[5] In the conditions Co-partisan and Political Opponent (see Section 3.4), we implemented a third comprehension question where the participant should indicate whether the rule was set by a US Prolific user who identifies as a Democrat or Republican to increase the salience of our treatment manipulations.

ticipant would receive if she pressed the space bar at that time point. With this approach, we want to emphasize (without explicitly stating it) that rule-breaking carries no negative monetary consequences; thus making the availability of the choice between breaking the rule and following it salient. Participants play the coins task exactly once.

3.2 Measuring political polarization

Two distinct social identities are political ideology (Iyengar et al., 2019; Huddy et al., 2015), i.e., a set of beliefs and values about politics and the society (Mason, 2023), and partisanship, i.e., a psychological attachment to a political party (Campbell et al., 1980). In the US context, for many people, political ideology is either liberal or conservative and partisanship is with either the Republican Party or the Democratic Party.

As these two identities become more entrenched, political polarization emerges as a prominent feature of political behavior. Political polarization can be divided into ideological polarization and affective polarization (Mason, 2023). The former describes an increase in the extremes of liberalism and conservatism (Mason, 2023), while the latter involves an intensification of distrust, aversion and animosity toward the opposing party (Campbell et al., 1980; Fasching et al., 2024), to a point where "partisans should be averse to entering into close interpersonal relations with their opponents" (Iyengar et al., 2019, p. 132).

To investigate *ideological polarization*, we follow Balliet et al. (2018) and use the following three items to measure political ideology: (i) agreement with the statement "When it comes to politics, I consider myself politically conservative." (1 = strongly agree, 7 = strongly disagree), (ii) agreement with the statement "When it comes to politics, I consider myself politically liberal." (1 = strongly agree, 7 = strongly disagree) and (iii) self-placement on a 7-point Likert scale ranging from "extremely left" (1) to "extremely right" (7). The first item is reverse-coded to align it with the latter two. As proposed by Balliet et al. (2018), we then construct a composite political ideology index by averaging the three items, so that lower values consistently indicate more liberal political orientations that are typically attributed to the Democratic Party; and higher values consistently indicate more conservative political orientations that are typically attributed to the Republican Party.

To assess *affective polarization*, we use the "Inclusion of Other in the Self" (IOS11) scale (Baader et al., 2024) which provides a robust and validated measure of subjective closeness toward someone else (Aron et al., 1992; Gächter et al., 2015; Baader et al., 2024)—in our case to someone from the Republican Party or the Democratic Party.[6] The IOS11 scale consists of a pair of circles. The degree of overlap can be varied by the participant on an 11-point Likert scale from not overlapping (1) to overlapping almost completely (11). Hereby, one of the circles represents the respondent, while the other one represents another group of people.

In our study, participants could determine the degree of overlap between the two circles by moving a slider on the screen. They received the instructions that they should interpret the degree of overlap as representing the relationship between themselves and "X", with "X" serving as a placeholder for US Prolific users who define themselves either as Republicans

[6] We deviate from the prominent workhorse measure for affective polarization, i.e., the feeling thermometer, because Iyengar et al. (2019, p. 132) point out that "[a] less obtrusive measure of partisan affect is social distance, the extent to which individuals feel comfortable interacting with out-group members in various settings. If partisanship is an important social identity in its own right, partisans should be averse to entering into close interpersonal relations with their opponents."

(R), Democrats (D) or Stranger (S, i.e., without disclosing party affiliation). Depending on the phase (introduced below in Section 3.5), participants either rated their subjective closeness toward both Democrats and Republicans (Pre-election Phase), or in addition to that also toward a Stranger (Election and Post-inauguration Phase). For an illustration, see Section SI-3.3.

For constructing a measure of affective polarization, we calculated the following difference in the responses of each participant by making use of their self-reported political party: *(IOS11 to co-partisan − IOS11 to political opponent)*, where co-partisan is someone who affiliates with the same political party as the participant, while political opponent is someone who affiliates with the opposing political party. The resulting index ranges from 10 (maximal subjective distance between someone from the same to the other party, i.e., I feel very close to a co-partisan and very distant to a political opponent) to -10 (maximal subjective distance between someone from the other to the same party, i.e., I feel very distant to a co-partisan and very close to a political opponent).

3.3 Eliciting social expectations and respect for rules

Social expectations are central for rule conformity (Cialdini & Goldstein, 2004; Bicchieri, 2006; Bicchieri & Xiao, 2009; Charness et al., 2025). Consistent with this insight, social expectations S also play a central role in the CRISP framework that Gächter et al. (2025a) developed to explain rule compliance. To investigate the role of social expectations (S), we collect two different forms of beliefs: normative and descriptive beliefs about rule compliance. The former reflects what participants believe is socially appropriate in the coins task, whereas the latter captures expectations about how others actually behaved. For both instances, we follow the wording in Gächter et al. (2025a). Specifically, we assess the second-order normative belief of rule compliance (resp. rule violation), where we asked participants to first assume that a US Prolific user had (not) followed the rule in the coins task and then to report how socially appropriate she thinks other US Prolific users would view this. We incentivized the normative belief elicitation following the coordination-based approach in Krupka and Weber (2013). The descriptive belief asked for the empirical expectation of how many participants have followed the rule. Here accuracy was incentivized using the quadratic scoring rule (Selten, 1998). In addition to these beliefs, we also elicited the general rule-following attitude by asking the participant what she thinks one should do in our coins task on a 4-point Likert scale (never (1), rarely (2), often (3) or always (4) follow the rule). This is a proxy for respect for rules (R) in the CRISP framework.

3.4 Treatment conditions: The political identity of the rule setter

Our experiment employs a between-subject design, consisting of three different origins of the rule. The rule is set either by (i) the experimenter as, for example, in Kimbrough and Vostroknutov (2016), Gächter et al. (2025a) or Suri et al. (2025); (ii) a US Prolific user who identifies as a Democrat; or (iii) who identifies as a Republican. Specifically, we tell the participant the following: "Note: The rule was chosen by a Prolific user from the USA who defines themselves as Democrat [Republican]."[7] We make use of the origin of the rule as well as the participant's

[7] Prior to Pre-election Phase, we ran a small pilot (n=291) to test for the duration and to ask participants if they would set the rule mentioned in Section 3.1. Out of those, 24.8% of Democrats and 40% of Republicans indicated that they would implement this rule.

self-reported political party affiliation to define our pre-registered exogenous treatment conditions as follows:

- EXPERIMENTER: No origin of the rule is disclosed.
- CO-PARTISAN: The rule is set by a US Prolific user who has the same party affiliation than the participant (e.g. the participant is a Democrat and the rule is set by a Democrat).
- POLITICAL OPPONENT: The rule is set by a US Prolific user who has the opposing party affiliation than the participant (e.g. the participant is a Democrat and the rule is set by a Republican).

In addition to these three treatments, in some waves we also had an additional treatment STRANGER where we provided the information that the rule setter is a US Prolific user but we did not indicate her partisanship. With this treatment, we can distinguish the impact coming from a common identity (being a US Prolific user) and a political identity (being a Democrat or Republican). More details will be provided in Section 4.3 below.

3.5 Procedures and timelines

The experiment was programmed using the o-Tree software (Chen et al., 2016). US participants were sampled on Prolific Academic in four waves, which differ with respect to the proximity to the 2024 US presidential election (that took place on November 5, 2024): May 21, 2024 (Wave 1, referred to as Pre-election Phase), October 29 and November 12, 2024 (Waves 2 and 3, referred to as Election Phase[8]) as well as April 1, 2025 (Wave 4, referred to as Post-inauguration Phase).

In each wave, we applied the following pre-registered selection criteria and implemented a balancing quota for political party affiliation and gender: at least 18 years old, native English speaker, citizen of the United States, currently living in the United States, not participated in a similar study by the authors of this study before and submitting consent to participate. The experiment was scheduled for 10 minutes.

After providing informed consent, participants conducted the coins task as outlined in Section 3.1. We then elicited the general rule-following attitude, followed by the descriptive and normative beliefs (see Section 3.3). Then, participants answered the "Inclusion of Other in the Self" (IOS11) scale (Baader et al., 2024, see Section 3.2). The experiment concluded with a brief socio-demographic questionnaire. Participants received a fixed participation fee of £1.50 and a bonus payment of up to £2.00. For determining the bonus payment, either the coins task or the belief elicitation was randomly selected and paid.

3.6 Our participant pool

We pre-registered a sample size of 600 participants for each treatment condition and wave. After application of our pre-registered exclusion criteria, we end up with sample sizes of n=1,727 in the Pre-election Phase, n=4,540 in the Election Phase (consisting of Waves 2 and 3) and

[8]For the sake of clarity, we pool Wave 2 and 3 data as there are no significant differences between both waves, neither w.r.t. political polarization measures and subjective closeness in general (see Table SI-2), nor do we find substantial differences in rule compliance (see Table SI-4).

Table 1: Overview of participants' characteristic

Characteristic	Pre-election Phase	Election Phase	Post-inauguration Phase	Full sample	p-value
Sample size	1,727	4,540	2,073	8,340	
Male	50%	49%	49%	49%	0.86
Democrat	51%	51%	53%	51%	0.24
Age: 18 to 34	34%	42%	43%	41%	<0.001
Age: 35 to 54	46%	41%	40%	42%	<0.001
Age: 55 and above	20%	16%	17%	17%	0.007
Northeast	19%	18%	19%	18%	0.60
Midwest	21%	19%	17%	19%	0.004
South	40%	43%	43%	42%	0.09
West	21%	20%	22%	21%	0.24

Notes: The table reports means for individual characteristics for the three phases and the full sample. The p-values are obtained from Pearson's chi-squared tests.

n=2,073 in the Post-inauguration Phase. Table 1 displays mean socio-demographic characteristics of each sample. Gender and political party affiliation are balanced across phases, as we invited participants accordingly. The samples mainly differ with regard to the age structure of our participants, i.e., on average younger US citizens participated in the Election Phase and Post-inauguration Phase, though there is still substantial variation. Furthermore, our participant pool includes individuals from all four major US regions—Northeast, Midwest, South, and West—–in roughly representative percentages.[9]

4 Results

In presenting our results, we will first examine in Section 4.1 the degree of political polarization in our Prolific sample of US citizens with respect to participants' (i) political ideology and (ii) subjective feelings of closeness toward both co-partisans and opposing partisans. As we will show, there is a large ideological gap between Democrats and Republicans already in the Pre-election Phase, and this gap remains stable across the Election Phase and Post-inauguration Phase. Likewise, we see that participants feel substantially distant from opposing partisans, and this affective gap also remains stable across all three phases. Having established a strong degree of ideological and affective polarization, in the next step we check in Section 4.2 if our treatment manipulations affect rule-following behavior.

Despite the strong political polarization, we find that rule compliance does not differ significantly across treatments. Regardless of whether the arbitrary rule is set by a co-partisan, an opposing partisan or the experimenter, compliance with the rule is remarkably high in all three treatments and remains stable across the three phases. This is not to say that the origin of the rule does not matter: we find in Section 4.3 that the subjective closeness to the rule setter does make a difference. Participants who feel more distant from the rule setter are significantly more likely to break the rule than when they feel close to the rule setter, independent

[9] US Census Bureau data (Decennial Census, 2020 DEC 118th Congressional District Summary File) indicates that 17.4% of the US population lives in the region Northeast, 20.8% in Midwest, 38.1% in South, and 23.7% in West.

of whether we control for the rule-setter's political orientation or not.

Consistent with this, we show in Section 4.4 that the same pattern exists in participants' normative and descriptive beliefs about rule-following and rule-breaking, and her general rule-following attitude. If the rule is set by a person that a participant reports feeling distant from, that participant (i) is more likely to believe that the rule should not be followed; (ii) expects fewer people to follow the rule; and (iii) is less likely to state that others consider rule-following as socially appropriate and rule-violation as socially inappropriate.

Lastly, we further investigate the null effect of our treatment conditions where we reveal the identity of the rule setter in Section 4.5. Using equivalence testing we show that the small and non-significant variations among Co-partisan, Political Opponent, and Stranger are behaviorally negligible.

All data analyses were conducted using the R software version 4.2.2 (R Core Team, 2022). Further figures and tables are displayed in Section SI-1 and Section SI-2 in the online Supplementary Information for explanatory and exploratory purposes. Throughout the paper and appendices, all reported p-values are based on two-sided hypothesis tests.

4.1 Ideological and affective polarization

To assess ideological polarization, we look at the gap in ideological intensity between Democrats and Republicans using the political ideology index by Balliet et al. (2018). Higher index values consistently indicate more conservative political orientations, while lower values consistently indicate more liberal orientations. Internal consistency of the index is high, with a Cronbach's alpha of 0.80, suggesting that these items reliably capture a common underlying ideological dimension.

Figure 2 shows the distribution of *ideological polarization* in the Pre-election Phase (top row, panel A), Election Phase (middle row, panel B) and Post-inauguration Phase (bottom row, panel C), separately for self-identified Democrats and Republicans. As one would expect, Democrats consistently report lower (i.e., more left-leaning, liberal) scores, while Republicans report higher (i.e., more right-leaning, conservative) scores. The difference in scores between Democrats and Republicans is significant in all three phases (Kruskal-Wallis tests, all p<.001). For Democrats, the average ideology score increases slightly over time, from a mean of 2.35 (Pre-election Phase) to 2.45 (Election Phase) and 2.73 (Post-inauguration Phase). Among Republicans, the average ideology score also shows a slight increase from 4.73 (Pre-election Phase) to 4.83 (Election Phase), followed by a modest decline to 4.75 (Post-inauguration Phase).

When examining the changes in ideological orientation across the three election phases among Democrats, there is a significant shift over time (p<.001, Kruskal–Wallis test across all 3 phases). This shift is also confirmed when we test for differences in the distributions using Kolmogorov–Smirnov tests (pairwise tests with Holm correction: Pre-election Phase vs. Election Phase: p =.34, Election Phase vs. Post-inauguration Phase: p =.006). Pairwise comparisons using Wilcoxon ranksum tests with Holm correction show no significant change from Pre-election Phase to Election Phase (+4.2%, p=.11), but a significant increase by 11.4% from Election Phase to Post-inauguration Phase (p<.001). In contrast, Republicans showed no significant change in ideological orientation across phases; neither when testing for changes across all three phases jointly (p=.30, Kruskal–Wallis test) nor in pairwise comparisons (Wilcoxon ranksum tests / Kolmogorov-Smirnov tests, both with Holm correction: Pre-election Phase

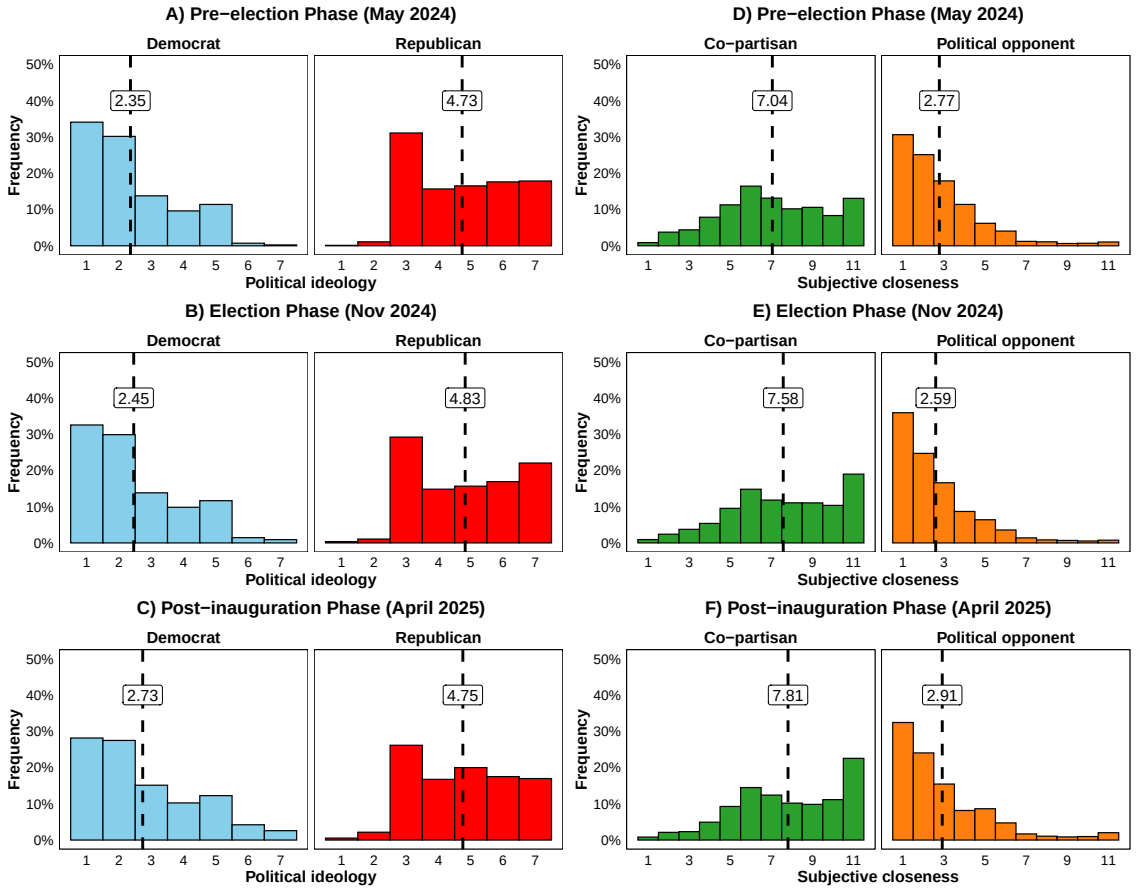


Figure 2: Political polarization across phases. The figure shows the distribution of ideological intensity (left) and subjective closeness (right) across phases (repeated cross-section). Ideological intensity is measured by the political ideology index (Balliet et al., 2018), where 1 means left/liberal and 7 means right/conservative. Subjective closeness is measured by the IOS11 score (Baader et al., 2024), where 1 means subjectively very distant and 11 means subjectively very close. Political ideology is shown separately for Democrats and Republicans, and subjective closeness is shown separately for co-partisans and opposing partisans. The dashed lines represent the respective means.

vs. Election Phase: +2.1%, p=.49 / p=.45, Election Phase vs. Post-inauguration Phase: -1.6%, p=.61 / p=.21), indicating ideological stability over time (see also the regression estimates in Table SI-3). The same holds when looking at the mean differences between Democrats and Republicans across phases: the ideological gap, on average, remains constant between Pre-election Phase and Election Phase, but closes slightly in the months after the 2024 US presidential election.

