

Ehret, Sönke; Constantino, Sara M.; Weber, Elke U.; Efferson, Charles; Vogt, Sonja

**Working Paper**

## Group Identities Make Fragile Tipping Points

CESifo Working Paper, No. 9737

**Provided in Cooperation with:**

Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

*Suggested Citation:* Ehret, Sönke; Constantino, Sara M.; Weber, Elke U.; Efferson, Charles; Vogt, Sonja (2022) : Group Identities Make Fragile Tipping Points, CESifo Working Paper, No. 9737, Center for Economic Studies and ifo Institute (CESifo), Munich

This Version is available at:

<https://hdl.handle.net/10419/260867>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

## Group Identities Make Fragile Tipping Points

*Sönke Ehret, Sara M. Constantino, Elke U. Weber, Charles Efferson, Sonja Vogt*

## **Impressum:**

CESifo Working Papers

ISSN 2364-1428 (electronic version)

Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH

The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute

Poschingerstr. 5, 81679 Munich, Germany

Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email [office@cesifo.de](mailto:office@cesifo.de)

Editor: Clemens Fuest

<https://www.cesifo.org/en/wp>

An electronic version of the paper may be downloaded

- from the SSRN website: [www.SSRN.com](http://www.SSRN.com)
- from the RePEc website: [www.RePEc.org](http://www.RePEc.org)
- from the CESifo website: <https://www.cesifo.org/en/wp>

# Group Identities Make Fragile Tipping Points

## Abstract

Social tipping can accelerate beneficial changes in behaviour in diverse domains from equality and social justice to climate change. Hypothetically, however, group identities might undermine tipping in ways policy makers do not anticipate. To examine this, we implemented an experiment around the 2020 U.S. elections. Participants faced consistent incentives to coordinate their choices. Once participants had established a coordination norm, an intervention created pressure to tip to a new norm. Our control treatment used neutral labels for choices. Our identity treatment used partisan political images. This simple payoff-irrelevant relabelling generated extreme differences. Control groups developed norms slowly before intervention but transitioned to new norms rapidly after intervention. Identity groups developed norms rapidly before intervention but persisted in a state of costly disagreement after intervention. Tipping was powerful but fragile. It supported striking cultural changes when choices and identity were unlinked, but even a trivial link destroyed tipping entirely.

JEL-Codes: Z100, Z130, Z180.

Keywords: social tipping, cultural evolution, behaviour change, coordination.

*Sönke Ehret\**  
*Faculty of Business and Economics*  
*University of Lausanne / Switzerland*  
*sonke.ehret@gmail.com*

*Sara M. Constantino\**  
*Princeton University*  
*Princeton / NJ / USA*  
*sara.constantino@gmail.com*

*Elke U. Weber*  
*Princeton University*  
*Princeton / NJ / USA*  
*eweber@princeton.edu*

*Charles Efferson\*\**  
*Faculty of Business and Economics*  
*University of Lausanne / Switzerland*  
*charles.efferson@unil.ch*

*Sonja Vogt\*\**  
*Faculty of Business and Economics*  
*University of Lausanne / Switzerland*  
*sonja.vogt@unil.ch*

\* Shared first authorship

\*\* These authors jointly supervised the research

25 April 2022: For helpful comments, we thank Heinrich Nax, Nikos Nikiforakis, Simon Siegenthaler, Paul Smaldino, Christian Zehnder, and seminar participants at the University of Lausanne, NYUAD, Oxford, and the University of Zurich. We thank the Swiss National Science Foundation (Nr. 100018 185417/1) for support.

## 1 Introduction

Social change can stagnate for long periods of time, and then suddenly it can unfold quickly and often unexpectedly. Binding the feet of girls and women persisted in China for centuries, only to disappear in a generation<sup>1</sup>. In the United States, longstanding attitudes towards same-sex marriage went from hostile to accepting in a few years<sup>2</sup>. Germany began subsidising solar panels on private homes in the 1990s. Social interactions among friends and neighbours soon accelerated the spread of the new technology, and residents of Germany were generating more solar power per capita than people in any other country by 2016<sup>3</sup>.

This kind of punctuated cultural change can occur when a population tips from one social norm to another<sup>1,4</sup>. Social tipping of this sort is a flamboyant form of cultural evolution in which people change en masse in terms of both how they behave and how they think about the behaviour of others<sup>5</sup>. Social tipping has generated enormous interest as an efficient means of triggering beneficial changes in behaviour<sup>6</sup> in a wide range of domains from public health<sup>7,8</sup> and social justice<sup>9,10</sup> to resource conservation<sup>11,12</sup> and climate change<sup>13–15</sup>. Given the widespread appeal of tipping, researchers and practitioners have a responsibility to investigate the conditions that might support or undermine tipping<sup>16–18</sup>. We do so here by examining group identities as a specific psychological mechanism hypothesised to interfere with tipping dynamics<sup>19–21</sup>.

Importantly, proof of concept exists for tipping. Observational data show that cultural evolutionary processes can support multiple distinct norms, and punctuated cultural change certainly occurs<sup>1,22–26</sup>. Recent experimental studies have also demonstrated that interventions can spark rapid transitions from one norm to another<sup>5,27</sup>. Nonetheless, empirical and theoretical studies of tipping in relation to gender-based violence<sup>10,28–30</sup>, political revolutions<sup>31</sup>, and lab experiments<sup>5</sup> suggest that our understanding of when tipping is possible and how to activate tipping is fundamentally limited<sup>18</sup>. The associated risk is that policy makers misinvest in poorly designed or pointless interventions centred around tipping in settings with little or no potential for social tipping to support behaviour change.

In settings that do have this potential, tipping dynamics hold clear implications for policy makers<sup>6</sup>. The key idea is that the same mechanisms contribute to both the slow and fast phases of punctuated cultural evolution<sup>1,22</sup>. Specifically, a psychological tendency to conform and incentives to coordinate one's choices with others motivate people to behave

like those around them. If a given behaviour is rare, conformity and coordination keep it rare. This is the slow phase. If the behaviour becomes sufficiently common, for whatever reason, conformity and coordination switch from obstructing to accelerating the diffusion of the behaviour in question. This is the fast phase. The upshot is that, once sufficient change occurs, the population crosses a tipping point and quickly transitions to a new cultural regime.

The policy maker's task is to initiate this social dynamic. When conformity and coordination incentives reinforce a status quo norm inconsistent with policy objectives, a policy maker can choose to promote an alternative norm. Alternative norms might include the abandonment of female genital cutting<sup>32</sup>, choosing not to smoke<sup>33</sup>, favouring electric cars over gas<sup>6</sup>, and eating chicken instead of pangolin<sup>12</sup>. To promote an alternative, the policy maker targets a subset of the population with an intervention that incentivises the policy maker's preferred behaviour. Interventions can take many forms ranging from taxes and subsidies<sup>5</sup> to entertaining narratives with educational messaging<sup>34–36</sup>.

If enough people in the targeted subset change behaviour, the population may cross a tipping point. If so, conformity and coordination switch from supporting the status quo choice to supporting the policy maker's alternative. This specifically means that individuals not exposed to the intervention change their evaluation of the choices available. They see many others changing behaviour, the harbinger of a new norm, and they conclude that some alternative choice has become preferable to the old status quo. Once this happens, the population should complete the transition to a new norm quickly, even if the policy maker has moved on to some other problem. Behaviour change is partly exogenous, because some people change their behaviour after direct exposure to the intervention, and partly endogenous, because of the effects of conformity and coordination after the population crosses the tipping point. In short, the direct effects of the intervention spill over and indirectly influence the choices of those never exposed to the intervention themselves.

Tipping thus represents an efficient use of resources because these spillovers imply that endogenous social forces produce much of the behaviour change. This is especially important given that many contemporary social problems are daunting in scale<sup>6,12</sup>. Moreover, promoting widespread social change is an attempt to engineer culture, and even policy makers with the purest of intentions cannot escape the practical and ethical dilemmas this implies. To the extent that tipping activates endogenous social forces, change originates from within the population. The hope is that this moderates concerns about

paternalistic intrusions in a society's culture and the associated risk of backlash<sup>9,20</sup>.

The challenge is that conformity and coordination incentives do not generally operate in isolation. Rather, they interact with many other social and psychological motives<sup>5, 18, 37</sup>, and these motives are often organised around group identities. The need to belong to a group often drives people to signal group membership and emphasise conformity to the ingroup<sup>38</sup>. Moreover, people often experience positive affect towards markers of ingroup affiliation and the values these markers represent, together with negative affect towards outgroup markers and associated values<sup>39, 40</sup>. When these affective responses become linked to policy-relevant behaviours, group identities might moderate tipping in situations where tipping would otherwise support socially beneficial change.