> ***Result 1a:*** *In all three phases, there is substantial ideological polarization, marked by consistently distinct political ideology scores for Democrats (liberal) and Republicans (conservative).*

With respect to *affective polarization*—i.e., the gap in psychological closeness toward a co-partisan and a political opponent—, we pre-registered that subjective closeness toward co-partisans and subjective distance toward political opponents will become more extreme with closer proximity to the US election (H2 in the pre-registration for phase 2).[10] To test the

[10] Similarly, we also hypothesized and pre-registered a difference between Election Phase and the Post-inauguration Phase (H2 in the pre-registration for phase 3).

hypothesis, we use our data from the IOS11 scale (Baader et al., 2024). Recall that participants use a slider to indicate perceived closeness to members of the same or opposing political party, with circle overlap representing subjective connection. As shown in the top (D), middle (E), and bottom (F) row in Figure 2, reported closeness to in-group members increased by 7.6% from the Pre-election Phase (mean of 7.04) to the Election Phase (7.58) and further by 3.0% after the inauguration (7.81), while out-group closeness showed a non-monotonic pattern, declining by 6.5% from Pre-election Phase (2.77) to Election Phase (2.59), but rebounding by +12.4% Post-inauguration Phase (2.91).

Statistical tests corroborate this temporal variation: For *in-group closeness*, a Kruskal–Wallis test finds differences across the three phases ($p<.001$), with all pairwise comparisons being significant in both Wilcoxon and Kolmogorov–Smirnov tests with Holm corrections (all $p<.01$). We find the same results separately for Democrats and Republicans (except for the comparison of Election vs. Post-inauguration Phase for Republicans where the Kruskal–Wallis and Kolmogorov-Smirnov tests with Holm correction yield $p>.28$). This indicates a gradual increase in in-group psychological attachment over the three phases.

In contrast, the *out-group closeness* is most pronounced around the time of the Election Phase. The Kruskal–Wallis test on the differences across the three phases yields significant differences ($p<.001$), and the pairwise comparisons show a significant decline from Pre-election Phase to Election Phase and increase again from Election Phase to Post-inauguration Phase (Wilcoxon and Kolmogorov–Smirnov tests with Holm corrections, all $p<.001$, with the exception of $p<.01$ for Pre-election Phase vs. Election Phase in the Kologorov-Smirnov test). The same pattern is visible for both Democrats and Republicans separately (Wilcoxon and Kolmogorov–Smirnov tests with Holm corrections, all $p<.026$, with the exception of $p=.07$ for Pre-election Phase vs. Election Phase in the Kologorov-Smirnov test for Republicans).

Looking more closely at the difference in psychological closeness toward co-partisans minus political opponents, Figure SI-1 displays the respective histograms. Average affective polarization significantly increases from the Pre-election Phase (mean of 4.28) to the Election Phase (4.98), but shrinks toward the Post-inauguration Phase (4.9); though the decrease is not statistically significant (Kologorov-Smirnov tests with Holm correction: Pre-election Phase vs. Election Phase: $p<.001$, Election Phase vs. Post-inauguration Phase: $p=.591$). Regression estimates further confirm these findings (see Table SI-3) that affective polarization was most pronounced around the time of the election, primarily driven by heightened in-group bonding and temporarily increased out-group distancing.

> ***Result 1b:*** *In all three phases, there is substantial affective polarization, marked in particular by a predominantly strong subjective distance to opposing partisans and a closeness to co-partisans.*

Overall, our data suggest very stable partisan self-identification over time across three key phases of the 2024 US presidential election cycle. There is strong polarization between co-partisans and political opponents, both in terms of ideological polarization and affective polarization. In the next subsections, we examine whether this polarized setting influences rule compliance in our coins task.

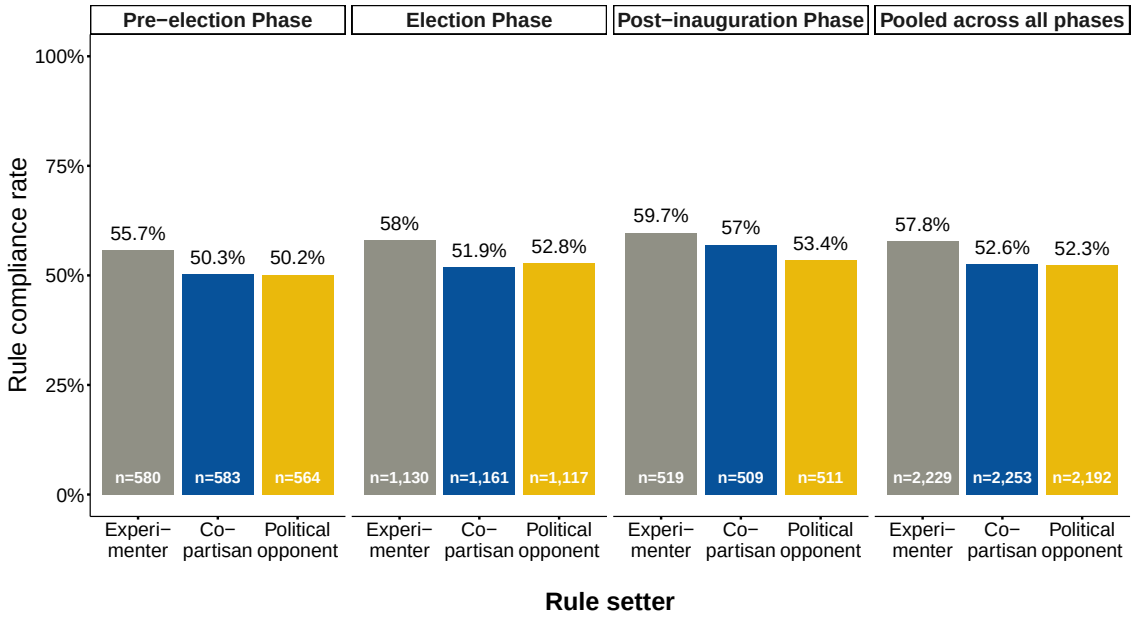


Figure 3: Rule compliance depending on identity of rule setter across phases. The figure shows the share of rule followers across the different treatment conditions and phases (repeated cross-section). A disaggregated overview of each wave is displayed in Table SI-4.

4.2 Rule compliance given party affiliation of the rule setter

To test whether rule-following is affected by our treatment manipulations of revealing the US party affiliation of the rule setter, we use data from our coins task. Recall that in this task, the rule is to prevent the displayed coins from disappearing by pressing the space bar only after a corresponding signal is given (see Figure 1). We therefore code rule-following as 1 if a participant stops the coins after the signal, and 0 if the space bar is pressed prior to the signal. Between treatments, we varied if the rule was set by the experimenter (Experimenter), a co-partisan (Co-partisan) or an political opponent (Political Opponent). We pre-registered that we expect rule compliance rates to be lowest when the rule is set by an opposing partisan, highest when it is set by a co-partisan, and in between these two rates when it is set by the experimenter (H1 in our pre-registrations: compliance(Political Opponent) $\leq$ compliance(Experimenter) $\leq$ compliance(Co-partisan)).

Figure 3 shows the rule compliance rates across the three rule-setter conditions and across the three election phases. Contrary to the hypothesized ordering, the observed compliance rates are only partially consistent with our directional expectations. In each phase, rule compliance in Co-partisan and Political Opponent are of comparable size, while rule compliance is always highest in Experimenter. However, Pearson's chi-squared tests that compare rule compliance rates for a given phase across the three treatment conditions are only significant in the Election Phase ($p<.01$ in Election Phase, $p=.099$ in Pre-election Phase, $p=.12$ in Post-inauguration Phase). In that phase, the pairwise Fisher's exact tests with Holm correction show a significant difference only for the difference between Experimenter and Co-partisan ($p=.034$), but not for the other two comparisons (Experimenter vs. Political Opponent: $p=.069$, Co-partisan vs. Political Opponent: $p=.62$).[11] In fact, all pairwise comparisons of

[11] Corresponding linear probability regression estimates yield qualitatively same results, even after including control variables (see Table SI-8). In that case, also the comparison between Experimenter and Political Opponent becomes significant.

Co-partisan vs. Political Opponent are insignificant, even when using non-adjusted Pearson's chi-squared test (Pre-election Phase p=1, Election Phase p=.67, Post-inauguration Phase p=.28).[12]

Taken together, rule compliance rates suggest that, despite the strong ideological polarization, the rules are followed even when those rules are set by political opponents. Even in the weeks directly surrounding the elections—when political ideology might be expected to matter most and ideological polarization might be most salient—rule compliance is high across all our treatments. Indeed, we found no significant differences across the three phases for a given treatment condition (Pearson's chi-squared test, p=.39 in Experimenter, p=.06 in Co-partisan, p=.49 in Political Opponent). Across all three phases, the average rule compliance rates are 57.8% in Experimenter, 52.6% in Co-partisan, and 52.3% in Political Opponent.[13] Moreover, zooming in more closely on differences in rule compliance between Democrats and Republicans, we find—except for a couple of instances—similar rule-following rates between participants of both party affiliations (see Table SI-6).

> ***Result 2:*** *Rule-following rates are high and robust: Irrespective of phase and rule-setting treatment, and despite the strong ideological polarization, 54.4% decide to follow the rule in the coins task.*

When it comes to the political identity of the rule setter per se, rule-following rates are remarkably robust and stable.[14] A second factor of partisan affect is affective polarization, which—as we have shown in Section 4.1—is pronounced in each of the phases as well. Thus, in the next section we turn to our measure of affective polarization, i.e., psychological closeness toward co-partisans and political opponent. As we will show, this actually seems to matter for rule-following behavior.

4.3 Rule compliance given subjective closeness toward the rule setter

Turning to the social proximity toward co-partisans and political opponents, a more nuanced pattern of rule compliance emerges. We again use data from our coins task, but now check whether the subjective distance to the rule setter matters for rule-following behavior. We pre-registered that we expected rule compliance to decrease if the rule setter is perceived to be socially more distant (H3 in our pre-registrations for the Election Phase and Post-inauguration Phase, resp. secondary analysis in the Pre-election Phase). To test this hypothesis, we first split our groups into a Close and a Distant subgroup by conducting (pre-registered) median splits according to the respective group's distribution of IOS11 scores (randomly allocating ties into either of the two, in a replicable manner).[15] Then, we compare rule compliance between the resulting Close and Distant subgroups.

Table 2 lists the corresponding rule compliance rates separately for each phase. Across the full sample, individuals in the Close subgroup consistently exhibited higher compliance

[12] An overview of all (pairwise) comparisons is displayed in Table SI-5 and the insignificant difference between Political Opponent and Co-partisan is discussed further in Section 4.5.

[13] Note that the difference in rule conformity of 5.2 to 5.5 percentage points between the treatment condition Experimenter and the conditions Co-Partisan respectively Political Opponent in the pooled sample are statistically significant (Pearson's chi-squared test, both p<.001). We discuss this further in Section 4.5.

[14] In Tables SI-8 and SI-9, we see that this also holds true when controlling for other factors that might potentially affect rule-following rates, since the coefficients for Co-partisan and Political Opponent in the estimated regression models are virtually indistinguishable.

[15] The replication package will provide comprehensive details regarding this procedure for the interested reader.

Table 2: Rule compliance w.r.t. subjective closeness to rule setter across phases

	Pre-election Phase	Election Phase	Post-inauguration Phase
Close	52.54%	54.68%	60.94%
Distant	47.39%	50.67%	51.51%
Co-partisan Close	58.70%	54.85%	58.33%
Co-partisan Distant	42.67%	49.28%	55.20%
Political Opponent Close	53.92%	56.08%	56.29%
Political Opponent Distant	44.83%	49.14%	49.28%
Stranger Close	-	56.17%	63.11%
Stranger Distant	-	51.66%	49.33%

Notes: The table reports mean rule compliance w.r.t. close and distant subjective closeness, using Median splits, across the different phases (repeated cross-section) and sub-groups. The Stranger treatment condition was not implemented in the Pre-election Phase. A disaggregated overivew of each wave is displayed in Table SI-4.

than those in the Distant subgroup, with differences becoming more pronounced over time (Pre-election Phase: 52.5% vs. 47.4%, Post-inauguration Phase: 60.9% vs. 51.5%).[16] Pearson's chi-squared tests indicate that these differences are statistically significant during the Election Phase (p=.02) and Post-inauguration Phase (p<.001), but insignificant in the Pre-election Phase (p=.094).[17]

Disaggregation by treatment reveals that this affective compliance gap holds across all rule-setter identities. This is clearly visible in Figure 4, which again shows rule compliance rates, but now pooled over all three phases. Compliance is consistently higher among Close participants compared to their Distant counterparts—whether the rule was set by a co-partisan (Co-partisan, close vs. distant: p<.001), an opposing partisan (Political Opponent, close vs. distant: p<.001), or an anonymous stranger (Stranger, close vs. distant: p=.002).

The latter treatment Stranger was only conducted in the Election Phase and Post-inauguration Phase. It follows the exact same procedures as Co-partisan and Political Opponent. The only change is that, when the rule setter is mentioned in the instructions, subjects are only informed that the rule was set by a US Prolific user, but they are not informed whether this user identifies herself as Democrat or Republican.[18] Correspondingly, the IOS11 score was also elicited toward a Prolific user for whom no political identification was provided. Using this score to compare rule compliance between close and distant Stranger thus allows us to see the mere impact of subjective distance toward the rule setter, devoid of any concerns about political ideology.

Regression analyses further corroborate how psychological closeness is correlated with

[16] Since we did only ask subjects for their social distance toward other Prolific users, but not toward the experimenter, we do not have IOS11 scores toward the rule setter in Experimenter. The analyses and numbers reported in this subsection thus exclude data from treatment Experimenter.

[17] An overview of all (pairwise) comparisons is displayed in Table SI-7.

[18] See also Table SI-1 for a detailed overview of all conditions and Section SI-3.1 for the exact wording.

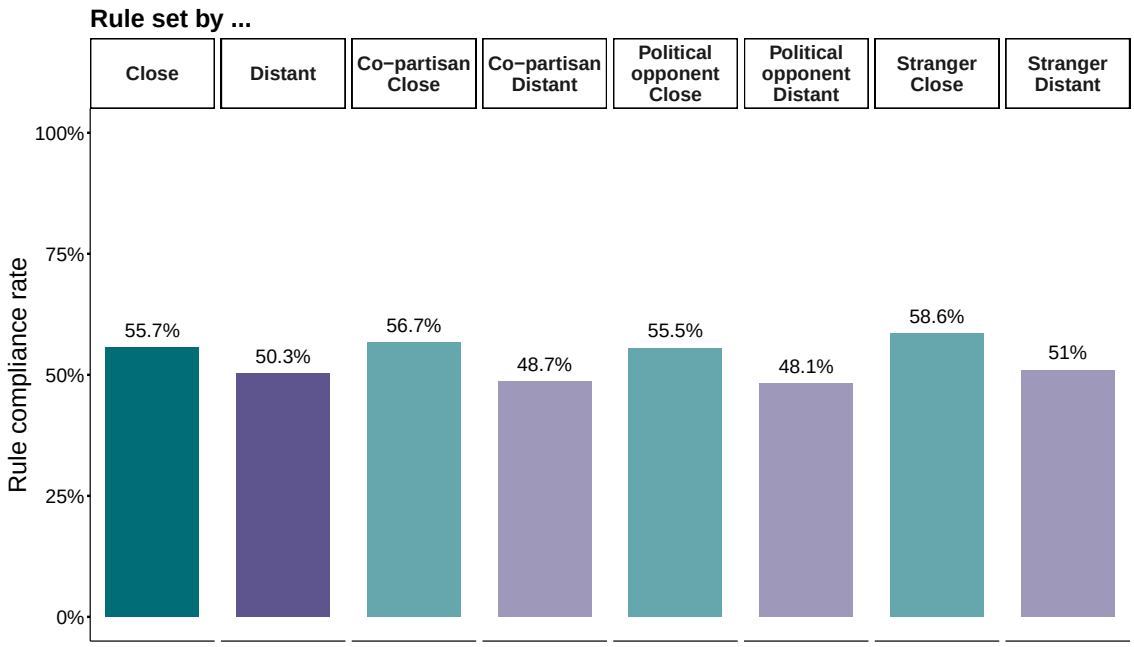


Figure 4: Pooled rule compliance and subjective closeness. The figure shows the share of rule followers across the different levels of subjective closeness as displayed in Table 2. Rule compliance is pooled across all phases.

rule compliance. Table 3 presents the estimation results for linear probability models where the dependent variable is rule compliance, and the key independent variable is subjective closeness to the rule setter. In both specifications in Columns 1 and 2, subjective closeness has a positive and statistically significant effect on rule compliance: a one-unit increase in perceived closeness (11-point Likert scale) is associated with an increase of one percentage point in the probability of complying with the rule ($p<.001$). This effect holds both in the baseline specification (Column 1), where we control for the phase and subjects' political ideology, and when adding further control variables in Column 2, like gender and age. This suggests that the observed relationship between rule compliance and closeness to the rule setter is also robust to standard individual-level controls. The magnitude of the effect is substantively meaningful, given that the closeness scale is bounded and given the modest overall variation in rule compliance rates. These results align with our descriptive findings and reinforce the interpretation that affective proximity to the rule setter correlates with greater behavioral alignment to the given rule (see also the linear probability models displayed in Table SI-9 which use dummy variables for the close-distant classifications in Table 2).

> ***Result 3:*** *Social distance toward the rule setter is significantly correlated with lower rule compliance, irrespective of phase and political ideology.*

These findings provide strong support for the pre-registered hypothesis that perceived social distance is correlated with lower compliance with arbitrary rules, reinforcing the role of affective polarization in shaping behavioral responses to political out-groups and even non-partisan actors. However, as can be seen in the estimation results of Column (3) in Table 3, subjective closeness to the rule setter turns insignificant once we control for general rule-following attitude, and normative and descriptive beliefs about rule-following behavior in the coins task. In the next subsection, we will thus examine these beliefs in detail.

Table 3: Impact of subjective closeness to rule setter on rule compliance

	Dependent variable: Rule compliance		
	(1)	(2)	(3)
Constant (Election Phase)	0.45***	0.48***	−0.11***
	(0.02)	(0.04)	(0.03)
Subjective closeness to rule setter	0.01***	0.01***	−0.002
	(0.002)	(0.002)	(0.002)
Pre-election Phase	−0.02	−0.02	0.01
	(0.02)	(0.02)	(0.01)
Post-inauguration Phase	0.03	0.03	−0.01
	(0.02)	(0.02)	(0.01)
Political ideology	0.01	0.01	−0.003
	(0.003)	(0.004)	(0.004)
General rule-following attitude			0.12***
			(0.01)
Normative belief (rule compliance)			−0.002
			(0.01)
Normative belief (rule violation)			−0.05***
			(0.01)
Descriptive belief			0.005***
			(0.0003)
Control variables	No	Yes	Yes
Observations	5,869	5,858	5,858
Adjusted R^2	0.01	0.01	0.21

Notes: The table reports coefficient estimates from linear probability models. The dependent variable is rule compliance. No data from the EXPERIMENTER condition is used due to non-existence of subjective closeness measures to the rule setter in this condition. Control variables include political party affiliation, gender, age and US regions dummies. The full estimation results are in Table SI-9. Robust standard errors are in parentheses. Levels of significance: *p<0.05, **p<0.01, ***p<0.001.

4.4 Respect for rules and social expectations about rule compliance

According to the CRISP framework (Gächter et al., 2025a), respect for rules (R) and social expectations (S, i.e., normative and descriptive beliefs) are the only motives that can explain rule compliance in our setting (because $I < 0$ and P is mute). We therefore expected and pre-registered (as the only secondary analysis in the pre-registrations for all three phases) that the likelihood to follow the rule in our study will, *ceteris paribus*, be higher for (i) people who think that others should follow rules in decision-making situations like this one (general rule-following attitude); (ii) people who think that others think one should follow the rule in our study (normative belief); and (iii) people who expect more others to have followed the rule in our study (descriptive belief). Model (3) in Table 3 evaluates these hypotheses.

The results strongly support the predictive power of *R*- and *S*-based mechanisms for rule compliance (*C*). A general rule-following attitude (a proxy for *R*)—that is, whether participants personally believe that one should follow the rule in this task (see, e.g., Bašić & Verrina, 2024)—is significantly positively associated with rule compliance (p<.001). The same holds true for descriptive beliefs (p<.001), capturing descriptive social expectations about others' compliance rates. Stronger social normative beliefs about the appropriateness of violating the rule—i.e., participants' beliefs about how others view the social appropriateness of a rule violation—are associated with significantly lower compliance rates (p<.001).

> ***Result 4a:*** *Consistent with the CRISP framework, respect for rules (internalized norms) and conformity with perceived social expectations are both correlated with rule compliance—although in our setting incentives suggest breaking the rule and social preferences are muted.*

Can these results on the impact and importance of beliefs be aligned with the previous result on the subjective distance toward the rule setter? The potential answer can be seen in Figure 5, where general rule-following attitude and social expectations (normative and descriptive beliefs) are shown separately for situations where the rule setter is the Experimenter, a politically identified or non-identified Prolific user who is perceived as Close or a user who is perceived to be Distant according to the median split on the respective IOS11 scores (see the previous section for details). Participants reporting greater psychological distance from the rule setter not only exhibit significantly lower compliance, they also exhibit reduced descriptive and normative beliefs about rule-following in the coins task. It seems that the (initially) vague norms in the coins task enable self-serving belief distortion about rule compliance, which is most pronounced if the rule setter is perceived to be socially distant.

In line with this, additional non-parametric analyses reveal that these belief measures systematically vary with the perceived social distance to the rule setter. For general rule-following attitudes (Kruskal-Wallis test, p<.001), we observe no difference between participants who are assigned a close rule setter and those governed by the experimenter (mean score of 3.28 for Experimenter and 3.3 for Close, Wilcoxon ranksum test with Holm correction, p=.66), but both groups display significantly stronger pro-rule attitudes than those exposed to distant rule setters (mean of 3.05, Experimenter vs. Distant, p<.001, and Close vs. Distant, p<.001). The same pattern is also visible in the descriptive beliefs (Kruskal-Wallis test: p<.001; with Wilcoxon ranksum test with Holm correction on pairwise comparisons yielding p=.2 for Experimenter vs. Close, and p<.001 for both pairwise comparisons with Distant), and likewise in both types of normative beliefs: participants report stronger perceived compliance norms and weaker perceived rule-violation norms when rule setters are Close compared to Distant. These differences are statistically significant across all three groups (in both cases Kruskal-Wallis p<.001; respective pairwise comparisons using Wilcoxon ranksum tests with Holm correction all p<.001, except for rule violation where Experimenter vs. Distant gives p=.14).

> ***Result 4b:*** *General rule-following attitudes and social expectations (descriptive and normative beliefs) about rule compliance are significantly shaped by social distance toward the rule setter, in particular when the rule setter is perceived as psychologically distant.*

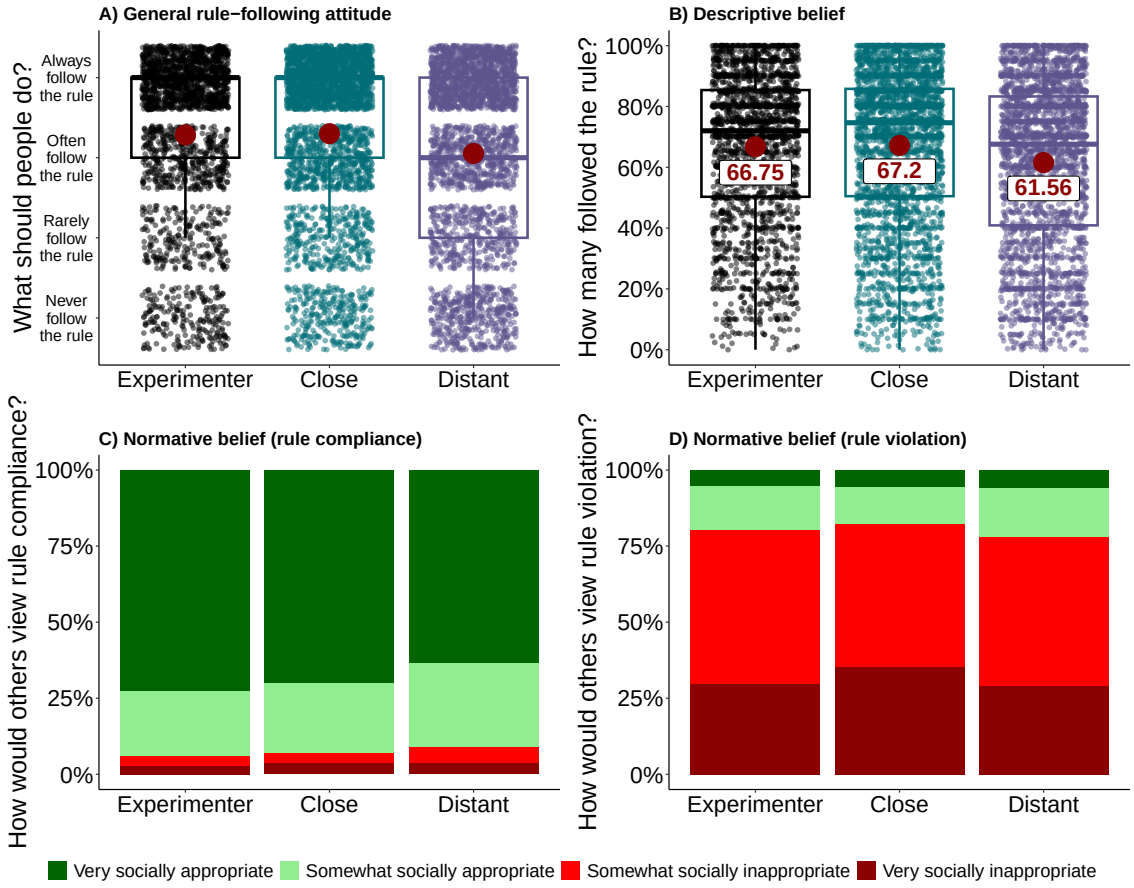


Figure 5: Beliefs across levels of subjective closeness. The figure shows the general rule-following attitude (A), descriptive (B) and normative (C and D) beliefs. The red dot in A) and B) represents the mean. A disaggregated overview of each wave, also w.r.t. our treatment manipulation, is displayed in Figures SI-2 to SI-9.

It seems plausible that these differences in beliefs then translate into lower rule compliance rates in DISTANT compared to CLOSE, as we saw in the previous section. Taken together, our results thus suggest that psychological closeness does not only shape rule compliance directly, but also indirectly by influencing the beliefs individuals hold about the rule-setting context. In terms of the CRISP framework, this supports a possible mechanism in which perceived social distance weakens norm internalization (intrinsic respect for rules) and the conformity with social expectations about others' compliance.

4.5 The role of the rule-setter's identity: Re-assessing the null results

We now turn to a final, exploratory, analysis that focuses on the general effect of revealing the identity of the rule setter. As shown before in Section 4.2, we find no significant differences in rule compliance between the CO-PARTISAN and POLITICAL OPPONENT conditions, neither in any phase nor for the pooled sample. Here we investigate how noisy the estimated null effects are, since statistically insignificant estimates may reflect large but noisy treatment differences (and hence type-II errors) and the true effect may be positive or negative. The question is, how big these possible effects are and whether we should care about these effects once we understand their potential effect sizes. For this assessment, we use "equivalence testing" (explained below).

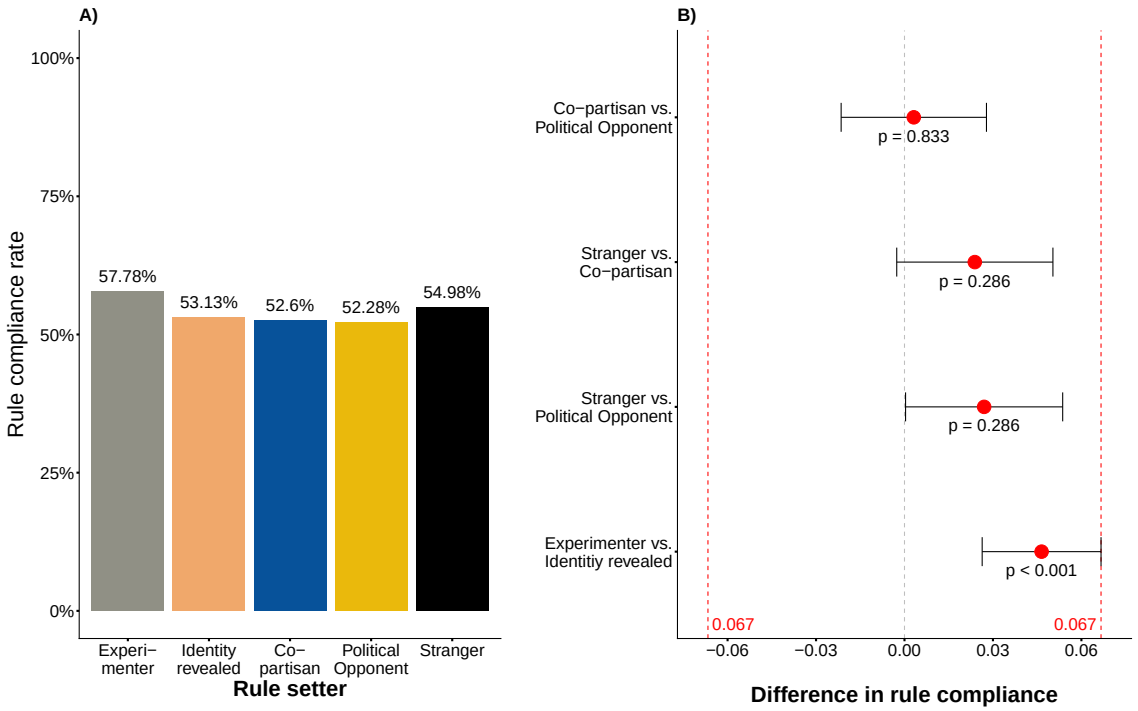


Figure 6: Differences in rule compliance across treatment conditions. Panel (A) shows mean rule compliance for each treatment condition. Panel (B) presents the 90% confidence intervals for pairwise differences between rule-setter identities. "Identity revealed" represents the pooled data for the Co-partisan, Political Opponent, and Stranger conditions. In panel (B), the red dotted lines indicate the maximum region of practical equivalence for differences defined by the "experimenter-induced demand effect". P-values are based on two-sided Wald tests and are adjusted for multiple comparisons using the Holm method. A disaggregated overview by wave is displayed in Figure SI-10.

For our analysis (and for expositional ease) we pool observations across survey phases and display rule compliance rates for all exogenously manipulated treatment conditions in Figure 6A. The column labeled "Identity revealed" reports the mean compliance rate across the three conditions in which the identity of the rule setter was disclosed (i.e., all except Experimenter). This pooled compliance rate is 53.13%, which is 4.65 percentage points lower than in the Experimenter condition (57.78%). This difference is statistically significant (Pearson's chi-squared test, p<.001) and the 90% confidence interval is [0.026, 0.067]. By contrast, we observe no statistically significant differences in compliance among the conditions Co-partisan, Political Opponent, and Stranger (see the difference estimates and associated p-values in Figure 6B). We refer to the 4.65-percentage-point compliance gap between the Experimenter condition and the identity-revealed conditions as the "experimenter-induced demand effect".

To understand the "experimenter-induced demand effect", consider the nature of "demand effects" in our context. An "experimenter demand effect" refers "to changes in behavior by experimental subjects due to cues about what constitutes appropriate behavior (behavior 'demanded' from them)" (Zizzo, 2010, p. 75; see also de Quidt et al., 2018; Fleming and Zizzo, 2015). Notice that a rule is a demand because it asks for a certain behavior, in our case, waiting for the "plus"-sign to appear. Therefore, rule compliance could be seen as a demand effect that satisfies the request of the rule setter. As Figure 6A shows, this demand is satisfied more often when the rule setter is the experimenter than when the rule setter is a co-partisan, a political

opponent, or another US citizen whose partisanship is not revealed. The average difference of 4.65 percentage points (with an upper bound in the 90% confidence interval of 6.7 percentage points) is therefore an experimenter-induced demand effect: the rate of rule compliance that can be attributed to the fact that an experimenter sets the rule.

We will use the experimenter-induced rule compliance rate in the following as a benchmark—*the smallest effect size of interest* (SESOI, Lakens et al., 2018)—to interpret our null results regarding the differences in rule compliance across rule-setter identities. We believe that, in the context of our study, the statistically significant experimenter-induced demand effect is a meaningful and natural SESOI against which to measure the noisiness—and hence the potentially true effect sizes—of rule-setter identity for rule compliance.

We evaluate our null results using "equivalence testing" (e.g., Lakens et al., 2018; Fitzgerald, 2024) where, as argued above, we use the maximal experimenter-induced demand effect of 6.7 percentage points as the SESOI in our analysis. Equivalence testing requires to do two one-sided tests (TOSTs) that check whether the observed values of the corresponding 90% confidence interval are significantly different from the SESOI benchmark (0.067 in our case), which implies a *region of practical equivalence* (ROPE), indicated by the red dashed lines in Figure 6B.[19] If the 90% confidence interval lies fully within ROPE we declare the effect as behaviorally negligible.

The results of this approach are presented in Figure 6B. Focusing first on the primary comparison of interest—Co-partisan vs. Political Opponent—the estimated difference in compliance is small (0.3 percentage points), and its 90% confidence interval lies entirely within ROPE set by the experimenter-induced demand effect. Accordingly, the TOSTs indicate statistical equivalence in compliance between these two conditions.

The same pattern holds for the remaining pairwise comparisons among the identity-revealed conditions. In all cases, estimated differences are small (<2.8 percentage points), their 90% confidence intervals fall fully within ROPE and the corresponding TOSTs indicate statistical equivalence. Taken together, these results suggest that while revealing the rule setter's identity reduces compliance relative to the Experimenter condition, the specific type of revealed identity does not meaningfully affect compliance behavior—provided that a maximum difference in rule compliance of 6.7 percentage points is considered "too small" to be of interest.

At first glance, a difference of 6.7 percentage points may appear behaviorally meaningful. However, we argue that variation of this magnitude is modest in the context of existing evidence derived from the CRISP framework, for two reasons. First, rule compliance in situations comparable to the Experimenter condition varies across studies, ranging from 58% to 65% in the literature (Kimbrough & Vostroknutov, 2016, 2018; Molleman et al., 2023; Kimbrough et al., 2024; Gächter et al., 2025a; Suri et al., 2025). This 7–percentage-point spread mirrors the equivalence bound used here.