We hypothesized that group identities induce important forms of heterogeneity that can undermine tipping and limit socially beneficial changes in behaviour. Broadly speaking, heterogeneity may or may not hinder tipping based on a number of details like, for example, the distribution of preferences in the population and variation in how strongly people respond to the information about the choices of others<sup>5, 20, 37, 41–47</sup>. People differ in many dimensions critical to behaviour change, and these differences shape the potential to activate tipping. Heterogeneity based specifically on group identities holds particular interest because human psychology has a strong parochial streak, arguably based on an evolutionary history of ingroup cooperation and coordination combined with outgroup hostility<sup>40, 48, 49</sup>. Models suggest that this kind of heterogeneity can have an outsized influence on cultural evolutionary dynamics and limit the potential for tipping<sup>19–21</sup>.

To capture the intuition behind this theoretical notion, imagine a population that is subdivided into two groups. Each group has its own norm in some domain. For example, perhaps one group practises female genital cutting, and the other does not. A policy maker steps in to promote a specific norm for the entire population, in this example a norm based around not cutting, and by extension the policy maker's preference is inconsistent with the norm in one of the two groups. Depending on how strongly diverse group members link cutting to being a valued member of the group, the policy maker's efforts can represent more than simple measures that promote a specific behaviour. They can also represent a kind of existential threat to one's identity as a group member. In extreme cases, the policy maker's efforts can even strengthen the tie between traditional practice and group identity. This would mean, for example, that the policy maker adds value to cutting for members of the cutting group and actually increases their resistance to behaviour change<sup>50</sup>. Our study,

though not about cutting, focuses on these kinds of situations in which a policy maker’s intervention promotes a norm that is inconsistent with group identity.

To examine the hypothesised influence of group identities, we implemented an incentivised online experiment in the time surrounding the contentious 2020 election for President of the United States. A U.S. sample participated in repeated play of coordination games. We designed our control treatment to be maximally favourable for tipping and rapid transitions from one norm to another. The experimental treatment was identical with one exception. We relabelled choice options with images designed to activate partisan political identities (Fig. 1). Partisan loyalties provide an important component of identity in contemporary U.S. politics<sup>51,52</sup>, and party affiliation has increasingly become a matter of strong sectarian emotions based on ingroup favouritism and outgroup derogation<sup>53,54</sup>. Crucially, our partisan images had no explicit consequences in terms of material incentives. They simply provided a labelling system for choice options (Supplementary Fig. 2), and in this sense our treatment manipulation was minimal.

## **Experimental Design**

Regardless of treatment, all sessions had the following common structure. Based on results from a survey conducted shortly before the experiment, we created groups of 12 participants who were either all Republicans or all Democrats (Methods). Participants played a symmetric two-player coordination game with random rematching each period with members of their own group. Participants were anonymous and unable to communicate. Thus, to coordinate consistently they had to establish a group norm via repeated play with feedback. In addition to private feedback related to one’s own choices, we provided public feedback at the beginning of each period by sharing the distribution of choices in the previous period among 10 randomly selected group members (Methods). To minimise any sense of shared group identity before participants actually started the experiment, we did not tell participants they were in a group with other supporters of the same party.

Each session had a pre-intervention phase and a post-intervention phase. The pre-intervention phase lasted between 10 and 20 periods based on how long it took to establish an initial coordination norm (Methods). Payoffs simply favoured coordinating; they did not favour coordinating on any specific option (Table 1a). Once a group had established a norm, we implemented an intervention (Methods). We targeted a random sample of par-



ticipants, typically half of the group, and we changed payoffs for these participants. The new payoffs favored choosing the alternative option regardless of a partner’s choice (Table 1b; see Supplementary Information, Section 4.2, for why the intervention was strong). Participants targeted by the intervention faced this new incentive structure for the remainder of the session. Non-targeted participants retained their original payoffs (Table 1c), but we told them that payoffs had changed for others. Groups continued to play with random rematching for another 25 periods after the intervention. Behaviour change after intervention was socially beneficial in the sense that, if everyone in a group would have adopted the behaviour favoured by the intervention, no one would have experienced a decline in payoffs, and some would have experienced a strict increase. As explained in greater detail below, we refer to the norm in place just before the intervention as the “status quo” norm, and we refer to the behaviour promoted by the intervention as the “alternative”.

Behaviour change happens for two reasons in settings of this sort. First, targeted individuals may change behaviour because the intervention directly incentivises them to do so. Second, all individuals in the group interact, and all face incentives to coordinate their choices. If potential game partners change behaviour, a focal player, whether targeted or not, may follow along as she observes others abandoning the status quo for the alternative. This second effect represents the central idea behind policy applications of tipping points, the idea that social interactions within a society can accelerate and amplify transitions to new norms.

Our experiment consisted of two treatments in a between-subjects design. In the **neutral** treatment (35 groups), choice options in the game had neutral symbols, namely @ and #, as labels. In the **identity** treatment (33 groups), choice options had one of two images as labels. These images were a drawing of a victorious Joe Biden sitting on Donald Trump and a drawing of a victorious Donald Trump sitting on Joe Biden (Fig. 1). Neutral symbols and political images were simply labels for the two choice options in the sense that they were embedded in the on-screen buttons participants had to press to make a choice (Supplementary Figs. 2 and 3). They had no other role or significance in the experiment.

Importantly, our intervention created heterogeneity within groups. After intervention, targeted participants faced material incentives that made the alternative choice dominant (Table 1b) and thus clearly favoured behaviour change. Non-targeted participants faced material incentives that simply favoured behaving like others. In this material sense,

targeted individuals after intervention favoured the equilibrium associated with the alternative choice, while non-targeted individuals were indifferent over equilibria. Material incentives were heterogeneous, but in a way that supported behaviour change. The use of neutral symbols in our neutral treatment ensured that identity concerns were irrelevant.

The use of political labels in our identity treatment added a second currency of potential value to the two choice options and thus a second potential source of heterogeneity. To illustrate the possible effects of political labels, imagine a group in the identity treatment that converges on choosing the partisan image consistent with the party affiliation of group members. Imagine a group of Democrats, for example, who initially converge on choosing the image of a victorious Biden. In this case, the intervention incentivises targeted participants to change to the image of a victorious Trump. Given this scenario, what happens after intervention? The answer should depend on how participants trade money against identity concerns.

First, consider an extreme group in which money dominates identity concerns for everyone. In this case, all targeted participants should change behaviour because the intervention introduces material incentives that strongly favour doing so. To continue with our example group of Democrats, all targeted participants should switch to the choice associated with the Trump sitting on Biden label. Once this happens, all non-targeted participants should also change behaviour. Indeed, experimental results<sup>27</sup> suggest that, if targeted participants change behaviour, interventions of the size we implemented are easily large enough to induce non-targeted participants to follow. More generally, if we consider only monetary incentives, the long-run behaviour change in the neutral treatment should be the same as in the identity treatment. This outcome should hold because we are considering the case in which money dominates identity for everyone, and labels have no monetary consequences.

Alternatively, consider an extreme group in which identity concerns dominate money for everyone. In this case, behaviour change should not occur at all in the identity treatment. Because the intervention involves material incentives only, targeted players should maintain the pre-intervention norm consistent with their identity as the post-intervention phase unfolds. In particular, targeted players should not incur the identity-based cost of changing to the alternative behaviour. Non-targeted players should be similar because we are assuming that identity dominates money for everyone, and non-targeted players only differ from targeted players in terms of monetary incentives. Finally, consider a group

in which people trade money against identity concerns in heterogeneous ways, a case between the two extremes. This kind of heterogeneity implies that some players in the identity treatment might change behaviour, others might not, and the result might ultimately be no norm at all.

## Results

To analyse behaviour change in our experiment with a common framework across sessions, we need to distinguish between the status quo norm in place before intervention and the alternative norm promoted by the intervention. The pre-intervention phase lasted a minimum of 10 periods. Subject to this constraint, the pre-intervention phase ended when at least 90% of the group chose the same option in a period or when the group had played 20 periods, whichever came first. Although not all groups reached the 90% threshold by period 20, all groups had a well-defined majority choice in place by the end of the pre-intervention phase. We treat the option chosen by the majority as the status quo behaviour in a group. Intuitively, the status quo was a descriptive norm that held when the intervention occurred.