Second, a closer examination of rule compliance as summarized in Gächter et al. (2025a, Table 1) reveals considerable heterogeneity. Estimated compliance rates are 22% and 23% in situations in which no one disapproves of rule violations and no one complies with the rule, respectively—both significantly different from zero. This implies that unconditional respect

[19] See, e.g., Lakens et al. (2018) and Fitzgerald (2024) for good explanations and applications in economics, and psychology, respectively.

for rules (*R* in the CRISP framework) explains almost a quarter of rule compliance rates. Moreover, regarding the effect of conformity with social expectations (*S* in the CRISP framework), the authors report differences of 20 to 23 percentage points between the lowest and highest quintiles of conditional compliance rates. Against this backdrop, differences on the order of 6.7 percentage points appear behaviorally rather small. Hence, we conclude that the null effects of the role of political identity in rule compliance are also behaviorally negligible.

5 Discussion and conclusion

In this study, we tested the resilience of rule compliance in a highly polarized society: the United States of America. Because ideological and affective polarization is very high in the US as a whole and also among our participants, our setting of partisan rule setters during the 2024 US election cycle is a particularly strong stress test for the resilience of rule compliance. Surprisingly, the strong polarization we observe among our participants did not translate into lower rule compliance even when the rule originates from an opposing partisan.[20] Instead, and although the level of rule compliance was very high in general, it was subjective social closeness, rather than political ideology *per se*, that predicted rule compliance: individuals were more likely to break rules set by those they feel distant from. In terms of the CRISP framework (Gächter et al., 2025a), this relationship seemed to be mediated by general attitudes about the respect for rules (*R*), and conformity with social expectations (*S*, i.e., normative and descriptive beliefs) about others' compliance with the rule, all of which deteriorated as subjective social distance to the rule setter increased.

Our findings contribute to the growing literature on the consequences of political polarization by highlighting a critical distinction: while ideological polarization may be pervasive, its behavioral consequences seem to be limited when explicitly stated rules are in place (at least in our context). Participants followed arbitrary rules at similarly high rates across rule setters—whether the rule setter was a political co-partisan, a political opponent, politically neutral or the experimenter. This might seem surprising in view of recent findings (Iyengar & Westwood, 2015; Dimant, 2024; Feldhaus et al., 2024; Freundt & Herz, 2024), which show that political identity meaningfully shapes cooperative and trusting behavior. In our view, these differences might be due to three potential reasons. First, some studies use Trump lover/Trump hater as the cue for political ideology (Dimant, 2024; Feldhaus et al., 2024), which arguably might be perceived as an affective rather than an ideological cue; and for the former, we do observe behavioral effects, too. Second, while in some studies the co-partisan or political opponent is directly affected by the participant's actions (Iyengar & Westwood, 2015; Dimant, 2024; Feldhaus et al., 2024), the rule-following behavior of a participant in our design has no immediate impact on the co-partisan or opposing-party individual. Third, in view of the findings on the importance of rules (see, e.g., Kimbrough & Vostroknutov, 2016; Gächter et al., 2025a), another potential explanation lies in our study's framing: unlike "restrictions", "suggestions" or "requests" in other experiments, we explicitly refer to a "rule", which may trigger different, stronger internalized norms of compliance.

In contrast, affective polarization appears far more consequential than the ideological di-

[20] Recall that in our coins task, participants can freely choose whether to comply with an arbitrary rule, and that control questions are used to ensure that any rule breaking is understood to entail no negative consequences.

vide. Rule compliance was found to drop significantly in our study when participants perceived the rule setter as socially distant, regardless of their political identity. Moreover, we found that the decline in compliance was accompanied by shifts in beliefs: participants who felt distant from the rule setter were less likely to believe that the rule should be followed, less likely to think that others follow the rule, and less likely to perceive social pressure to comply. It seems that these belief distortions might have facilitated motivated reasoning and self-serving behavior. Strikingly, these dynamics were also observed in our STRANGER condition, indicating that even non-political affective distance undermines rule compliance.

The link between affective polarization and rule compliance is technically correlational: our experiment is not designed to identify a causal effect of social closeness on rule compliance. Nonetheless, several considerations make a causal interpretation plausible. First, social closeness typically reflects relationships that form and stabilize over longer horizons (e.g., Gächter et al., 2015), whereas the experimental interaction is very brief. It is therefore unlikely that behavior in the coins task materially reshapes the underlying relationship. Second, reverse causality is implausible. A single instance of rule-breaking in the coins task lasting a few moments—even if attributed to a salient out-group characteristic like partisanship—is unlikely to systematically increase perceived social distance toward the counterpart. Third, prior research documents robust associations between social closeness and behavior in cooperation settings (e.g., Gächter et al., 2015, 2022). Fourth, and most importantly, Tufano et al. (mimeo) leverage random assignment to naturally occurring groups find a causal effect of social closeness on cooperation. Taken together, these points suggest that perceived social closeness is more likely to affect rule compliance than the reverse.

The significant impact of affective polarization is concerning, not only due to increasing dislike of opposing social and political actors, but particularly given that our data reveal substantial heterogeneity in subjective distance even toward co-partisans. The emergence of this heterogeneity may be linked to the observation that, within strongly polarized societies—like the US population in our study—individuals initially face diminishing social repercussions for the expression of negative attitudes directed at political opponents (Iyengar & Westwood, 2015). This erosion of social constraints may then extend to the expression of negative attitudes even toward co-partisans, contributing to the observed heterogeneity in subjective distance within political groups. Intra-party dynamics, such as divisive primary debates or personalized attacks, could amplify these effects. As political discourse becomes more personalized and affectively charged, rather than ideologically grounded, the risks of affective polarization may become increasingly severe. Ultimately, affective polarization may pose a greater threat to democratic societies than the ideological divide.

Our findings suggest at least three important directions for future research. A first promising avenue would be to explore the causal mechanisms underlying affective polarization, such as how different types of political discourse (e.g., ideological vs. affective) influence subjective closeness toward others and, consequently, rule compliance behavior; or whether policies designed to strengthen national identity could foster greater citizen compliance with rules by mitigating affective polarization, even with rules originating from political opponents.

A second avenue would be to explore, for a given degree of affective polarization, the potential of behavioral interventions to dampen the negative effect of affective distance—be it in the form of behavioral nudges (e.g., like in Dimant, 2024, who studies a default nudge and an

information nudge) or strong institutions (like in Robbett & Matthews, 2023, where punishment mechanisms are shown to foster cooperation in mixed-party groups, that are otherwise less cooperative than homogeneous groups).

The third avenue would be to explore procedural details more intensely. While we deliberately focused on the legitimacy of a given rule with respect to the rule-setter's identity to have a clean behavioral measure of rule compliance without introducing potential confounds due to participant's preferences, future studies might explore the interplay between the process of rule formation and the identity of individuals involved in passing those rules. In our setting, participants may find it difficult to understand why a co-partisan or political opponent would implement such an arbitrary rule, as rules in political discourse are typically perceived as serving some underlying interest. For example, a rule could introduce positive (or negative) externalities associated with compliance (or non-compliance), such as a donation to the rule setter's preferred political party. Along these lines, corresponding experiments might also check to what extent the exact framing does affect observed behavioral results (e.g., is a "law" equally effective as a "rule"), or test the scope of setting up clear rules in strategic situations like cooperative and trusting environments that were used in the closely related studies on political polarization and behavior.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Declaration of AI and AI-assisted technologies in the writing process

During the preparation of this work, we used ChatGPT in order to improve readability and language. After using this service, we reviewed and edited the content as needed and take full responsibility for the content of the publication.

Authorship contribution statement

Dominik Suri: Conceptualization, Data curation, Formal analysis, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Simon Gächter:** Conceptualization, Formal analysis, Writing – original draft, Writing – review & editing. **Sebastian Kube:** Conceptualization, Formal analysis, Writing – original draft, Writing – review & editing. **Johannes Schultz:** Conceptualization, Formal analysis, Writing – review & editing.

References

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of Other in the Self Scale and the Structure of Interpersonal Closeness. *Journal of Personality and Social Psychology*, *63*(4), 596–612. https://doi.org/10.1037/0022-3514.63.4.596.

Baader, M., Starmer, C., Tufano, F., & Gächter, S. (2024). Introducing IOS11 as an extended interactive version of the 'Inclusion of Other in the Self' scale to estimate relationship closeness. *Scientific Reports*, *14*(1), 8901. https://doi.org/10.1038/s41598-024-58042-6.

Balliet, D., Tybur, J. M., Wu, J., Antonellis, C., & Van Lange, P. A. M. (2018). Political Ideology, Trust, and Cooperation: In-group Favoritism among Republicans and Democrats during a US National Election. *Journal of Conflict Resolution*, *62*(4), 797–818. https://doi.org/10.1177/0022002716658694.

Bašić, Z., & Verrina, E. (2024). Personal norms — and not only social norms — shape economic behavior. *Journal of Public Economics*, *239*, 105255. https://doi.org/10.1016/j.jpubeco.2024.105255.

Bicchieri, C. (2006). *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press.

Bicchieri, C., Dimant, E., Gächter, S., & Nosenzo, D. (2022). Social proximity and the erosion of norm compliance. *Games and Economic Behavior*, *132*, 59–72. https://doi.org/10.1016/j.geb.2021.11.012.

Bicchieri, C., Gächter, S., Molleman, L., & Nosenzo, D. (2025). Group identity and peer effects in rule-following. *Journal of Economic Behavior & Organization*, *239*, 107264. https://doi.org/10.1016/j.jebo.2025.107264.

Bicchieri, C., & Xiao, E. (2009). Do the right thing: But only if others do so. *Journal of Behavioral Decision Making*, *22*(2), 191–208. https://doi.org/10.1002/bdm.621.

Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., Heikensten, E., Holzmeister, F., Imai, T., Isaksson, S., Nave, G., Pfeiffer, T., Razen, M., & Wu, H. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, *351*(6280), 1433–1436. https://doi.org/10.1126/science.aaf0918.

Campbell, A., Converse, P. E., Miller, W. E., & Stokes, D. E. (1980). *The American Voter*. University of Chicago Press.

Charness, G., Dimant, E., Gneezy, U., & Krupka, E. (2025). *Experimental Methods: Eliciting and Measuring Social Norms* (SSRN Working Paper No. 5194790). https://doi.org/10.2139/ssrn.5194790.

Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, *9*, 88–97. https://doi.org/10.1016/j.jbef.2015.12.001.

Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology*, *55*(1), 591–621. https://doi.org/10.1146/annurev.psych.55.090902.142015.

Croson, R., & Gächter, S. (2010). The science of experimental economics. *Journal of Economic Behavior & Organization*, *73*(1), 122–131. https://doi.org/10.1016/j.jebo.2009.09.008.

Dana, J., Weber, R. A., & Kuang, J. X. (2007). Exploiting moral wiggle room: Experiments demonstrating an illusory preference for fairness. *Economic Theory*, *33*, 67–80. https://doi.org/10.1007/s00199-006-0153-z.

Daston, L. (2022). *Rules: A short history of what we live by*. Princeton University Press.

de Quidt, J., Haushofer, J., & Roth, C. (2018). Measuring and bounding experimenter demand. *American Economic Review*, *108*(11), 3266–3302. https://doi.org/10.1257/aer.20171330.

Desmet, P. T., & Engel, C. (2021). People are conditional rule followers. *Journal of Economic Psychology*, *85*, 102384. https://doi.org/10.1016/j.joep.2021.102384.

Dimant, E. (2024). Hate Trumps Love: The Impact of Political Polarization on Social Preferences. *Management Science, 70*(1), 1–31. https://doi.org/10.1287/mnsc.2023.4701.

Dimant, E., & Kimbrough, E. O. (2024). Polarization in multidisciplinary perspective. *PNAS Nexus, 3*(10), pgae425. https://doi.org/10.1093/pnasnexus/pgae425.

Dreber, A., & Johannesson, M. (2025). A framework for evaluating reproducibility and replicability in economics. *Economic Inquiry, 63*(2), 338–356. https://doi.org/10.1111/ecin.13244.

Elster, J. (1989). *The Cement of Society. A Study of Social Order*. Cambridge University Press.

Enke, B. (2020). Moral Values and Voting. *Journal of Political Economy, 128*(10), 3679–3729. https://doi.org/10.1086/708857.

Fasching, N., Iyengar, S., Lelkes, Y., & Westwood, S. J. (2024). Persistent polarization: The unexpected durability of political animosity around US elections. *Science Advances, 10*(36), eadm9198. https://doi.org/10.1126/sciadv.adm9198.

Feldhaus, C., Reinhardt, L., & Sutter, M. (2024). *Trump ante Portas: Political Polarization Undermines Rule-Following Behavior* (SSRN Working Paper No. 5019206). https://doi.org/10.2139/ssrn.5019206.

Fitzgerald, J. (2024). *The Need for Equivalence Testing in Economics* (I4R Discussion Paper Series No. 125).

Fleming, P., & Zizzo, D. J. (2015). A simple stress test of experimenter demand effects. *Theory and Decision, 78*(2), 219–231. https://doi.org/10.1007/s11238-014-9419-2.

Freundt, J., & Herz, H. (2024). *From Partisanship to Preference: How Identity Shapes Dependence Aversion* (SSRN Working Paper No. 4991802). https://doi.org/10.2139/ssrn.4991802.

Gächter, S., Molleman, L., & Nosenzo, D. (2025a). Why people follow rules. *Nature Human Behaviour, 9*, 1342–1354. https://doi.org/10.1038/s41562-025-02196-4.

Gächter, S., Starmer, C., Thöni, C., Tufano, F., & Weber, T. O. (2022). Social closeness can help, harm and be irrelevant in solving pure coordination problems. *Economics Letters, 216*, 110552. https://doi.org/10.1016/j.econlet.2022.110552.

Gächter, S., Starmer, C., & Tufano, F. (2015). Measuring the closeness of relationships: A comprehensive evaluation of the 'inclusion of the other in the self' scale. *PLOS ONE, 10*(6), e0129478. https://doi.org/10.1371/journal.pone.0129478.

Gächter, S., Starmer, C., & Tufano, F. (2025b). Measuring Group Cohesion to Reveal the Power of Social Relationships in Team Production. *The Review of Economics and Statistics, 107*(2), 539–554. https://doi.org/10.1162/rest_a_01283.

Garzia, D., & da Silva, F. F. (2022). The Electoral Consequences of Affective Polarization? Negative Voting in the 2020 US Presidential Election. *American Politics Research, 50*(3), 303–311. https://doi.org/10.1177/1532673X221074633.

Gelfand, M. (2019). *Rule makers, rule breakers: Tight and loose cultures and the secret signals that direct our lives*. Scribner.

Gidron, N., Adams, J., & Horne, W. (2020). *American affective polarization in comparative perspective*. Cambridge University Press.

Guriev, S., & Papaioannou, E. (2022). The Political Economy of Populism. *Journal of Economic Literature, 60*(3), 753–832. https://doi.org/10.1257/jel.20201595.

Hobbes, T. (1996). *Leviathan* (R. Tuck, Ed.). Oxford University Press. (Original work published 1651)

Hoyer, K., Sijtsema, J., Kogler, C., van den Bos, W., & Molleman, L. (2025). Determinants of rule-breaking in adolescence. *Journal of Adolescence*, *97*(8), 2185–2197. https://doi.org/10.1002/jad.70031.

Huddy, L., Mason, L., & Aarøe, L. (2015). Expressive Partisanship: Campaign Involvement, Political Emotion, and Partisan Identity. *American Political Science Review*, *109*(1), 1–17. https://doi.org/10.1017/S0003055414000604.

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The Origins and Consequences of Affective Polarization in the United States. *Annual Review of Political Science*, *22*(1), 129–146. https://doi.org/10.1146/annurev-polisci-051117-073034.

Iyengar, S., & Westwood, S. J. (2015). Fear and Loathing across Party Lines: New Evidence on Group Polarization. *American Journal of Political Science*, *59*(3), 690–707. https://doi.org/10.1111/ajps.12152.

Karakostas, A., & Zizzo, D. J. (2016). Compliance and the power of authority. *Journal of Economic Behavior & Organization*, *124*, 67–80. https://doi.org/10.1016/j.jebo.2015.09.016.

Kimbrough, E. O., Krupka, E. L., Kumar, R., Murray, J. M., Ramalingam, A., Sánchez-Franco, S., Sarmiento, O. L., Kee, F., & Hunter, R. F. (2024). On the stability of norms and norm-following propensity: A cross-cultural panel study with adolescents. *Experimental Economics*, *27*(2), 351–378. https://doi.org/10.1007/s10683-024-09821-5.

Kimbrough, E. O., & Vostroknutov, A. (2016). Norms make preferences social. *Journal of the European Economic Association*, *14*(3), 608–638. https://doi.org/10.1111/jeea.12152.

Kimbrough, E. O., & Vostroknutov, A. (2018). A portable method of eliciting respect for social norms. *Economics Letters*, *168*, 147–150. https://doi.org/10.1016/j.econlet.2018.04.030.

Kliemt, H. (2020). Economic and sociological accounts of social norms. *Analyse & Kritik*, *42*(1), 41–96. https://doi.org/10.1515/auk-2020-0003.

Krupka, E. L., & Weber, R. A. (2013). Identifying Social Norms Using Coordination Games: Why Does Dictator Game Sharing Vary? *Journal of the European Economic Association*, *11*(3), 495–524. https://doi.org/10.1111/jeea.12006.

Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence Testing for Psychological Research: A Tutorial. *Advances in Methods and Practices in Psychological Science*, *1*(2), 259–269. https://doi.org/10.1177/2515245918770963.

Maniadis, Z., Tufano, F., & List, J. A. (2014). One swallow doesn't make a summer: New evidence on anchoring effects. *American Economic Review*, *104*(1), 277–90. https://doi.org/10.1257/aer.104.1.277.

Martin, D., & Nai, A. (2024). Deepening the rift: Negative campaigning fosters affective polarization in multiparty elections. *Electoral Studies*, *87*, 102745. https://doi.org/10.1016/j.electstud.2024.102745.

Mason, L. (2023). Political Identities. In *The Oxford Handbook of Political Psychology* (3rd ed., pp. 886–917). Oxford University Press.

Molleman, L., Nosenzo, D., & Venema, T. (2023). *Ambiguity induces opportunistic rule breaking and erodes social norms* (Utrecht University Preprint). https://doi.org/10.31234/osf.io/n8bjp.

Posner, E. A. (2000). *Law and Social Norms*. Harvard University Press.

Puryear, C., Kubin, E., Schein, C., Bigman, Y. E., Ekstrom, P., & Gray, K. (2024). People believe political opponents accept blatant moral wrongs, fueling partisan divides. *PNAS Nexus*, *3*(7), pgae244. https://doi.org/10.1093/pnasnexus/pgae244.

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria.

Robbett, A., & Matthews, P. H. (2023). Polarization and Group Cooperation. *Quarterly Journal of Political Science*, *18*(2), 215–241. https://doi.org/10.1561/100.00021036.

Selten, R. (1998). Axiomatic characterization of the quadratic scoring rule. *Experimental Economics*, *1*, 43–61. https://doi.org/10.1023/A:1009957816843.

Smith, V. L. (1994). Economics in the laboratory. *Journal of Economic Perspectives*, *8*(1), 113–131. https://doi.org/10.1257/jep.8.1.113.

Sood, G., & Iyengar, S. (2016). *Coming to dislike your opponents: The polarizing impact of political campaigns* (SSRN Working Paper No. 2840225). https://doi.org/10.2139/ssrn.2840225.

Suri, D., Kube, S., & Schultz, J. (2025). Evaluating Online Data Collection Platforms Using A Simple Rule-Following Task. *Economics Letters*, *255*, 112509. https://doi.org/10.1016/j.econlet.2025.112509.

Tufano, F., Weber, T. O., Thöoni, C., Starmer, C., & Gächter, S. (mimeo). *How social relationships shape group cooperation and its foundations: Evidence from randomly assigned real groups* (Working Paper).

Weber, M. (1978). *Economy and Society: An Outline of Interpretitive Sociology* (G. Roth & C. Wittich, Eds.). University of California Press. (Original work published 1922)

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, *13*(1), 75–98. https://doi.org/10.1007/s10683-009-9230-z.

Supplementary Information

for

The resilience of rule compliance in a polarized society

Dominik Suri*, Simon Gächter, Sebastian Kube, Johannes Schultz

*Corresponding author. E-mail: dsuri@uni-bonn.de

Contents

SI-1 Supplementary Figures

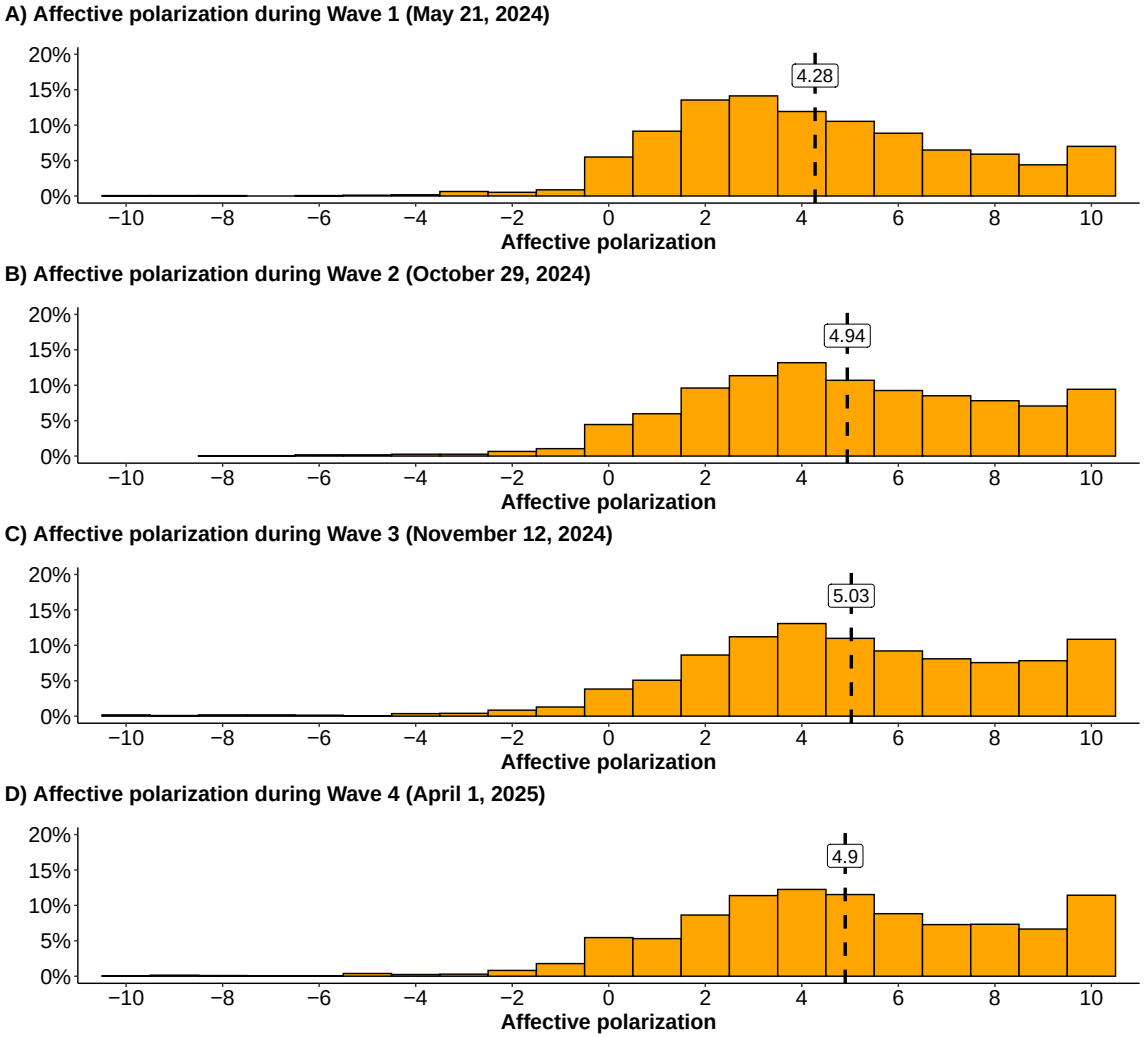


Figure SI-1: Affective polarization across phases. The figure shows the distribution of affective polarization across phases. Affective polarization is measured by subtracting the IOS11 score toward the political opponent from the co-partisan score. Negative values indicate a higher subjective closeness to someone from the opposing party compared to someone from the same party. The dashed line represents the mean. Mean affective polarization for Election Phase, i.e., combining Wave 2 and 3, is 4.98.

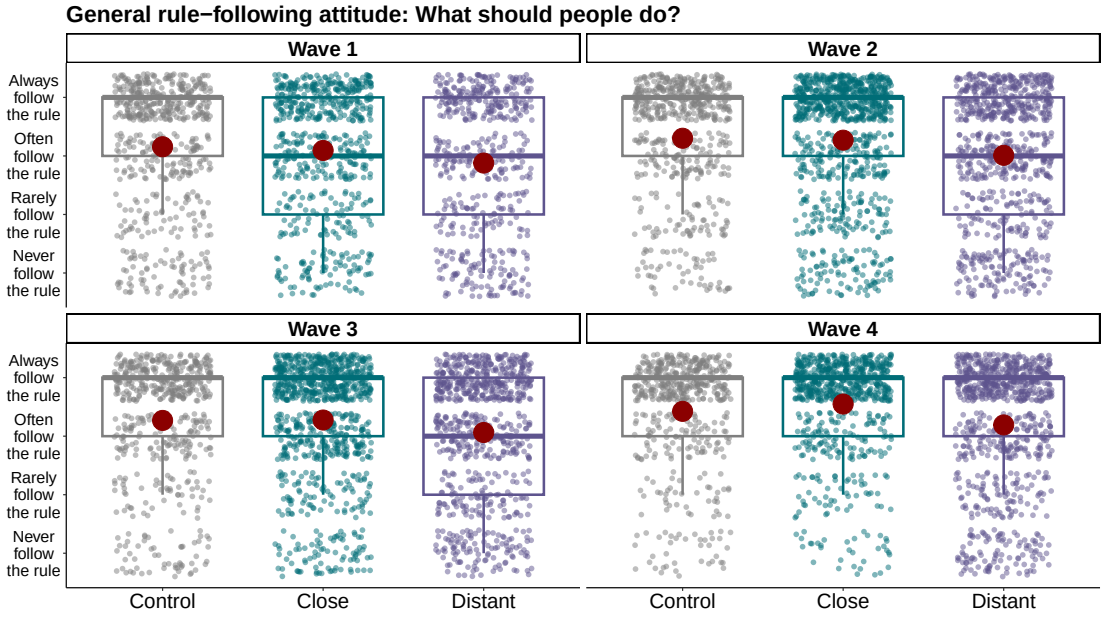


Figure SI-2: General rule-following attitude w.r.t. subjective closeness across waves. The figure shows the general rule-following attitude for EXPERIMENTER, CLOSE and DISTANT across waves. The exact wording of the general rule-following attitude is displayed in Section SI-3.2. The red dot indicates the mean.

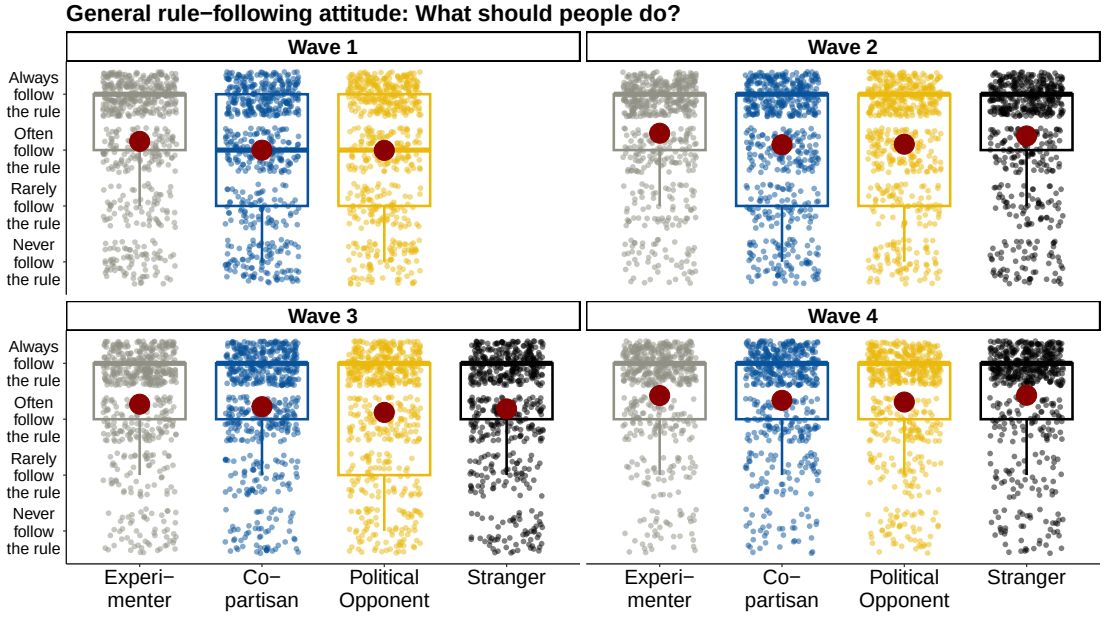


Figure SI-3: General rule-following attitude w.r.t. treatment conditions across waves. The figure shows the general rule-following attitude for EXPERIMENTER, CO-PARTISAN, POLITICAL OPPONENT and STRANGER across waves. The exact wording of the general rule-following attitude is displayed in Section SI-3.2. The red dot indicates the mean.

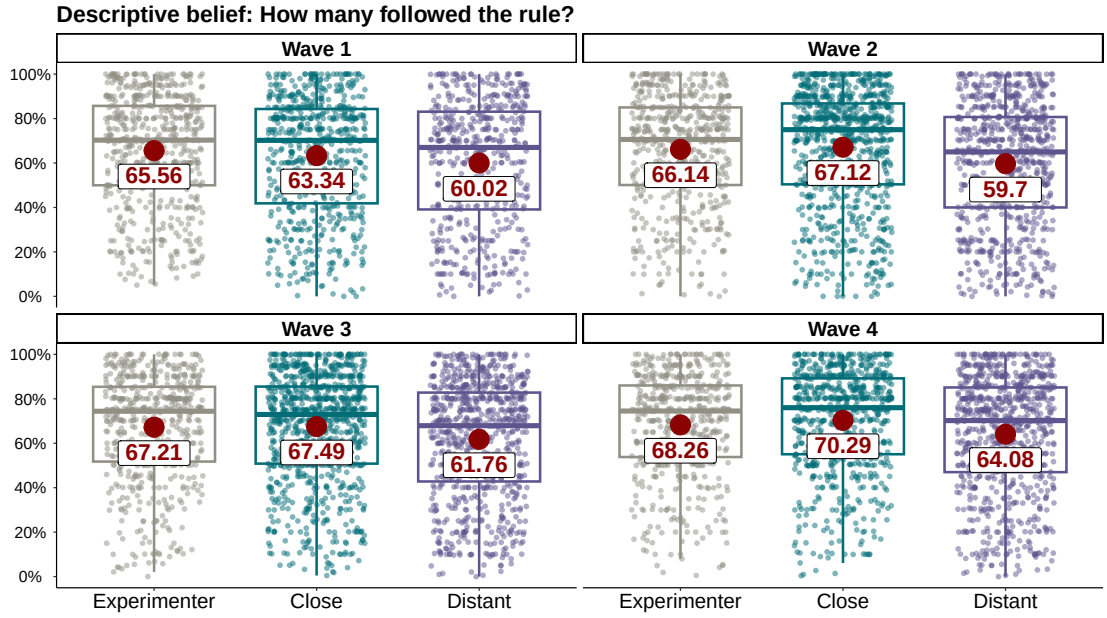


Figure SI-4: Descriptive belief w.r.t. subjective closeness across waves. The figure shows the descriptive belief for EXPERIMENTER, CLOSE and DISTANT across waves. The exact wording of the descriptive belief is displayed in Section SI-3.2. The red dot indicates the mean.

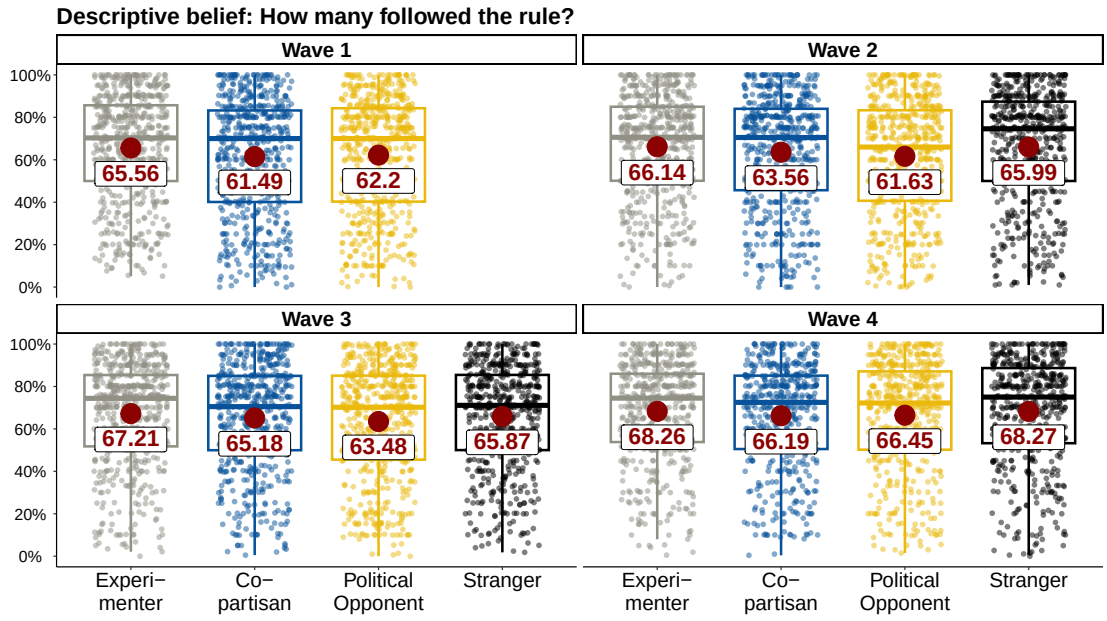


Figure SI-5: Descriptive belief w.r.t. treatment conditions across waves. The figure shows the descriptive belief for EXPERIMENTER, CO-PARTISAN, POLITICAL OPPONENT and STRANGER across waves. The exact wording of the descriptive belief is displayed in Section SI-3.2. The red dot indicates the mean.