In neutral sessions, some groups converged on a status quo norm of choosing @, while other groups converged on #. The status quo norm was unrelated to the political affiliations of the groups ( $\chi^2(1, N = 35) = 2.08, p = 0.15$ ). In identity sessions, although the same kind of flexibility was theoretically possible, in practice all Republican groups converged on victorious Trump as the status quo norm, while all Democrat groups converged on victorious Biden. With this definition of the status quo in place, the alternative behaviour in a group was simply the choice option that did not emerge as a norm before intervention and was thus favoured by the intervention (Table 1).

We begin with pre-registered analyses of spillovers (Methods), where spillovers are changes in aggregate behaviour that cannot be accounted for by the intervention<sup>20</sup>. We define spillovers as a normalised measure of how the long-run distribution of behaviours in a group deviates from the size of the intervention (Methods). Negative spillovers are in  $[-1, 0)$ , and they arise when the proportion choosing the alternative behaviour in the long run is less than the proportion of group members targeted by the intervention. Positive spillovers are in  $(0, 1]$  and arise when the long-run proportion choosing the alternative behaviour is larger than the proportion of group members targeted by the intervention. A

spillover of zero means the long-run proportion choosing the alternative is exactly the same as the proportional size of the intervention. Spillovers measure net aggregate endogenous change, where endogenous change can work against the intervention, amplify the effects of the intervention, or have no net effect.

Spillovers were large and highly significantly positive in our neutral treatment. In contrast, our identity treatment produced a large and highly significant reduction in spillovers relative to this benchmark (Fig. 2 and Table 2). Indeed, spillovers were not significantly different from zero in our identity treatment (Table 2 linear combination, Intercept + Identity = 0,  $F(1, 66) = 1.7$ , 95% CI =  $[-0.38, 0.06]$ ,  $p = 0.20$ , Cohen's  $f = 0.16$ ). A core principle associated with tipping points is that they reflect settings in which a tendency for people to behave like others can dramatically amplify the effects of some event that sets behaviour change in motion. This happened in our neutral treatment. Conditions were highly conducive to spillovers, and amplification reached 69% of the maximum conceivable value (Table 2, Intercept). However, the labelling of choice options in ways that misaligned behaviour change with group identity concerns destroyed spillovers entirely (Table 2, Identity).

To examine the underlying reasons for this striking difference in aggregate outcomes, we turn to individual decisions. Specifically, in a given period a participant could choose the status quo behaviour or the alternative behaviour. Based on one's own choice and the choice of one's partner, the outcomes possible included coordinating on the status quo, coordinating on the alternative, or miscoordinating.

Compared to neutral labels, political labels dramatically facilitated coordination before intervention (Fig. 3). In the neutral treatment, 23% of groups spent the full 20 periods of the pre-intervention phase without reaching the 90% threshold for triggering the intervention, and the pre-intervention phase lasted at least 15 periods in 43% of groups. In stark contrast, all groups in the identity treatment reached the 90% threshold by period 10. The political labels thus acted as focal points that supported coordination<sup>55,56</sup>. In effect, the pre-intervention game involved material incentives that favoured neither of the two pure-strategy equilibria, and so players faced an equilibrium-selection problem. Neutral labels did not help with this problem, and groups simply had to develop idiosyncratic local norms via repeated play with feedback. Political labels provided group members with a shared pre-existing basis for ranking the equilibria, and this allowed groups to pick an equilibrium with minimal fuss.

In addition, just as surely as political labels facilitated coordination before intervention, they hindered coordination after intervention. When the intervention was introduced, groups in the neutral condition immediately started changing behaviour, and the alternative behaviour was overwhelmingly dominant in most groups by the end of the post-intervention phase (Fig. 3a). Groups with political labels, however, persisted in a state of chronic disagreement after intervention. In any given period some players chose the status quo behaviour, some chose the alternative, and miscoordination was common and persistent (Fig. 3b).

To examine these treatment differences in detail, we used pre-registered models of individual choices (Methods) in the final periods of the pre-intervention and post-intervention phases (Table 3; see Supplementary Information, Section 2, for additional analyses). Under neutral labels before intervention, targeted and non-targeted participants exhibited the same tendency to choose the alternative behaviour (Table 3, Model 1, (Neutral,T,Pre-int)). This validates random assignment to the targeted subset, which happened at the beginning of sessions and thus well before intervention. Under political labels, both targeted (Table 3, Model 1, (Identity,T, Pre-int)) and non-targeted participants (Table 3, Model 1, (Identity,NT, Pre-int)) showed highly significant reductions in the probability of choosing the alternative behaviour relative to non-targeted participants in the neutral treatment before intervention. This result confirms that political labels facilitated coordination in the pre-intervention phase. Analogous to the neutral treatment, targeted and non-targeted participants in the identity treatment exhibited the same choices on average before intervention (Table 3, Model 1 linear combination, (Identity,NT,Pre-int) - (Identity,T,Pre-int) = 0,  $F(1, 1538) = 0.004$ , 95% CI =  $[-0.02, 0.02]$ ,  $p = 0.95$ , Cohen's  $f = 0.001$ ).

Post-intervention, both targeted (Table 3, Model 1, (Neutral,T,Post-int)) and non-targeted (Table 3, Model 1, (Neutral,NT,Post-int)) participants in the neutral treatment exhibited an increased probability of choosing the alternative behaviour relative to non-targeted participants in the neutral treatment before intervention. Targeted players showed a larger increase than non-targeted players (Table 3, Model 1 linear combination, (Neutral, NT,Post-int) - (Neutral,T,Post-int) = 0,  $F(1, 1538) = 16.27$ , 95% CI =  $[-0.26, -0.09]$ ,  $p < 0.001$ , Cohen's  $f = 0.10$ ), but the large and highly significant increase among non-targeted players demonstrates the power of endogenous social interactions to amplify the effects of a delimited intervention. Targeted participants in the identity treatment also exhibited highly significant changes in behaviour (Table 3, Model 1, (Identity,T,Post-int)) in

the wake of the intervention, but the effect was weaker than it was among targeted participants in the neutral treatment (Table 3, Model 1 linear combination, (Neutral,T,Post-int) - (Identity,T,Post-int) = 0,  $F(1, 1538) = 29.45$ , 95% CI = [0.18, 0.37],  $p < 0.001$ , Cohen's  $f = 0.14$ ). These results suggest that targeted participants in the identity treatment varied in terms of how they traded money against identity concerns. For some, switching to the alternative choice in the identity treatment was sufficiently aversive to prevent behaviour change, but for others this was not the case.

Strikingly, non-targeted participants in the identity treatment exhibited little behaviour change after intervention. These participants chose the alternative behaviour at a rate that was indistinguishable from non-targeted participants in the neutral treatment before intervention (Table 3, Model 1, (Identity,NT,Post-int)). In particular, they were highly significantly less likely to choose the alternative behaviour than their non-targeted counterparts in neutral sessions (Table 3, Model 1 linear combination, (Neutral,NT,Post-int) - (Identity,NT,Post-int) = 0,  $F(1, 1538) = 46.98$ , 95% CI = [0.39, 0.70],  $p < 0.001$ , Cohen's  $f = 0.17$ ) and their targeted counterparts in identity sessions (Table 3, Model 1 linear combination, (Identity,NT,Post-int) - (Identity,T,Post-int) = 0,  $F(1, 1538) = 108.04$ , 95% CI = [0.36, 0.53],  $p < 0.001$ , Cohen's  $f = 0.26$ ).

Altogether, these results show that social tipping contributed strongly to changes in behaviour under neutral labels, but tipping was essentially absent under political labels. Social tipping dynamics provided a spectacularly powerful route to behaviour change in the neutral treatment, but they proved to be equivalently fragile in the identity treatment. This difference also had stark consequences for participant payoffs. In our neutral treatment, players took a relatively long time to take advantage of the opportunities available to them pre-intervention. The absence of a focal point<sup>56</sup> meant that players needed time to develop idiosyncratic local norms to coordinate. These same groups, however, were able to transition rapidly to an alternative welfare-improving norm when circumstances changed. Indeed, tipping and its payoff consequences are plain to see when inspecting the rapid increase in payoffs that immediately followed intervention in neutral sessions (Fig. 4a). Exactly the opposite happened, however, in the identity treatment. Political labels provided players with a way to establish coordination quickly before intervention. However, with the appearance of a socially beneficial alternative that ran counter to the group's identity, players were completely unable to respond as a group. In the long run, group members accumulated substantial opportunity costs as a result (Fig. 4b).