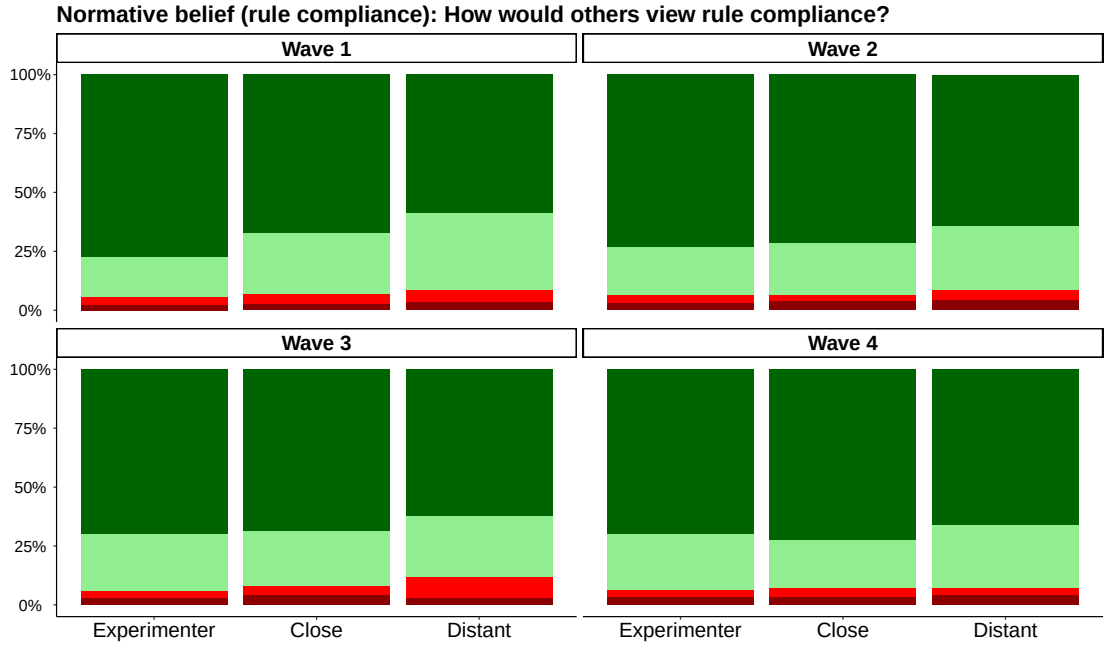


Figure SI-6: Normative belief (rule compliance) w.r.t. subjective closeness across waves. The figure shows the normative belief (rule compliance) for EXPERIMENTER, CLOSE and DISTANT across waves. The exact wording of the normative belief (rule compliance) is displayed in Section SI-3.2. The red dot indicates the mean.

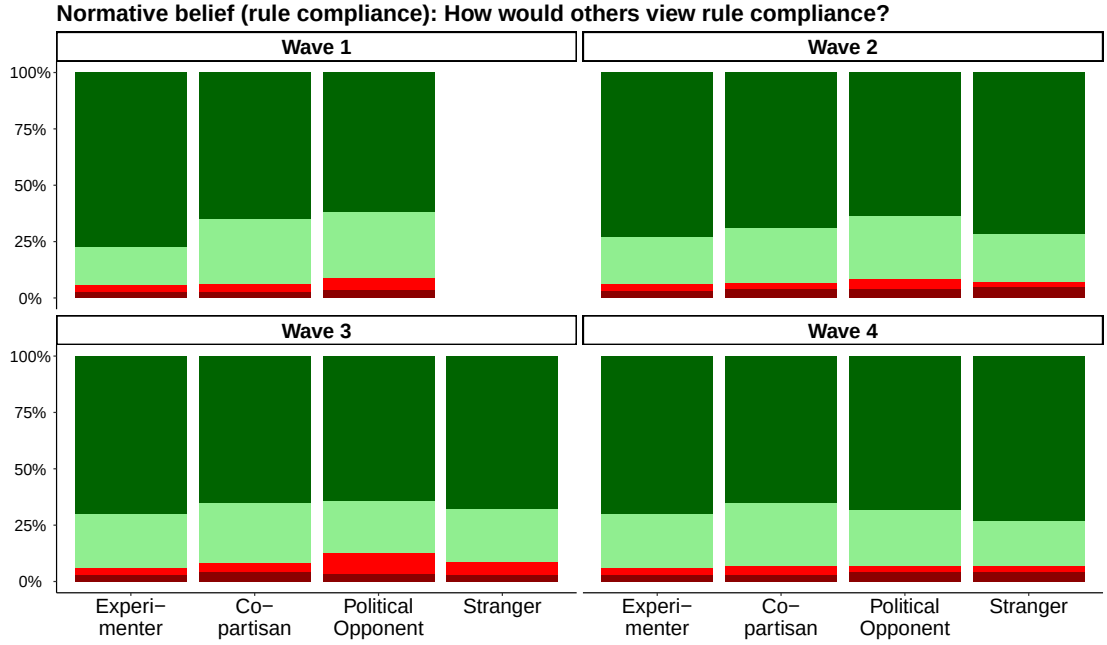


Figure SI-7: Normative belief (rule compliance) w.r.t. treatment conditions across waves. The figure shows the normative belief (rule compliance) for EXPERIMENTER, CO-PARTISAN, POLITICAL OPPONENT and STRANGER across waves. The exact wording of the normative belief (rule compliance) is displayed in Section SI-3.2. The red dot indicates the mean.

Figure SI-8: Normative belief (rule violation) w.r.t. subjective closeness across waves. The figure shows the normative belief (rule violation) for EXPERIMENTER, CLOSE and DISTANT across waves. The exact wording of the normative belief (rule violation) is displayed in Section SI-3.2. The red dot indicates the mean.

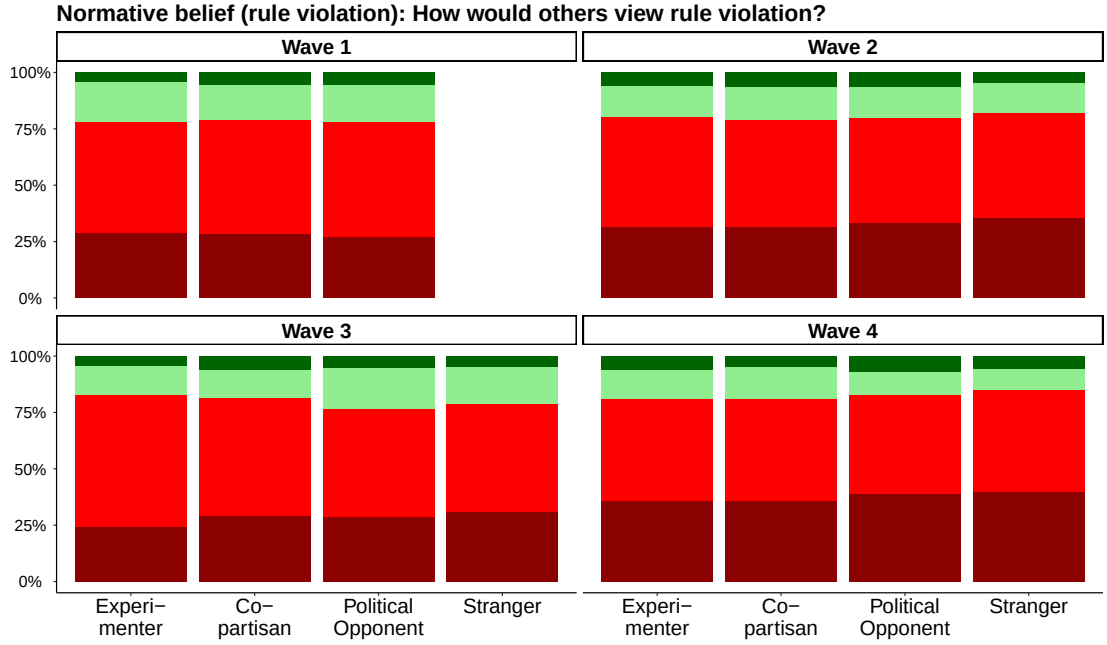


Figure SI-9: Normative belief (rule violation) w.r.t. treatment conditions across waves. The figure shows the normative belief (rule violation) for EXPERIMENTER, CO-PARTISAN, POLITICAL OPPONENT and STRANGER across waves. The exact wording of the normative belief (rule violation) is displayed in Section SI-3.2. The red dot indicates the mean.

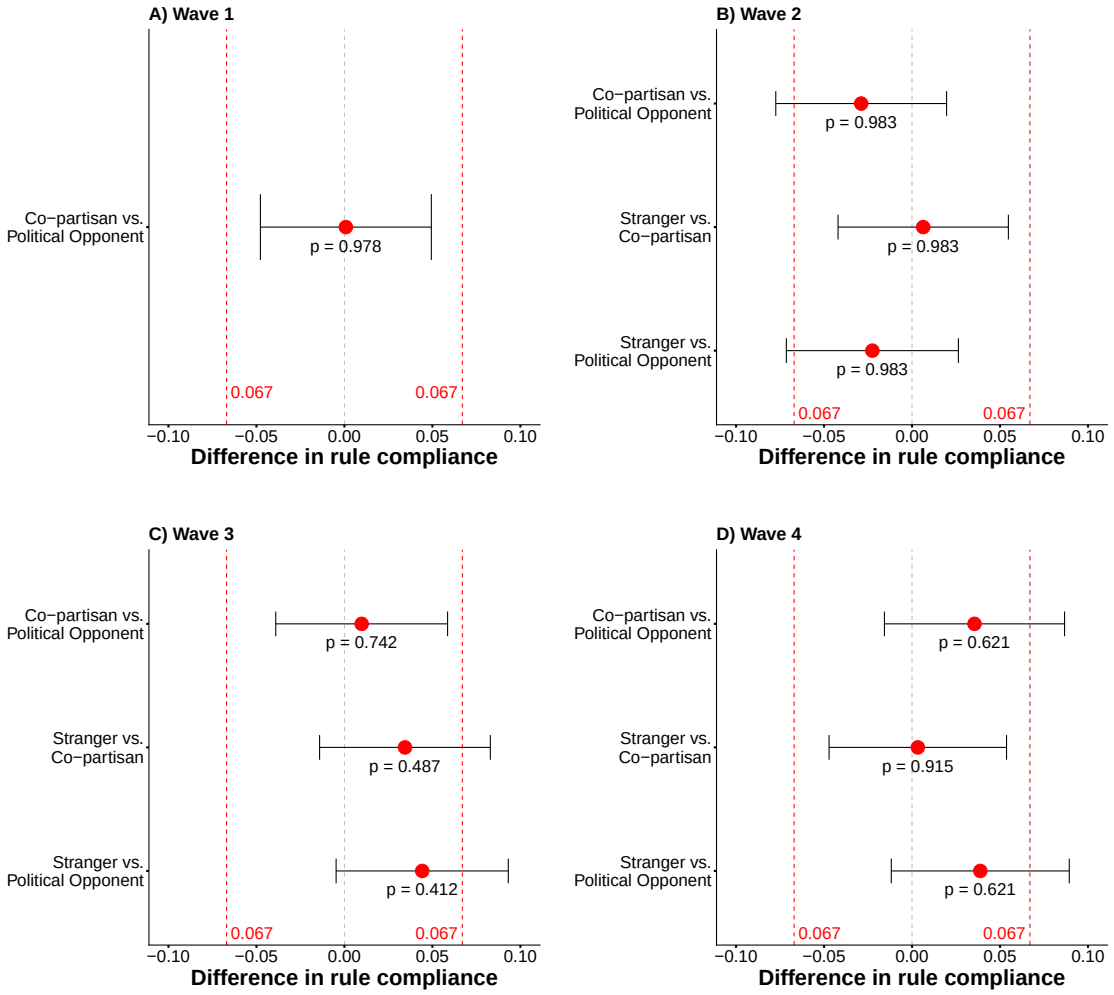


Figure SI-10: 90% confidence intervals for pairwise differences between rule-setter identities separated by wave. The figure shows the 90% confidence intervals for pairwise differences between rule-setter identities separated by wave. The red dotted lines indicate the maximum region of practical equivalence for differences defined by the "experimenter-induced demand effect". P-values are based on two-sided Wald tests and are adjusted for multiple comparisons using the Holm method.

SI-2 Supplementary Tables

Table SI-1: Setting overview

Setting	Description
Experimenter	Rule is (implicitly) set by the experimenter.
Stranger	Rule is set by a US Prolific user without revealing their political party affiliation.
Co-partisan	Rule is set by someone from the same party (e.g. if the participant is a Democrat and the rule is set by a Democrat).
Political Opponent	Rule is set by someone from the opposing party (e.g. if the participant is a Democrat and the rule is set by a Republican).
Close	Rule is set by a Prolific user from the US who has a high subjective closeness to the participant.
Distant	Rule is set by a Prolific user from the US who has a low subjective closeness to the participant.
Co-partisan Close	Rule is set by someone from the same party who has a high subjective closeness to the participant.
Co-partisan Distant	Rule is set by someone from the same party who has a low subjective closeness to the participant.
Political Opponent Close	Rule is set by someone from the opposing party who has a high subjective closeness to the participant.
Political Opponent Distant	Rule is set by someone from the opposing party who has a low subjective closeness to the participant.
Stranger Close	Rule is set by a stranger who has a high subjective closeness to the participant.
Stranger Distant	Rule is set by a stranger and who has a low subjective closeness to the participant.

Notes: We did not exogenously manipulate subjective closeness toward the rule setter, thus our exogenous treatment manipulations are Experimenter, Stranger, Co-partisan and Political Opponent.

Table SI-2: Sample characteristics with difference tests in means for Election Phase, which consists of Wave 2 (October 29, 2024) and Wave 3 (November 12, 2024), separated by political party affiliation

	Democrats			Republicans		
	Wave 2	Wave 3	p-value	Wave 2	Wave 3	p-value
Male	0.485 (0.500)	0.494 (0.500)	0.714	0.495 (0.500)	0.491 (0.500)	0.879
Age	41.21 (13.52)	39.33 (13.62)	0.001	40.34 (13.43)	38.29 (13.88)	<0.001
Political ideology	2.424 (1.419)	2.483 (1.458)	0.333	4.780 (1.568)	4.872 (1.528)	0.167
Affective polarization	4.784 (2.913)	4.844 (3.200)	0.641	5.103 (3.316)	5.210 (3.407)	0.453
Subjective closeness						
IOS11 to Democrat	7.267 (2.516)	7.362 (2.541)	0.369	2.700 (1.894)	2.671 (1.887)	0.710
IOS11 to Republican	2.483 (1.841)	2.518 (1.956)	0.656	7.804 (2.630)	7.881 (2.641)	0.490
IOS11 to Stranger	6.042 (2.634)	6.123 (2.684)	0.465	6.374 (2.804)	6.597 (2.840)	0.061
US regions						
West	0.236 (0.425)	0.215 (0.411)	0.244	0.191 (0.394)	0.164 (0.371)	0.108
Midwest	0.207 (0.405)	0.182 (0.386)	0.153	0.181 (0.385)	0.192 (0.394)	0.516
South	0.374 (0.484)	0.412 (0.492)	0.067	0.475 (0.500)	0.461 (0.499)	0.512
Northeast	0.183 (0.387)	0.190 (0.393)	0.681	0.153 (0.360)	0.183 (0.387)	0.065

Notes: The table reports means and standard deviations (in parentheses) for individual characteristics for Wave 2 and Wave 3 separated by political party affiliation. p-values are derived from two-sided Welch two sample t-tests for continuous variables and Pearson's chi-square tests for binary variables. Political ideology follows the 3-item index of Balliet et al. (2018). Affective polarization is the difference between the IOS11 score toward a co-partisan less the IOS11 score toward someone from the opposing party. The IOS11 score measures subjective closeness toward someone else and its implementation follows Baader et al. (2024).

Table SI-3: Impact of proximity to 2024 US presidential election on political polarization

	Dependent variable: Political Polarization					
	Ideological polarization			Affective polarization		
	(1)	(2)	(3)	(4)	(5)	(6)
Constant (Election Phase)	5.19***	4.39***	4.81***	4.98***	2.44***	4.54***
	(0.02)	(0.07)	(0.07)	(0.05)	(0.20)	(0.14)
Pre-election Phase	0.01	0.06	−0.01	−0.71***	−0.69***	−0.73***
	(0.04)	(0.04)	(0.04)	(0.09)	(0.09)	(0.09)
Post-inauguration Phase	−0.16***	−0.17***	−0.18***	−0.08	−0.02	−0.09
	(0.04)	(0.04)	(0.04)	(0.09)	(0.09)	(0.09)
Affective polarization		0.09***				
		(0.01)				
Ideological polarization					0.42***	
					(0.03)	
Democrat		0.72***	0.70***		−0.47***	−0.19**
		(0.03)	(0.03)		(0.07)	(0.07)
Age		0.003*	0.004**		0.01***	0.01***
		(0.001)	(0.001)		(0.003)	(0.003)
Male		−0.15***	−0.16***		−0.08	−0.18**
		(0.03)	(0.03)		(0.07)	(0.07)
Northeast		−0.07	−0.04		0.30**	0.29*
		(0.05)	(0.05)		(0.11)	(0.11)
South		−0.11**	−0.09*		0.31***	0.25**
		(0.04)	(0.04)		(0.09)	(0.09)
Midwest		0.03	0.05		0.23*	0.24*
		(0.05)	(0.05)		(0.11)	(0.11)
Observations	8,010	7,991	7,991	8,340	7,991	8,321
Adjusted R^2	0.002	0.10	0.06	0.01	0.05	0.01

Notes: The table reports coefficient estimates from linear regression models. The dependent variable is ideological polarization (Columns 1-3) and affective polarization (Columns 4-6). For comparability, we reverse-coded the 3-item political ideology index from Balliet et al. (2018) for Democrats, such that the value 7 captures the extreme positions, i.e., left/liberal for Democrats and right/conservative for Republicans. Affective polarization is the difference between the IOS11 score toward a co-partisan less the IOS11 score toward someone from the opposing party. The IOS11 score measures subjective closeness toward someone else and its implementation follows Baader et al. (2024). Robust standard errors are in parentheses. Levels of significance: *p<0.05, **p<0.01, ***p<0.001.

Table SI-4: Rule compliance overview

	Pre-election Phase (Wave 1)	Wave 2	Wave 3	Election Phase (Waves 2 & 3)	Post-inauguration Phase (Wave 4)	Pooled
Experimenter	55.69%	56.94%	59.03%	57.96%	59.73%	57.78%
Stranger	-	51.32%	56.48%	53.89%	57.30%	54.98%
Co-partisan	50.26%	50.68%	53.04%	51.85%	56.97%	52.60%
Political Opponent	50.18%	53.57%	52.06%	52.82%	53.42%	52.28%
Close	52.54%	53.05%	56.35%	54.68%	60.94%	55.68%
Distant	47.39%	50.38%	50.96%	50.67%	51.51%	50.33%
Co-partisan Close	58.70%	54.85%	55.13%	54.85%	58.33%	56.73%
Co-partisan Distant	42.67%	47.28%	51.28%	49.28%	55.20%	48.66%
Political Opponent Close	53.92%	55.86%	56.29%	56.08%	56.29%	55.55%
Political Opponent Distant	44.83%	51.11%	47.06%	49.14%	49.28%	48.14%
Stranger Close	-	53.43%	58.87%	56.17%	63.11%	58.64%
Stranger Distant	-	49.32%	54.09%	51.66%	49.33%	51.00%

Notes: The table reports mean rule compliance across settings, phases and waves (repeated cross-section).

Table SI-5: Overview of pairwise comparisons of rule compliance w.r.t. our treatments

Pre-election Phase			
Pearson's chi-square test for differences across treatments: p=.099			
Experimenter	Experimenter		
Co-partisan	p=.199	Co-partisan	
Political Opponent	p=.199	p=1	Political Opponent
Election Phase			
Pearson's chi-square test for differences across treatments: p=.007			
Experimenter	Experimenter		
Co-partisan	p=.01	Co-partisan	
Political Opponent	p=.031	p=.645	Political Opponent
Post-inauguration Phase			
Pearson's chi-square test for differences across treatments: p=.123			
Experimenter	Experimenter		
Co-partisan	p=.516	Co-partisan	
Political Opponent	p=.132	p=.516	Political Opponent
Pooled across all phases			
Pearson's chi-square test for differences across treatments: p<.001			
Experimenter	Experimenter		
Co-partisan	p<.001	Co-partisan	
Political Opponent	p<.001	p=.834	Political Opponent
Experimenter			
Pearson's chi-square test for differences across phases: p=.394			
Pre-election Phase	Pre-election Phase		
Election Phase	p=.761	Election Phase	
Post-inauguration Phase	p=.538	p=.761	Post-inauguration Phase
Co-partisan			
Pearson's chi-square test for differences across phases: p=.066			
Pre-election Phase	Pre-election Phase		
Election Phase	p=.543	Election Phase	
Post-inauguration Phase	p=.086	p=.11	Post-inauguration Phase
Political Opponent			
Pearson's chi-square test for differences across phases: p=.497			
Pre-election Phase	Pre-election Phase		
Election Phase	p=.898	Election Phase	
Post-inauguration Phase	p=.898	p=.898	Post-inauguration Phase

Notes: Pairwise comparisons between treatment conditions or phases are conducted using Fisher's exact test with Holm correction for multiple hypotheses.

Table SI-6: Overview of pairwise comparisons of rule compliance between Democrats and Republicans

Wave/Phase	Treatment	Democrats	Republicans	p-value
Wave 1/Pre-election	Experimenter	54.48%	56.99%	0.601
Wave 1/Pre-election	Co-partisan	52.01%	48.42%	0.433
Wave 1/Pre-election	Political Opponent	49.11%	51.24%	0.674
Wave 1/Pre-election	*pooled*	51.93%	52.18%	0.955
Wave 2	Experimenter	55.89%	58.06%	0.659
Wave 2	Co-partisan	47.08%	54.24%	0.099
Wave 2	Political Opponent	51.58%	55.64%	0.380
Wave 2	Stranger	49.48%	53.19%	0.422
Wave 2	*pooled*	51.03%	55.26%	0.047
Wave 3	Experimenter	61.01%	57.04%	0.388
Wave 3	Co-partisan	53.45%	52.63%	0.910
Wave 3	Political Opponent	50.69%	53.53%	0.559
Wave 3	Stranger	51.07%	61.84%	0.013
Wave 3	*pooled*	54.01%	56.29%	0.297
Election	Experimenter	58.36%	57.55%	0.830
Election	Co-partisan	50.26%	53.45%	0.304
Election	Political Opponent	51.13%	54.60%	0.272
Election	Stranger	50.26%	57.52%	0.017
Election	*pooled*	52.51%	55.77%	0.030
Wave 4/Post-inauguration	Experimenter	59.48%	60.00%	0.975
Wave 4/Post-inauguration	Co-partisan	63.27%	49.58%	0.002
Wave 4/Post-inauguration	Political Opponent	52.98%	53.91%	0.904
Wave 4/Post-inauguration	Stranger	57.80%	56.75%	0.874
Wave 4/Post-inauguration	*pooled*	58.41%	55.16%	0.148
pooled	Experimenter	57.60%	57.97%	0.895
pooled	Co-partisan	53.81%	51.32%	0.253
pooled	Political Opponent	51.07%	53.55%	0.263
pooled	Stranger	52.77%	57.28%	0.071
pooled	*pooled*	53.90%	54.88%	0.383

Notes: Pairwise comparisons are conducted using Pearson's chi-square tests.

Table SI-7: Overview of pairwise comparisons of rule compliance w.r.t. subjective closeness

Pre-election Phase			
Pearson's chi-square test for differences across treatments: p=.022			
Experimenter	Experimenter		
Close	p=.273	Close	
Distant	p=.019	p=.171	Distant
Election Phase			
Pearson's chi-square test for differences across treatments: p<.001			
Experimenter	Experimenter		
Close	p=.08	Close	
Distant	p<.001	p=.042	Distant
Post-inauguration Phase			
Pearson's chi-square test for differences across treatments: p<.001			
Experimenter	Experimenter		
Close	p=.681	Close	
Distant	p=.007	p<.001	Distant
Pooled across all phases			
Pearson's chi-square test for differences across treatments: p<.001			
Experimenter	Experimenter		
Close	p=.126	Close	
Distant	p<.001	p<.001	Distant
Close			
Pearson's chi-square test for differences across phases: p=.003			
Pre-election Phase	Pre-election Phase		
Election Phase	p=.355	Election Phase	
Post-inauguration Phase	p=.006	p=.008	Post-inauguration Phase
Distant			
Pearson's chi-square test for differences across phases: p=.313			
Pre-election Phase	Pre-election Phase		
Election Phase	p=.436	Election Phase	
Post-inauguration Phase	p=.436	p=.699	Post-inauguration Phase

Notes: Pairwise comparisons between settings or phases are conducted using Fisher's exact test with Holm correction for multiple hypotheses.

Table SI-8: Impact of rule setters on rule compliance separated by phases

	Dependent variable: Rule compliance					
	Pre-election Phase		Election Phase		Post-inauguration Phase	
	(1)	(2)	(3)	(4)	(5)	(6)
Constant (Experimenter)	0.56***	0.64***	0.58***	0.62***	0.60***	0.54***
	(0.02)	(0.07)	(0.01)	(0.04)	(0.02)	(0.06)
Co-partisan	−0.05	−0.04	−0.06**	−0.06**	−0.03	−0.03
	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)
Political Opponent	−0.06	−0.05	−0.05*	−0.05*	−0.06*	−0.08*
	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)
Stranger			−0.04	−0.03	−0.02	−0.04
			(0.02)	(0.02)	(0.03)	(0.03)
Affective Polarization		0.01		0.004		−0.003
		(0.004)		(0.002)		(0.003)
Political Ideology		−0.01		0.004		0.01
		(0.01)		(0.01)		(0.01)
Democrat		−0.03		−0.02		0.05
		(0.03)		(0.02)		(0.03)
Age		−0.0001		−0.001		0.0003
		(0.001)		(0.001)		(0.001)
Male		−0.03		−0.03*		−0.02
		(0.02)		(0.02)		(0.02)
Northeast		−0.04		−0.02		0.05
		(0.04)		(0.02)		(0.03)
South		−0.05		−0.01		0.01
		(0.03)		(0.02)		(0.03)
Midwest		−0.04		−0.01		−0.02
		(0.04)		(0.02)		(0.04)
Observations	1,727	1,664	4,540	4,362	2,073	1,965
Adjusted R^2	0.002	0.002	0.002	0.003	0.001	0.003

Notes: The table reports coefficient estimates from linear probability models. The dependent variable is rule compliance. Columns 1-2 display data from the Pre-election Phase. Columns 3-4 display data from the Election Phase. Columns 5-6 display data from the Post-inauguration Phase. Political ideology follows the 3-item index of Balliet et al. (2018). Affective polarization is the difference between the IOS11 score toward a co-partisan less the IOS11 score toward someone from the opposing party. Robust standard errors are in parentheses. Levels of significance: *p<0.05, **p<0.01, ***p<0.001.

Table SI-9: Impact of rule setters on rule compliance with phase dummies

	Dependent variable: Rule compliance					
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.58***	0.59***	0.58***	0.60***	0.58***	0.61***
	(0.01)	(0.03)	(0.01)	(0.03)	(0.01)	(0.03)
Co-partisan	−0.05***	−0.05**				
	(0.01)	(0.02)				
Political Opponent	−0.06***	−0.06***				
	(0.01)	(0.02)				
Stranger	−0.04*	−0.03				
	(0.02)	(0.02)				
Close			−0.02	−0.02		
			(0.01)	(0.01)		
Distant			−0.08***	−0.08***		
			(0.01)	(0.01)		
Co-partisan Close					−0.01	−0.01
					(0.02)	(0.02)
Co-partisan Distant					−0.09***	−0.09***
					(0.02)	(0.02)
Political Opponent Close					−0.02	−0.02
					(0.02)	(0.02)
Political Opponent Distant					−0.10***	−0.10***
					(0.02)	(0.02)
Stranger Close					−0.0001	0.004
					(0.02)	(0.02)
Stranger Distant					−0.07***	−0.07***
					(0.02)	(0.02)
Pre-election Phase	−0.02	−0.02	−0.03	−0.02	−0.02	−0.02
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Post-inauguration Phase	0.03*	0.03*	0.03*	0.03*	0.02	0.02
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)

< to be continued on the next page >

Table SI-9: Impact of rule setters on rule compliance with phase dummies (continued from previous page)

	Dependent variable: Rule compliance					
	(1)	(2)	(3)	(4)	(5)	(6)
Affective Polarization		0.002		0.002		0.001
		(0.002)		(0.002)		(0.002)
Political Ideology		0.004		0.003		0.003
		(0.004)		(0.004)		(0.004)
Democrat		−0.004		−0.003		−0.002
		(0.01)		(0.01)		(0.01)
Age		−0.0004		−0.0004		−0.0004
		(0.0004)		(0.0004)		(0.0004)
Male		−0.03**		−0.03**		−0.03**
		(0.01)		(0.01)		(0.01)
Northeast		−0.002		−0.001		−0.002
		(0.02)		(0.02)		(0.02)
South		−0.01		−0.01		−0.01
		(0.01)		(0.01)		(0.01)
Midwest		−0.02		−0.02		−0.02
		(0.02)		(0.02)		(0.02)
Observations	8,340	7,991	8,340	7,991	8,340	7,991
Adjusted R^2	0.003	0.003	0.005	0.01	0.01	0.01

Notes: The table reports coefficient estimates from linear probability models. The dependent variable is rule compliance. Political ideology follows the 3-item index of Balliet et al. (2018). Affective polarization is the difference between the IOS11 score toward a co-partisan less the IOS11 score toward someone from the opposing party. Robust standard errors are in parentheses. Levels of significance: *p<0.05, **p<0.01, ***p<0.001.

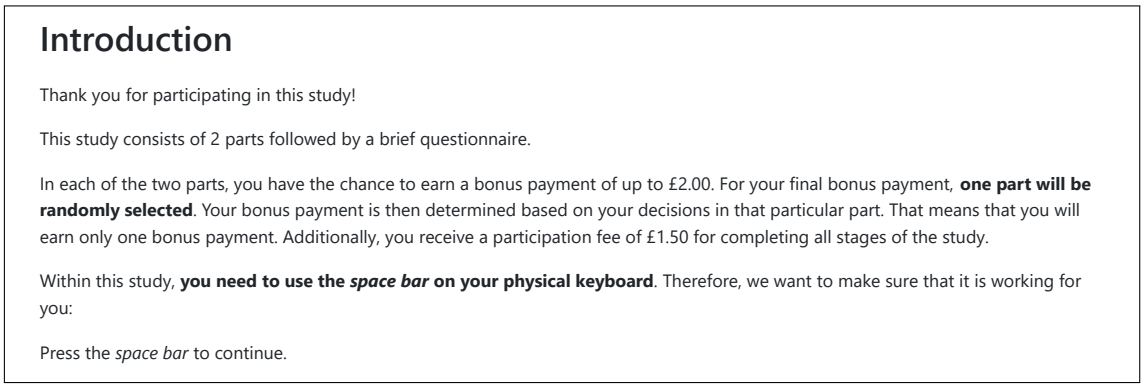
Table SI-10: Impact of subjective closeness to rule setter on rule compliance

	Dependent variable: Rule compliance		
	(1)	(2)	(3)
Constant (Election Phase)	0.45***	0.48***	−0.11***
	(0.02)	(0.04)	(0.03)
Subjective closeness to rule setter	0.01***	0.01***	−0.002
	(0.002)	(0.002)	(0.002)
Pre-election Phase	−0.02	−0.02	0.01
	(0.02)	(0.02)	(0.01)
Post-inauguration Phase	0.03	0.03	−0.01
	(0.02)	(0.02)	(0.01)
Political ideology	0.01	0.01	−0.003
	(0.003)	(0.004)	(0.004)
General rule-following attitude			0.12***
			(0.01)
Normative belief (rule compliance)			−0.002
			(0.01)
Normative belief (rule violation)			−0.05***
			(0.01)
Descriptive belief			0.005***
			(0.0003)
Democrat		0.003	0.04*
		(0.02)	(0.02)
Male		−0.03*	0.01
		(0.01)	(0.01)
Age		−0.0003	−0.001**
		(0.0005)	(0.0004)
Northeast		−0.003	−0.01
		(0.02)	(0.02)
Midwest		−0.02	−0.01
		(0.02)	(0.02)
South		−0.003	−0.02
		(0.02)	(0.02)
Observations	5,869	5,858	5,858
Adjusted R^2	0.01	0.01	0.21

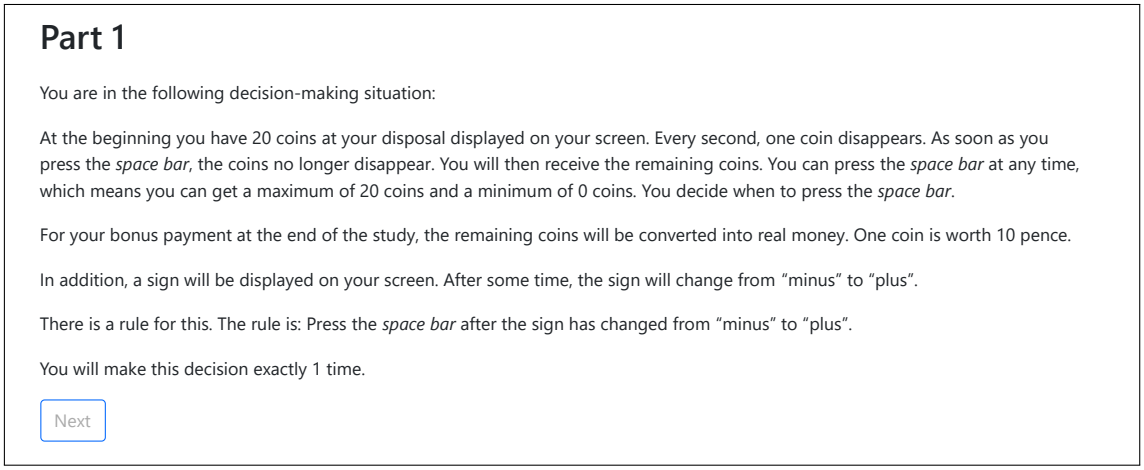
Notes: The table reports coefficient estimates from linear probability models. The dependent variable is rule compliance. No data from the EXPERIMENTER condition is used due to non-existence of subjective closeness measures to the rule setter in this condition. Political ideology follows the 3-item index of Balliet et al. (2018). Subjective closeness to the rule setter is implemented via the IOS11 score (Baader et al., 2024). Information about the beliefs is displayed in Section SI-1 and Section SI-3.2. Robust standard errors are in parentheses. Levels of significance: *p<0.05, **p<0.01, ***p<0.001.

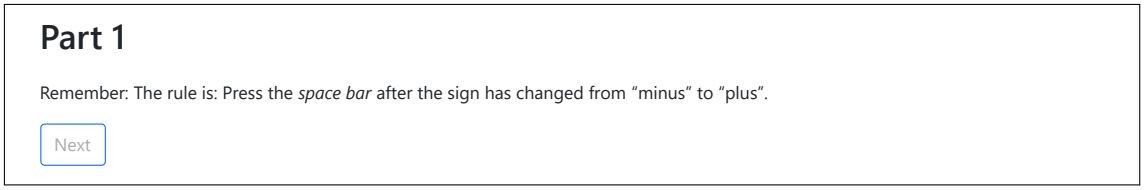
SI-3 Experimental Instructions

This section presents the experimental instructions as original screenshots shown to participants. Each frame indicates a new screen. Where appropriate, we add comments *** *in italic text between asterisks* ***. These comments were not shown to participants. Prior to the instructions shown below, we asked participants for their Prolific ID, to provide informed consent and checked whether they can use the space bar on a physical keyboard for conducting the coins task. After asking the participant for the reasoning behind their choice, we provided information on the payment as well as the completion code to be entered into Prolific.



SI-3.1 The coins task





*** *The content above was only shown in the Experimenter condition.* ***

Part 1

Remember: The rule is: Press the *space bar* after the sign has changed from "minus" to "plus".

Note: The rule was chosen by a Prolific user from the USA who defines themselves as Democrat.

Next

*** *The content above was only shown in the Democrat condition.* ***

Part 1

Remember: The rule is: Press the *space bar* after the sign has changed from "minus" to "plus".

Note: The rule was chosen by a Prolific user from the USA who defines themselves as Republican.

Next

*** *The content above was only shown in the Republican condition.* ***

Part 1

Remember: The rule is: Press the *space bar* after the sign has changed from "minus" to "plus".

Note: The rule was chosen by a Prolific user from the USA.