Because we ran the study from late October through mid-December 2020, we were also able to analyse if and how choice and tipping dynamics changed before and after November 7, the day major news networks called the election for Joe Biden. We had no preregistered hypotheses about the effects of the election, but multiple possibilities exist. For example, one might imagine that after the election the actual outcome of the election would have provided all participants with a shared focal point<sup>56</sup> rooted in reality. If so, all groups in the identity treatment, whether Democrat or Republican, would have converged before intervention on victorious Biden, an especially compelling possibility given that we did not tell participants they were in groups with others of the same party. In addition, one might imagine that after the election participants in the identity treatment would have been more willing to change their behaviour after intervention. With the election settled, participants might have been less likely to interpret choosing a specific partisan image as an endorsement of the associated election result. This, in turn, might have allowed participants to disinvest in the images and simply treat them as a labelling system to facilitate coordination and make money. Alternatively, however, one might imagine that the conclusion of the election could have actually exacerbated outgroup animosity, with the winners gloating and the losers defensive. In this case, behaviour change in the identity treatment should have declined after the election because protecting group identities would have increased in importance relative to earning money.

None of this happened. As was true before the election, all groups in the identity treatment converged before intervention on the image consistent with the group's party affiliation. In particular, Republican groups did not use the actual outcome of the election as a new focal point; they continued to converge initially on the option labelled with the image of a victorious Trump. Moreover, when comparing before versus after the election, the average tendencies to choose the alternative behaviour conditional on treatment, targeted status, and pre- versus post-intervention were virtually identical (Table 3, Models 1 and 2).

Interestingly and quite surprisingly, however, we have some evidence that Republicans exhibited a reduced willingness to change behaviour in the neutral treatment after the election. Specifically, before the election, non-targeted Republicans in the neutral treatment showed a highly significant tendency to switch to the alternative behaviour after intervention. In effect, the strong spillovers that occurred in the neutral treatment (Fig. 2a) occurred in both Democrat groups and Republican groups (Supplementary Table 2). After the election, however, non-targeted Republicans in the neutral treatment showed a signif-

icant reduction in this tendency to switch to the alternative behaviour after intervention (Supplementary Table 2). One speculative explanation for this finding is that Republican participants, grappling with the loss of their candidate, exhibited a kind of post-election spite<sup>57</sup>. This spite manifested itself as a general resistance to change even when identity concerns were not an explicit part of the experiment.

This kind of result, however, was not true in general. In particular, targeted Republicans in the neutral treatment showed the same tendency to change behaviour before and after the election (Supplementary Table 2). Though a speculative interpretation, these results for Republicans in the neutral treatment suggest that post-election spite may have been strong enough to reduce endogenous behaviour change among non-targeted participants, but it was not enough to overcome the monetary incentives favouring change among targeted participants. As explained above, it was also not enough to create any detectable average difference in choices before versus after the election (Table 3, Models 1 and 2). We leave this as an unexpected paradox that illustrates how group affiliations can shape social tipping in subtle and surprising ways.

## Discussion

The results of our study show that even a seemingly superficial link between group identity and decision making can completely restructure cultural evolutionary dynamics and undermine social tipping that would otherwise occur. When group identities are linked to specific choices in the policy domain in question, the link adds implicit value to the status quo choice for some or all individuals in the population. This implicit value constrains behaviour change in general and endogenous spillovers due to social tipping specifically. Our results demonstrate how easily this can happen, and they suggest that, when group identity concerns are active, the policy maker might consider the value of an intervention before the intervention<sup>20</sup>.

The first intervention would aim to weaken the link between group identities and choice options in the policy domain at hand. If this works, the first intervention would thus lay the groundwork for the intervention proper. With identity concerns less relevant because of the first intervention, the intervention proper would then promote the alternative norm of primary interest. CNN adopted exactly this approach within a single ad about face masks during the Covid-19 pandemic (link). The ad first attempted to decouple masks



from the partisan baggage they had acquired in the U.S. in the early days of the pandemic. It began with a photo of a mask and said, “This is a mask. It prevents the spread of coronavirus. This is not a political statement. It’s a mask.” The ad then moved on to its primary behavioural objective and simply concluded with, “Please wear a mask.”

A more nuanced approach centres around strategies that attempt to transfer identity concerns from the policy-relevant domain of interest to some other domain. For example, a number of initiatives promoting the abandonment of female genital cutting involve attempts to provide girls and their families with alternative rites of passage<sup>58</sup>. The logic is that these alternative rites of passage provide families a way to integrate their daughters in society, and to signal this integration to others, without the harm of genital cutting. In effect, the hope is that families become increasingly willing to abandon cutting if they have substitute rituals that allow them to pass along the norms and values of their culture to their daughters.

In broad terms, we have shown that social tipping can offer a powerful but exceedingly fragile route to social change. This combination presents policy makers with an unusual challenge. Because tipping has such impressive potential, strategies to provoke tipping should and presumably will remain a part of the policy maker’s repertoire. However, because tipping is fragile, interventions designed to trigger tipping may easily fail to do so. The objective for researchers and practitioners is to develop a detailed and empirically grounded understanding of when tipping is possible and how to spark tipping to confront major social, economic, and environmental challenges.

Group identities, and the ingroup-outgroup distinctions they induce, are universal phenomena<sup>39</sup> reinforced by socio-technological innovations that encourage polarization along political, religious, and ethnic lines<sup>54</sup>. Group identities can have positive consequences<sup>59</sup>. They can even help people solve important coordination problems in diverse populations<sup>38</sup>. They can also, however, inhibit the effects of policy efforts designed to change cultural norms in ways that increase public welfare. An improved understanding of when and how group identities influence social tipping would allow for the design of interventions that appropriately consider the effects of identity concerns as we all confront the numerous formidable challenges facing contemporary human societies.

1. Young, H. P. The Evolution of Social Norms. *Annual Review of Economics* **7**, 359–387 (2015). URL <https://doi.org/10.1146/annurev-economics-080614-115322>.
2. Rosenfeld, M. J. Moving a Mountain: The Extraordinary Trajectory of Same-Sex Marriage Approval in the United States. *Socius* **3**, 2378023117727658 (2017). URL <https://doi.org/10.1177/2378023117727658>.
3. Rode, J. & Weber, A. Does localized imitation drive technology adoption? a case study on rooftop photovoltaic systems in germany. *Journal of Environmental Economics and Management* **78**, 38–48 (2016).
4. Bowles, S. *Microeconomics: Behavior, Institutions, and Evolution* (Princeton University Press, 2009).
5. Andreoni, J., Nikiforakis, N. & Siegenthaler, S. Predicting social tipping and norm change in controlled experiments. *Proceedings of the National Academy of Sciences* **118** (2021).
6. Nyborg, K. *et al.* Social norms as solutions. *Science* **354**, 42–43 (2016). URL <https://science.sciencemag.org/content/354/6308/42>. Publisher: American Association for the Advancement of Science Section: Policy Forum.
7. Christakis, N. A. & Fowler, J. H. The Spread of Obesity in a Large Social Network over 32 Years. *New England Journal of Medicine* **357**, 370–379 (2007). URL <https://doi.org/10.1056/NEJMs066082>.
8. Arnot, M. *et al.* How evolutionary behavioural sciences can help us understand behaviour in a pandemic. *Evolution, Medicine, and Public Health* **2020**, 264–278 (2020).
9. Cloward, K. *When Norms Collide: Local Responses to Activism against Female Genital Mutilation and Early Marriage* (Oxford University Press). URL <https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780190274917.001.0001/acprof-9780190274917>.
10. Platteau, J.-P., Camilotti, G. & Auriol, E. Eradicating women-hurting customs. *Towards gender equity in development* **319** (2018).
11. Castilla-Rho, J. C., Rojas, R., Andersen, M. S., Holley, C. & Mariethoz, G. Social tipping points in global groundwater management. *Nature Human Behaviour* **1**, 640–649 (2017).