Next

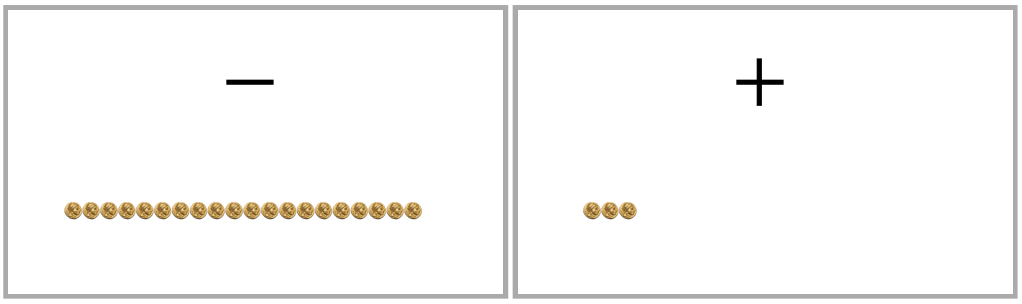
*** *The content above was only shown in the Stranger condition. This condition was not part of Wave 1.* ***

Part 1

Below you can see the previously described decision-making situation using two examples.

The left image shows the starting position. The sign is "minus" and there are 20 coins.

The right image shows a randomly chosen later position. The sign is "plus" and there are 3 coins.

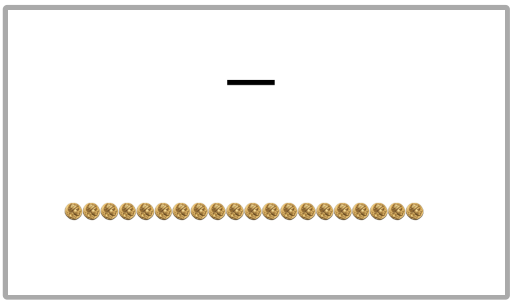


Next

Part 1

In the following, you will be asked to answer some short comprehension questions. We use these questions to check whether all participants have understood the decision-making situation.

Show instructions again



How many coins would you get if you pressed the *space bar* now?

How many coins would you get if you pressed the *space bar* now?

Who chose the rule?

- Republican
- Democrat

*** *In the content above, the third question was only shown in the Democrat and Republican conditions.* ***

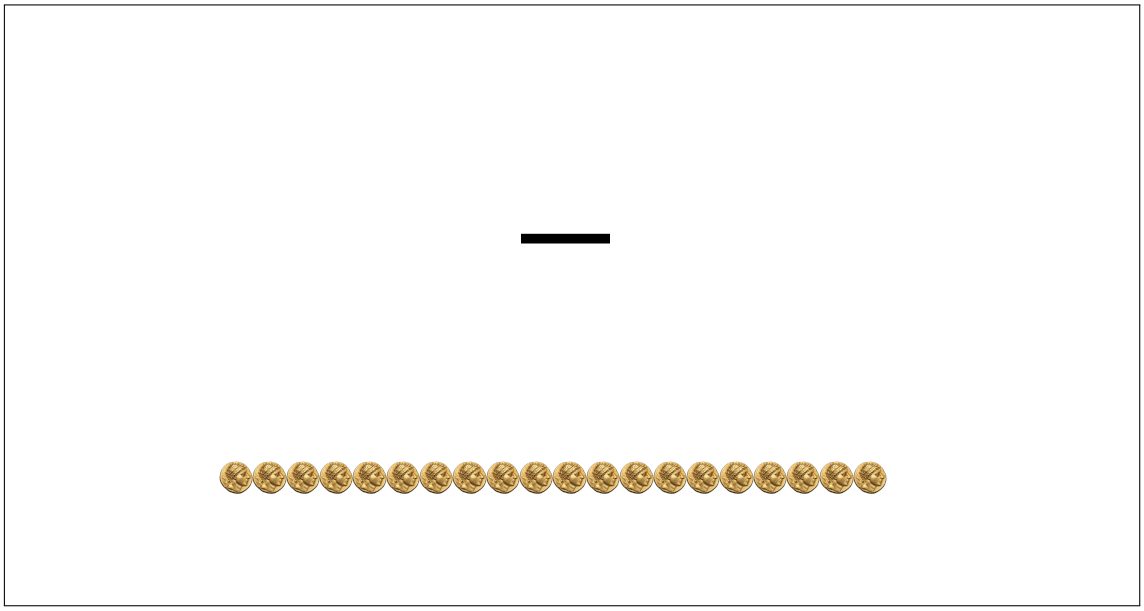
You have answered the comprehension questions correctly.

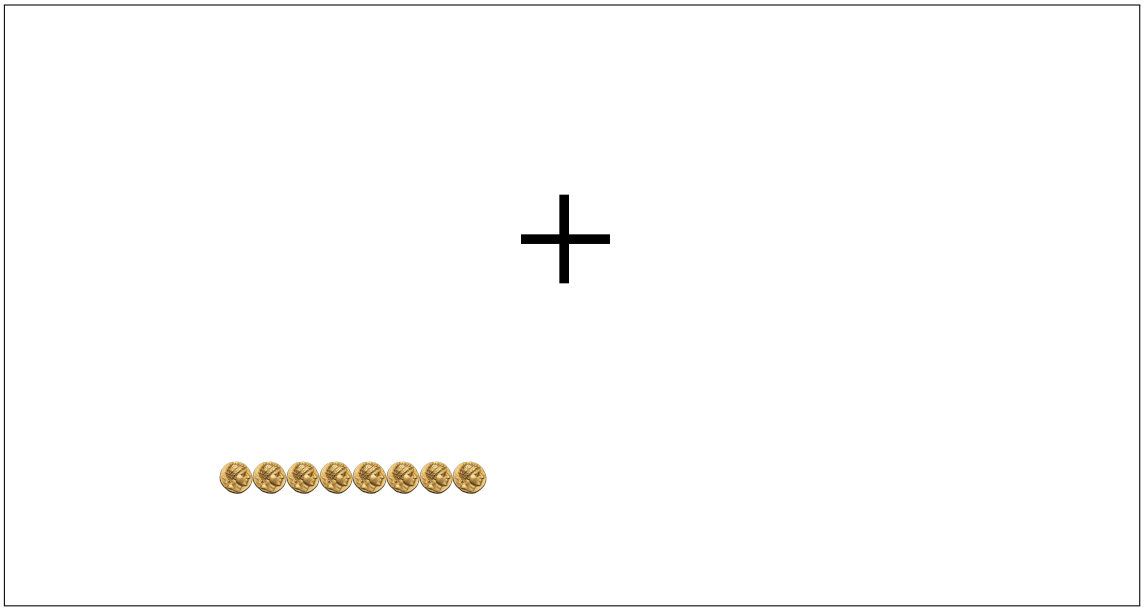
Press the *space bar* to start with the decision-making situation.

Remember: The rule is to press the *space bar* after the sign has changed from "minus" to "plus".

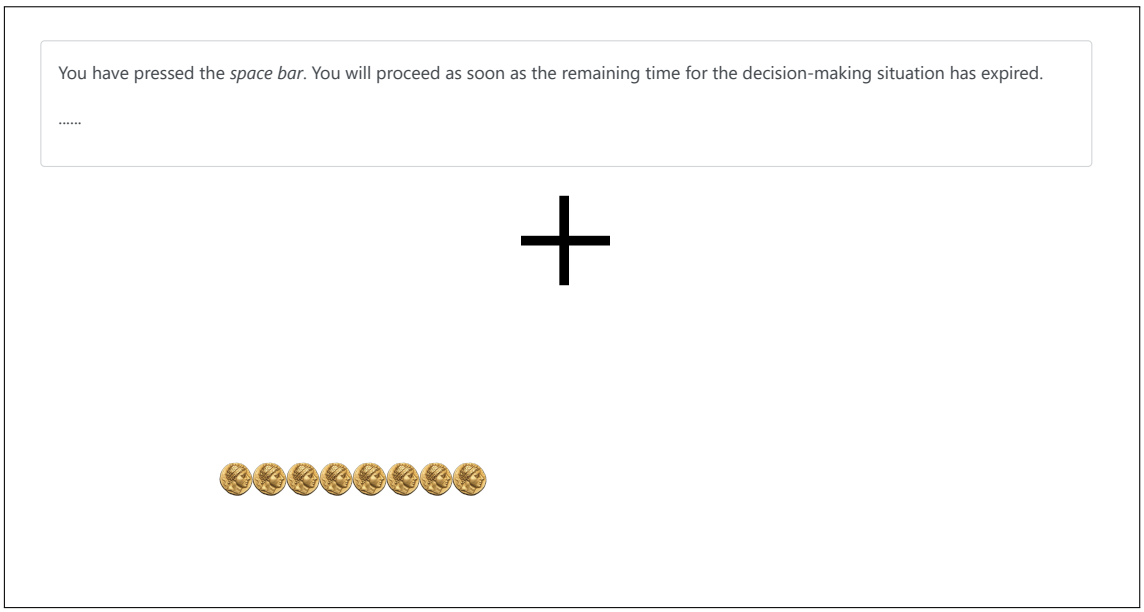
Note: The rule was chosen by a Prolific user from the USA who defines themselves as Democrat.

*** *In the content above, the sentence "Note: The rule was chosen by a Prolific user from the USA who defines themselves as Democrat." was only shown in the Democrat conditions. In the Republican condition, the sentence was "Note: The rule was chosen by a Prolific user from the USA who defines themselves as Republican.". In the Stranger condition, the sentence was "Note: The rule was chosen by a Prolific user from the USA.".* ***



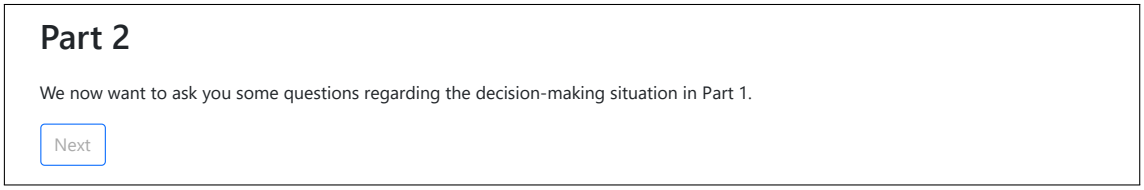


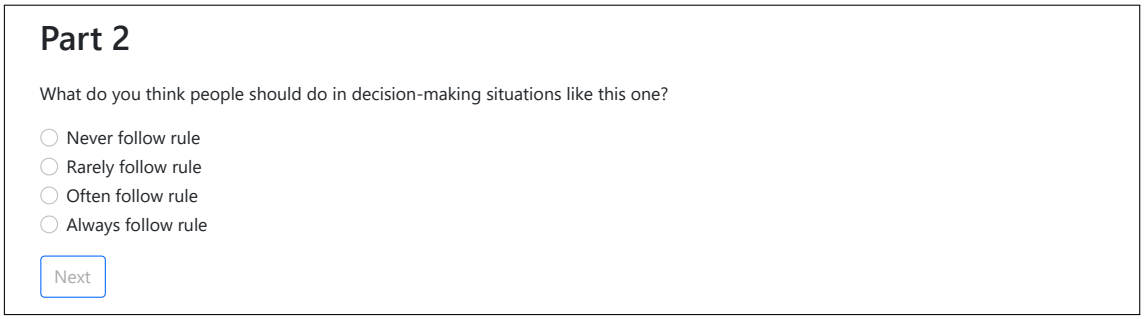
*** *In the content above, 12 seconds have already passed and the sign just changed. The participant has not pressed the space bar yet.* ***



*** *In the content above, 15 seconds have already passed. The sign change occurred after 12 seconds and the participant pressed the space bar directly after the sign change, thus no further coins disappear. A visual feedback indicates that the remaining seconds still continue to pass.* ***

SI-3.2 Belief elicitation





Part 2

The following two questions will ask how socially appropriate certain behavior is. By socially appropriate, we mean behavior that you think most other Prolific users from the USA would agree is the "correct" thing to do. Another way to think about what we mean is that if someone were to behave in a socially inappropriate way, then other people might be angry at them.

Your answers can influence your **bonus payment**. In order to earn this bonus payment, you need to identify for each of the two questions the answer that was given the most often by Prolific users from the USA who have participated in a similar study before. They faced the same decision-making situation with the same instructions including the same rule. For each correct answer, you can earn £0.50 as bonus payment.

Assume that a Prolific user from the USA **had followed** the rule in Part 1. How do you think most of the other Prolific users from the USA mentioned above would view this?

- ○ very socially inappropriate
- ○ somewhat socially inappropriate
- ○ somewhat socially appropriate
- ○ very socially appropriate

Assume that a Prolific user from the USA **had not followed** the rule in Part 1. How do you think most of the other Prolific users from the USA mentioned above would view this?

- ○ very socially inappropriate
- ○ somewhat socially inappropriate
- ○ somewhat socially appropriate
- ○ very socially appropriate

Next

Part 2

There were other Prolific users from the USA who have participated in a similar decision-making situation as described in Part 1. They had faced the same decision-making situation with the same instructions including the same rule. What do you think, how many out of these have followed the rule?

Your answer can influence your **bonus payment**. You can earn up to a £1.00 as a bonus payment the closer your answer is to the correct result. You can earn a maximum of £1.00 and a minimum of £0.00.

Show details on the calculation of the bonus payment

What do you think, how many have followed the rule?

0.0%

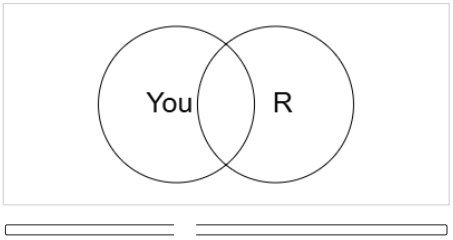
Note: For a higher precision of the slider, first click on it and then move the slider with your right and left arrow keys. The *next* button appears once you have clicked on the slider.

SI-3.3 IOS11

Questionnaire

Once you move the slider below, a pair of circles will appear in the box. The position of the slider will determine the extent to which the circles overlap. You should interpret the degree of overlap as representing the relationship between you and "R". "R" serves as a placeholder for Prolific users from the USA who define themselves as **Republicans**.

Please position the slider so that the circles indicate to what extent you and "R" are connected.



Note: The *next* button appears once you have clicked on the slider.

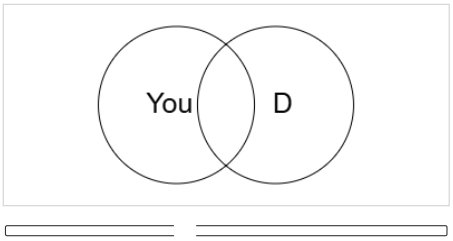
Next

*** *In the content above, the slider has been already clicked.* ***

Questionnaire

Once you move the slider below, a pair of circles will appear in the box. The position of the slider will determine the extent to which the circles overlap. You should interpret the degree of overlap as representing the relationship between you and "D". "D" serves as a placeholder for Prolific users from the USA who define themselves as **Democrats**.

Please position the slider so that the circles indicate to what extent you and "D" are connected.



Note: The *next* button appears once you have clicked on the slider.

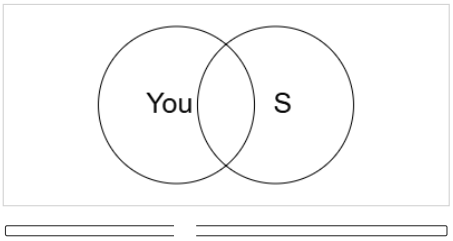
Next

*** *In the content above, the slider has been already clicked.* ***

Questionnaire

Once you move the slider below, a pair of circles will appear in the box. The position of the slider will determine the extent to which the circles overlap. You should interpret the degree of overlap as representing the relationship between you and "S". "S" serves as a placeholder for **Prolific users from the USA**.

Please position the slider so that the circles indicate to what extent you and "S" are connected.



Note: The *next* button appears once you have clicked on the slider.

Next

*** *In the content above, the slider has been already clicked. This question was not part of Wave 1.* ***

SI-3.4 Questionnaire

Questionnaire

Attention Check Question

Recall the decision-making situation from the beginning. Which sign change should occur after some time?

- from ? to !
- from 0 to 1
- from - to +
- from < to >

Next

Questionnaire

Socio-Demographic Information

How old are you (in years)?

Which gender do you identify with?

- ○ Female
- ○ Male
- ○ Other
- ○ Prefer not to answer

In which state do you live?

Are you generally an impatient person, or someone who always shows great patience? (0 = very impatient, 10 = very patient)

○ 0 ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10 ○ Prefer not to answer

Which political party do you identify with?

- ○ Democrats
- ○ Republicans
- ○ Other
- ○ Prefer not to answer

How would you describe your political orientation? (1 = extremely left, 7 = extremely right)

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ Prefer not to answer

Please indicate your level of agreement with the following two statements:

When it comes to politics, I consider myself politically conservative. (1 = strongly agree, 7 = strongly disagree)

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ Prefer not to answer

When it comes to politics, I consider myself politically liberal. (1 = strongly agree, 7 = strongly disagree)

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ Prefer not to answer

Next

*** *In the content above, the question on patience was not part of Wave 1.* ***

Questionnaire

Why did you decide to follow the rule in the decision-making situation in Part 1?

Next

SI-4 Supplementary References

Baader, M., Starmer, C., Tufano, F., & Gächter, S. (2024). Introducing IOS11 as an extended interactive version of the 'Inclusion of Other in the Self' scale to estimate relationship closeness. *Scientific Reports*, *14*(1), 8901. https://doi.org/10.1038/s41598-024-58042-6.

Balliet, D., Tybur, J. M., Wu, J., Antonellis, C., & Van Lange, P. A. M. (2018). Political Ideology, Trust, and Cooperation: In-group Favoritism among Republicans and Democrats during a US National Election. *Journal of Conflict Resolution*, *62*(4), 797–818. https://doi.org/10.1177/0022002716658694.