12. Travers, H., Walsh, J., Vogt, S., Clements, T. & Milner-Gulland, E. Delivering behavioural change at scale: What conservation can learn from other fields. *Biological Conservation* **257**, 109092 (2021).
13. Barrett, S. & Dannenberg, A. Sensitivity of collective action to uncertainty about climate tipping points. *Nature Climate Change* **4**, 36–39 (2014).
14. Kopp, R. E., Shwom, R. L., Wagner, G. & Yuan, J. Tipping elements and climate–economic shocks: Pathways toward integrated assessment. *Earth’s Future* **4**, 346–372 (2016).
15. Otto, I. M. *et al.* Social tipping dynamics for stabilizing Earth’s climate by 2050. *Proceedings of the National Academy of Sciences* **117**, 2354–2365 (2020). URL <https://www.pnas.org/content/117/5/2354>.
16. Bicchieri, C. & Dimant, E. Nudging with care: The risks and benefits of social information. *Public choice* 1–22 (2019).
17. Smith, S. R., Christie, I. & Willis, R. Social tipping intervention strategies for rapid decarbonization need to consider how change happens. *Proceedings of the National Academy of Sciences* **117**, 10629–10630 (2020).
18. Efferson, C. Policy to activate cultural change to amplify policy. *Proceedings of the National Academy of Sciences* **118** (2021).
19. Smaldino, P. E., Janssen, M. A., Hillis, V. & Bednar, J. Adoption as a social marker: Innovation diffusion with outgroup aversion. *The Journal of Mathematical Sociology* **41**, 26–45 (2017).
20. Efferson, C., Vogt, S. & Fehr, E. The promise and the peril of using social influence to reverse harmful traditions. *Nature Human Behaviour* **4**, 55–68 (2020). URL <https://www.nature.com/articles/s41562-019-0768-2>. Number: 1 Publisher: Nature Publishing Group.
21. Smaldino, P. E. & Jones, J. H. Coupled dynamics of behaviour and disease contagion among antagonistic groups. *Evolutionary Human Sciences* **3** (2021).
22. Henrich, J. Cultural Transmission and the Diffusion of Innovations: Adoption Dynamics Indicate That Biased Cultural Transmission Is the Predominate Force in Behavioral Change. *American Anthropologist* **103**, 992–1013 (2001).

URL <https://anthrosource.onlinelibrary.wiley.com/doi/abs/10.1525/aa.2001.103.4.992>.

23. Young, H. P. & Burke, M. A. Competition and custom in economic contracts: a case study of illinois agriculture. *American Economic Review* **91**, 559–573 (2001).
24. Rogers, E. M. *Diffusion of Innovations* (Simon and Schuster, 2010).
25. Eugster, B., Lalive, R., Steinhauer, A. & Zweimüller, J. The demand for social insurance: does culture matter? *The Economic Journal* **121**, F413–F448 (2011).
26. Eugster, B., Lalive, R., Steinhauer, A. & Zweimüller, J. Culture, work attitudes, and job search: Evidence from the swiss language border. *Journal of the European Economic Association* **15**, 1056–1100 (2017).
27. Centola, D., Becker, J., Brackbill, D. & Baronchelli, A. Experimental evidence for tipping points in social convention. *Science* **360**, 1116–1119 (2018). URL <https://science.sciencemag.org/content/360/6393/1116>. Publisher: American Association for the Advancement of Science Section: Report.
28. Bellemare, M. F., Novak, L. & Steinmetz, T. L. All in the family: Explaining the persistence of female genital cutting in West Africa. *Journal of Development Economics* **116**, 252–265 (2015).
29. Muthukrishna, M. Cultural evolutionary public policy. *Nature Human Behaviour* **4**, 12–13 (2020).
30. Novak, L. Persistent norms and tipping points: The case of female genital cutting. *Journal of Economic Behavior & Organization* **177**, 433–474 (2020).
31. Kuran, T. Now out of never: The element of surprise in the east european revolution of 1989. *World Politics: A Quarterly Journal of International Relations* 7–48 (1991).
32. Shell-Duncan, B. & Hernlund, Y. *Female “circumcision” in Africa: culture, controversy, and change* (Lynne Rienner Publishers, 2000).
33. Christakis, N. A. & Fowler, J. H. The Collective Dynamics of Smoking in a Large Social Network. *New England Journal of Medicine* **358**, 2249–2258 (2008). URL <https://doi.org/10.1056/NEJMSa0706154>.
34. DellaVigna, S. & La Ferrara, E. Economic and social impacts of the media. In *Handbook of Media Economics*, vol. 1, 723–768 (Elsevier, 2015).

35. La Ferrara, E. Mass media and social change: Can we use television to fight poverty? *Journal of the European Economic Association* **14**, 791–827 (2016).
36. Vogt, S., Zaid, N. A. M., Ahmed, H. E. F., Fehr, E. & Efferson, C. Changing cultural attitudes towards female genital cutting. *Nature* **538**, 506–509 (2016).
37. Granovetter, M. Threshold models of collective behavior. *American Journal of Sociology* **83**, 1420–1443 (1978).
38. Efferson, C., Lalive, R. & Fehr, E. The Coevolution of Cultural Groups and In-group Favoritism. *Science* **321**, 1844–1849 (2008). URL <https://science.sciencemag.org/content/321/5897/1844>.
39. Tajfel, H. *Human groups and social categories: Studies in social psychology* (Cup Archive, 1981).
40. De Dreu, C. K., Gross, J., Fariña, A. & Ma, Y. Group cooperation, carrying-capacity stress, and intergroup conflict. *Trends in Cognitive Sciences* (2020).
41. Young, H. P. Innovation diffusion in heterogeneous populations: Contagion, social influence, and social learning. *American Economic Review* **99**, 1899–1924 (2009).
42. Goeree, J. K. & Yariv, L. Conformity in the lab. *Journal of the Economic Science Association* **1**, 15–28 (2015).
43. Mesoudi, A., Chang, L., Dall, S. R. & Thornton, A. The evolution of individual and cultural variation in social learning. *Trends in Ecology & Evolution* **31**, 215–225 (2016).
44. Muthukrishna, M., Morgan, T. J. & Henrich, J. The when and who of social learning and conformist transmission. *Evolution and Human Behavior* **37**, 10–20 (2016).
45. Neary, P. R. & Newton, J. Heterogeneity in preferences and behavior in threshold models. *Journal of Mechanism and Institution Design* **2**, 1 (2017).
46. Kendal, R. L. *et al.* Social learning strategies: Bridge-building between fields. *Trends in Cognitive Sciences* **22**, 651–665 (2018).
47. Gavrillets, S. The dynamics of injunctive social norms. *Evolutionary Human Sciences* **2** (2020).

48. Choi, J.-K. & Bowles, S. The coevolution of parochial altruism and war. *Science* **318**, 636–640 (2007).
49. Handley, C. & Mathew, S. Human large-scale cooperation as a product of competition between cultural groups. *Nature Communications* **11**, 1–9 (2020).
50. Thomas, L. M. ‘Ngaitana (I will circumcise myself)’: Lessons from colonial campaigns to ban excision in Meru, Kenya (2000).
51. Iyengar, S., Sood, G. & Lelkes, Y. Affect, Not Ideology: A Social Identity Perspective on Polarization. *Public Opinion Quarterly* **76**, 405–431 (2012). URL <https://doi.org/10.1093/poq/nfs038>.
52. Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N. & Westwood, S. J. The Origins and Consequences of Affective Polarization in the United States. *Annual Review of Political Science* **22**, 129–146 (2019). URL <https://doi.org/10.1146/annurev-polisci-051117-073034>.  
\_eprint: <https://doi.org/10.1146/annurev-polisci-051117-073034>.
53. McConnell, C., Margalit, Y., Malhotra, N. & Levendusky, M. The Economic Consequences of Partisanship in a Polarized Era. *American Journal of Political Science* **62**, 5–18 (2018). URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/ajps.12330>.  
\_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/ajps.12330>.
54. Finkel, E. J. *et al.* Political sectarianism in America. *Science* **370**, 533–536 (2020). URL <https://science.sciencemag.org/content/370/6516/533>. Publisher: American Association for the Advancement of Science Section: Policy Forum.
55. Schelling, T. C. *The Strategy of Conflict* (Harvard University Press, 1960).
56. Crawford, V. P., Gneezy, U. & Rottenstreich, Y. The power of focal points is limited: Even minute payoff asymmetry may yield large coordination failures. *American Economic Review* **98**, 1443–58 (2008).
57. Levin, S. A. Public goods in relation to competition, cooperation, and spite. *Proceedings of the National Academy of Sciences* **111**, 10838–10845 (2014). URL <http://www.pnas.org/cgi/doi/10.1073/pnas.1400830111>.

58. Droy, L. *et al.* Alternative rites of passage in FGM/C abandonment campaigns in Africa: a research opportunity. *LIAS Working Paper Series* **1** (2018).
59. Chen, Y. & Li, S. X. Group identity and social preferences. *American Economic Review* **99**, 431–57 (2009).
60. Druckman, J. N., Klar, S., Krupnikov, Y., Levendusky, M. & Ryan, J. B. Affective polarization, local contexts and public opinion in America. *Nature Human Behaviour* **5**, 28–38 (2021). URL <https://www.nature.com/articles/s41562-020-01012-5>. Number: 1 Publisher: Nature Publishing Group.
61. Mason, L. Ideologues without Issues: The Polarizing Consequences of Ideological Identities. *Public Opinion Quarterly* **82**, 866–887 (2018). URL <https://doi.org/10.1093/poq/nfy005>.

**Acknowledgements** The study was funded by the Swiss National Science Foundation (Nr. 100018\_185417/1 to CE and SV). The funding agency played no role in the design of the study, data collection, data analysis and interpretation, or the writing and submission of the paper.

**Author Contributions** All authors designed the study. SE programmed the main experiment. SE and SV worked with a free-lance artist to develop the images of Biden and Trump. SE and SC pre-tested the images, ran initial surveys to identify partisan commitments, and ran the experimental sessions. SE, SC, and CE analysed the data. All authors interpreted the results. SE, SC, and CE wrote the paper with input from the other authors. SE, SC, and SV wrote the Supplementary Information with input from the other authors.

**Competing Interests** The authors declare that they have no competing financial interests.

**Data/code availability** We will make the data and code freely and publicly available at the time of publication. For requests before that time, please contact the corresponding authors.

**Correspondence** Correspondence should be addressed to [sonkeklaus.ehret@unil.ch](mailto:sonkeklaus.ehret@unil.ch), [sara.constantino@gmail.com](mailto:sara.constantino@gmail.com), [charles.efferson@unil.ch](mailto:charles.efferson@unil.ch), and [sonja.vogt@unil.ch](mailto:sonja.vogt@unil.ch).

## 2 Methods

**Participants.** We conducted the study with adult participants living in the U.S. between October 28 and December 16, 2020. The study was approved by the Institutional Review

Boards at the University of Lausanne, the University of Bern, and Princeton University. All participants provided informed consent.

We recruited participants online via Prolific Academic. We screened potential participants based on their self-reported political affiliations and responses to two questions about political preferences. Specifically, in an initial recruitment survey, we elicited responses to questions about Joe Biden and Donald Trump using a feeling thermometer<sup>51,52,60,61</sup>, and we used these responses to recruit participants to the main study. For the main study, we formed groups of either all partisan Republicans or all partisan Democrats. All sessions began with groups of size 12, and we relied on a number of protocols to minimise participant dropout during sessions. The Supplementary Information (Sections 5 and 6.1) provides additional details and analyses related to recruitment, sample composition, and dropout.

**Coordination game and treatments.** Regardless of treatment, participants repeatedly played coordination games for points in groups of 10 or 12 for up to 45 periods. Points were converted to dollars at a fixed rate at the end of a session, and the total payoff for each participant was calculated by summing over payoffs from five randomly selected periods. Participants were informed about payment and other procedures before the start of the game.

Players were anonymous, were not informed about the shared partisan commitments within their groups, and were unable to communicate with each other. We randomly paired players from the group in each period. Sessions were divided into a pre-intervention and a post-intervention phase. In the pre-intervention phase, everyone played the same symmetric coordination game (Table 1a). The intervention, however, introduced an important source of heterogeneity in the group by applying a new payoff matrix to a subset of players, while the remaining players retained their original incentives (Table 1b-1c). The intervention was randomly assigned to 50% of players in each group at the start of the session (Supplementary Information, Section 3.3). Because assignment to the targeted subset occurred at the beginning of sessions, occasional dropouts before intervention meant that the targeted subset sometimes consisted of 40% or 60% of the group instead of 50% (see Supplementary Information, Section 6.3, for associated robustness checks).

To provide feedback about evolving social dynamics, participants saw the following information at the start of each period except the first: the complete distribution of choices



in the previous period for 10 randomly selected group members, their partner’s choice in the previous period, and the points they earned in the previous period. All groups began with 12 participants. We were able to continue with a session even if someone dropped out without disturbing our feedback protocol because we provided feedback each period by randomly selecting 10 participants in the group. Because participants played in pairs, when one player dropped out we removed the player’s counterpart from that period, but only after the counterpart had entered a choice for the period in question. This resulted in some periods with 11 responses, and more broadly group sizes ranged from 10 to 12. If more than two players exited the group, for whatever reason, we ended the session. Dropouts were not systematically related to treatment (Supplementary Information, Section 6.1). Each period, participants indicated their choices by clicking an on-screen button that was integrated with the display of the player’s payoff matrix. Neutral or political labels were embedded in the buttons themselves (Supplementary Figs. 2 and 3). Apart from this difference in the on-screen buttons used to make choices, the treatments were identical.

**Analyses.** The initial data consisted of 28,303 observations from 908 participants in 77 groups. We removed nine groups that, due to dropouts, did not have at least one period in the post-intervention phase. This left 27,624 observations from 805 participants in 68 groups. Analyses were pre-registered on the Open Science Framework ([osf.io/6adbx](https://osf.io/6adbx)) unless otherwise indicated.

Table 2 presents an analysis of spillovers<sup>20</sup>, which we define as a normalised measure of net endogenous behaviour change at the level of the group. Specifically, let  $\phi_j$  be the proportion of decision makers in group  $j$  targeted by the intervention. Let  $\hat{q}_j$  be the proportion of decision makers choosing the alternative behaviour in the final period of the post-intervention phase. Spillovers in group  $j$  are defined as  $\Theta_j = [\hat{q}_j > \phi_j](\hat{q}_j - \phi_j)/(1 - \phi_j) + [\phi_j \geq \hat{q}_j](\hat{q}_j - \phi_j)/\phi_j$ , where  $[\cdot]$  are Iverson brackets. Thus, a positive spillover signifies a group in which the final effect of the intervention is larger than the proportional size of the intervention. A negative spillover signifies the opposite (Supplementary Information, Section 2.3).

Model 1 of Table 3 models the probability that an individual chooses the alternative behaviour for the group in question in the final periods of the pre-intervention and post-intervention phases. Focusing on the final periods minimises the role of transient dynamics. It instead focuses on the key idea behind policy applications of social tipping, namely the idea that tipping relates specifically to a transition between pure-strategy equilibria,

one consistent with policy objectives and the other inconsistent<sup>20</sup>. The Supplementary Information (Supplementary Table 2) includes robustness checks based on analyses over more periods.

Model 2 of Table 3 is an exploratory analysis identical to Model 1 except that it distinguishes between sessions conducted before and after the election was called. Although perhaps not immediately obvious, the right-hand sides of Models 1 and 2 are equivalent to difference-in-difference estimations with added distinctions between targeted and non-targeted participants (Models 1 and 2) and between pre- and post-election sessions (Model 2). In contrast to a typical difference-in-difference specification, we coded the right-hand side by forming a set of mutually exclusive dummy variables defined jointly over (i) the treatment (Neutral vs. Identity), (ii) whether the participant was targeted (T) or not (NT), and (iii) whether an observation was before (Pre-int) or after (Post-int) intervention. This coding allowed us to avoid three-way interactions in Model 1 and four-way interactions in Model 2.

**Table 1 | Participant payoffs.** Matrices show row player payoffs in points as a function of row and column choices. The status quo (SQ) choice was the choice associated with the norm that emerged before intervention in a session. Given a status quo choice, the alternative (Alt) was simply the other choice option. **a**, Payoffs were the same for everyone in the pre-intervention phase and did not favour any particular equilibrium. **b**, The intervention encouraged behaviour change by introducing new payoffs that favoured the alternative choice among targeted (T) players, and these payoffs held for the entire post-intervention phase. **c**, Non-targeted (NT) players retained their original payoffs post-intervention.

	<b>(a) Pre-int (all)</b>		<b>(b) Post-int (T)</b>		<b>(c) Post-int (NT)</b>	
	SQ	Alt	SQ	Alt	SQ	Alt
SQ	200	50	200	50	200	50
Alt	50	200	350	350	50	200

**Table 2 | Spillovers by treatment.** Spillovers<sup>20</sup> take values in  $[-1, 1]$  and provide a normalised measure of long-run socially beneficial changes in behaviour at the group level while accounting for the size of the intervention (Methods). Results are from an OLS regression that models spillovers as a function of treatment (Fig. 2). Robust standard errors are in parentheses. Spillovers were highly significantly positive in the neutral treatment (Intercept), and relabelling choice options in the identity treatment resulted in a large and highly significant reduction in beneficial spillovers.

	Spillovers
Intercept	0.69*** (0.07)
Identity	-0.82*** (0.12)

The  $p$  values are based on two-sided (Gaussian) z-tests.

\*  $p \in (0.01, 0.05]$

\*\*  $p \in (0.001, 0.01]$

\*\*\*  $p \leq 0.001$

**Table 3 | Participant chooses the alternative behaviour.** Linear probability models for individual choices in the final periods of the pre-intervention and post-intervention phases. Cluster-robust standard errors are clustered at the group level. Election is a dummy indicating sessions after 7 November 2020, which was the day major news networks called the election for Joe Biden. Composite dummies are defined jointly over (i) whether a group was in the neutral or identity treatment, (ii) whether the participant was targeted (T) or not (NT) by the intervention, and (iii) whether the period was the final period of the pre-intervention phase or of the post-intervention phase. (Neutral,NT,Pre-int) is the omitted category. Model 1 was pre-registered. Model 2 is exploratory and additionally distinguishes between before (omitted category) and after the election.

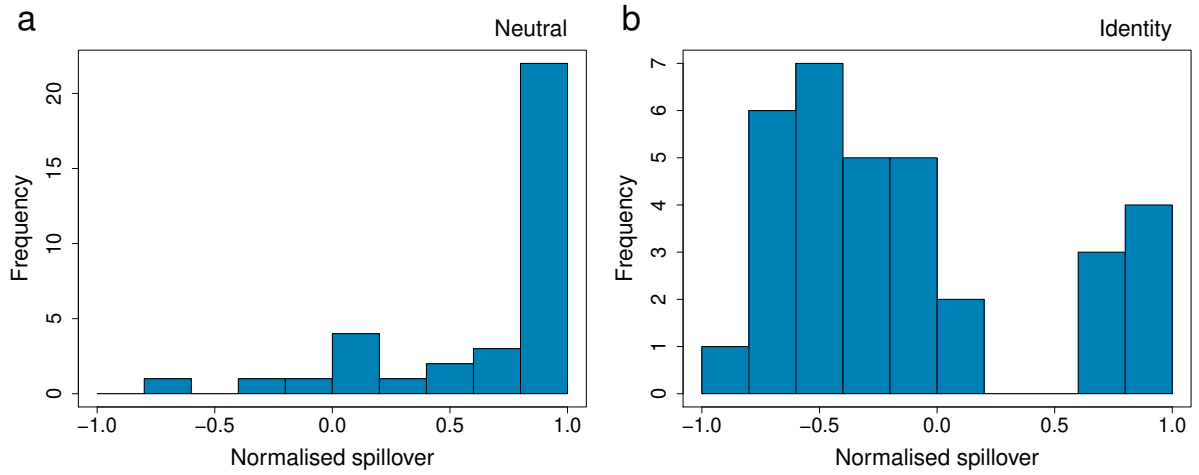
	Choose alternative behaviour	
	Model 1	Model 2
Intercept	0.13*** (0.02)	0.14*** (0.02)
Election		-0.02 (0.03)
(Neutral,T,Pre-int)	-0.03 (0.02)	-0.07 (0.04)
(Neutral,NT,Post-int)	0.63*** (0.05)	0.73*** (0.06)
(Neutral,T,Post-int)	0.81*** (0.03)	0.80*** (0.04)
(Identity,NT,Pre-int)	-0.12*** (0.02)	-0.12*** (0.02)
(Identity,T,Pre-int)	-0.12*** (0.02)	-0.13*** (0.02)
(Identity,NT,Post-int)	0.09 (0.06)	0.03 (0.07)
(Identity,T,Post-int)	0.53*** (0.05)	0.54*** (0.06)
Election×(Neutral,T,Pre-int)		0.06 (0.05)
Election×(Neutral,NT,Post-int)		-0.15 (0.09)
Election×(Neutral,T,Post-int)		0.01 (0.05)
Election×(Identity,NT,Pre-int)		0.001 (0.03)
Election×(Identity,T,Pre-int)		0.02 (0.03)
Election×(Identity,NT,Post-int)		0.10 (0.12)
Election×(Identity,T,Post-int)		-0.02 (0.09)

The  $p$  values are based on two-sided (Gaussian) z-tests.

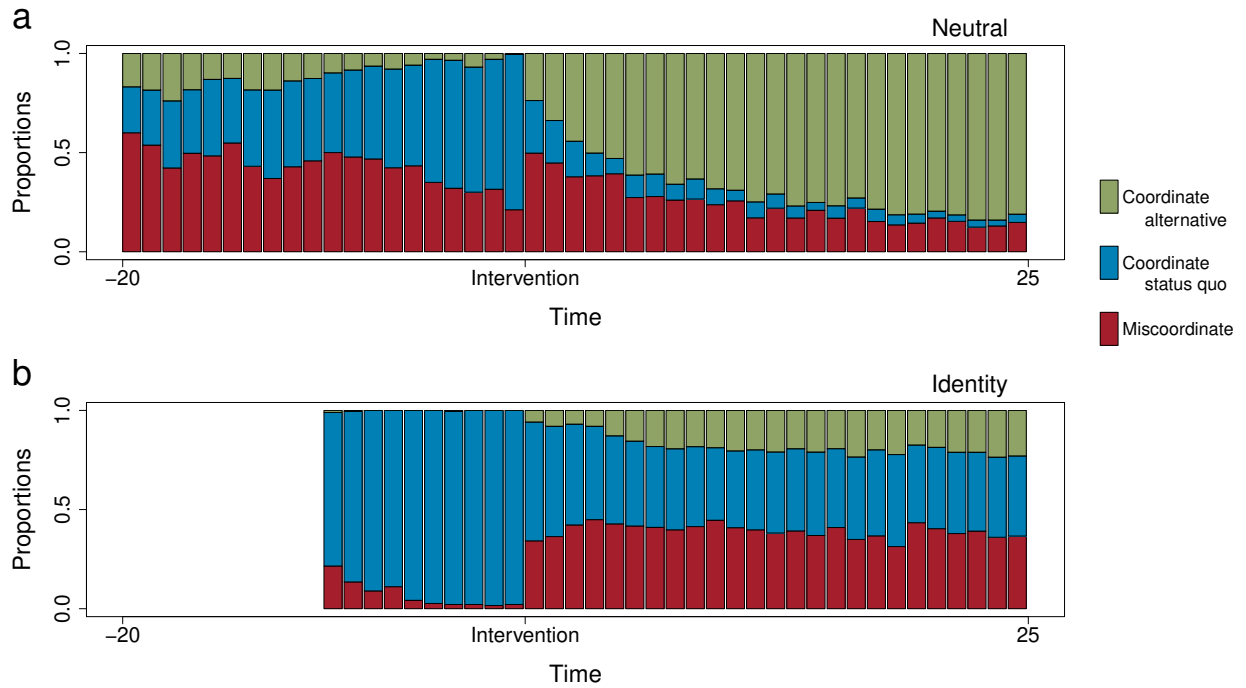
\*  $p \in (0.01, 0.05]$  \*\*  $p \in (0.001, 0.01]$  \*\*\*  $p \leq 0.001$



**Figure 1 | The two images used to label buttons in the identity treatment.** Specifically, instead of clicking on a button labelled with @ or #, as in the neutral treatment, participants in the identity treatment had to choose by clicking on one of two buttons with these images embedded in the buttons themselves (Supplementary Figs. 2 and 3).

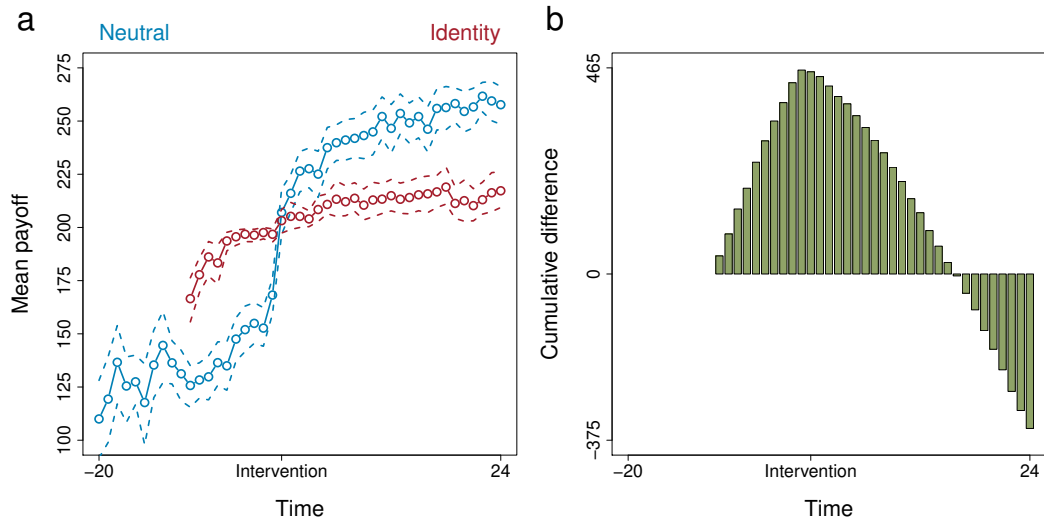


**Figure 2 | Distributions of normalised spillovers by treatment.** The spillover<sup>20</sup> is a normalised measure of endogenous behaviour change at the group level (Methods), and it can take any value in  $[-1, 1]$ . Negative values occur when the final proportion of the group choosing the alternative behaviour is less than the proportional size of the intervention. Positive values occur when the final proportion choosing the alternative behaviour is greater than the proportional size of the intervention. **a**, The distribution of spillovers in the neutral treatment. **b**, The distribution of spillovers in the identity treatment. The difference in spillovers by treatment is large and highly significant (Table 2).



**Figure 3 | Choice dynamics by treatment.** The status quo behaviour was the choice associated with the norm that emerged in the pre-intervention phase of a session. With a status quo established, the alternative behaviour was simply the other choice option, which was always favoured by the intervention (Table 1). Here we show, by period, the proportions of participants coordinating on the status quo, coordinating on the alternative, or miscoordinating. **(a)** In neutral sessions, groups were relatively slow to converge before intervention and relatively fast to converge on the alternative norm after intervention. **(b)** In identity sessions, groups converged quickly before intervention and persisted in a state of chronic disagreement after intervention.





**Figure 4 | Payoff dynamics.** **a**, Mean payoffs by treatment and period. Dashed lines are 95% confidence intervals from a bootstrapping algorithm clustered at the group level. Compared to the neutral treatment, political labels in the identity treatment provided a ready focal point<sup>56</sup> that allowed groups to converge on a norm quickly before intervention. After intervention, however, chronic disagreement (Fig. 3) prevented participants in the identity treatment from taking advantage of the new opportunities provided by the intervention. **b**, The accumulated difference in mean payoffs, identity minus neutral, over periods the two treatments had in common.

# Supplementary Information for “Group identities make fragile tipping points”

Sönke Ehret<sup>1,†,\*</sup>, Sara M. Constantino<sup>2,3,†,\*</sup>, Elke U. Weber<sup>2,3,4</sup>, Charles Efferson<sup>1,‡,\*</sup>, & Sonja Vogt<sup>1,5,6‡,\*</sup>

<sup>1</sup>*Faculty of Business and Economics, University of Lausanne, Switzerland*

<sup>2</sup>*School of Public and International Affairs, Princeton University, USA*

<sup>3</sup>*Andlinger Center for Energy and the Environment, Princeton University, USA*

<sup>4</sup>*Department of Psychology, Princeton University, USA*

<sup>5</sup>*Centre for Development and Environment, University of Bern, Switzerland*

<sup>6</sup>*Nuffield College, University of Oxford, United Kingdom*

†*Shared first authorship.*

‡*These authors jointly supervised the research.*

\**Address correspondence to sonke.ehret@gmail.com, sara.constantino@gmail.com, sonja.vogt@soz.unibe.ch, and charles.efferson@unil.ch.*

## Contents

<b>1</b>	<b>Definitions of Key Terms</b>	<b>3</b>
<b>2</b>	<b>Additional Analytical Methods and Results</b>	<b>4</b>
2.1	Political Labels & Coordination: Analysis of Individual Behaviour . . . . .	4
2.2	Political Labels, Political Preferences & Coordination: Analysis of Individual Behaviour . . . . .	5
2.3	Political Labels & Spillovers: Analysis of Group Behaviour . . . . .	9
2.4	Political Labels, Coordination & a Federal Election: Analysis of Individual Behaviour	10

<b>3</b>	<b>Experimental Procedure and Design Details</b>	<b>12</b>
3.1	Experimental Design . . . . .	12
3.2	Experimental Treatment Condition . . . . .	15
3.3	Pre- and Post-Intervention Phase . . . . .	16
3.4	Pre-Testing and Selection of Identity Labels . . . . .	17
<b>4</b>	<b>Coordination Game Payoff and Parameter Selections</b>	<b>22</b>
4.1	Introducing the Targeted Intervention . . . . .	23
4.2	Setting the Intervention Amounts . . . . .	24
<b>5</b>	<b>Data Collection</b>	<b>26</b>
5.1	Pilot Sessions . . . . .	27
5.2	Participant Panel and Sample Selection . . . . .	27
5.3	Study Scheduling . . . . .	29
5.4	Experimental Sample: Cleaning and Overview of the Group Data . . . . .	31
5.5	Communication with Subjects During the Experiment . . . . .	32
5.6	Randomization and Sample Composition by Treatment Group . . . . .	33

5.7	Participant Payments . . . . .	34
5.8	Ethics and Preregistration . . . . .	34
<b>6</b>	<b>Robustness Checks</b>	<b>34</b>
6.1	Dropouts . . . . .	34
6.2	Participant Authentication . . . . .	35
6.3	Additional Robustness Analyses . . . . .	36
<b>7</b>	<b>Instructions and Questionnaires</b>	<b>41</b>
7.1	Experimental Instructions for the Coordination Game . . . . .	41
7.2	Questionnaire Items: Recruitment and Image Pre-test . . . . .	52
<b>8</b>	<b>Supplementary References</b>	<b>82</b>
<b>1</b>	<b>Definitions of Key Terms</b>	

- **Alternative Behaviour:** The minority choice in the final period of the pre-intervention phase. Alternately, the policy-maker’s objective.
- **Identity Treatment:** Between-subjects treatment condition with political labels designed to cue partisan group identities.
- **Neutral Label:** The symbols “@” or “#”, which were embedded into the choice options in the neutral treatment.

- **Neutral Treatment:** Between-subjects treatment condition with neutral labels.
- **Partisan Feelings:** Political preferences measured using “thermometer” scores towards the two main candidates of the 2020 election, Donald Trump and Joe Biden.
- **Period:** A decision-period in the coordination game.
- **Political Labels:** Image of Donald Trump or Joe Biden, which were embedded into the choice options in the identity treatment.
- **Post-Intervention Phase:** The period in the coordination game where the intervention payoff has been introduced.
- **Pre-Intervention Phase:** The period in the coordination game before the intervention payoff has been introduced.
- **Norm:** The majority option that the group has converged on in either of the pre- or post-intervention phases.
- **Spillover:** Proportion of non-targeted participants that shift from the status quo to the alternative behaviour, accounting for the number of individuals who were targeted.
- **Status Quo Behaviour:** The majority choice in the final period of the pre-intervention phase.
- **Targeted Individuals:** Subset of individuals whose payoffs changed following the intervention.

## 2 Additional Analytical Methods and Results

**2.1 Political Labels & Coordination: Analysis of Individual Behaviour** Our primary model of interest in the main paper is an analysis of the probability of an individual switching from the status quo behaviour to the alternative behaviour in the post-intervention phase (Main Study, Table

2). We analysed the effect of the identity treatment (“political labels”) on the probability of an individual switching from one choice option to another by looking at individual choices in the final periods of phases 1 and 2 of the coordination game. We specified a difference-in-difference linear probability model that is fully saturated with respect to (i) the experimental design and (ii) whether player  $i$  was targeted by the intervention.

$$\begin{aligned}
c_i &= \beta_0 + \beta_1[u_i = 1] + \beta_2 z_i + \beta_3[u_i = 1]z_i \\
&\quad \beta_4 \tau_i + \beta_5[u_i = 1]\tau_i + \beta_6 z_i \tau_i + \\
&\quad \beta_7[u_i = 1]z_i \tau_i + \epsilon_i
\end{aligned} \tag{1}$$

Where  $i \in \{1, 2, \dots, I\}$  is a unique index for each participant and  $\tau_i \in \{0, 1\}$  is a time index.  $\tau_i = 0$  indicates the final period of the pre-intervention phase and  $\tau_i = 1$  indicates the final period of the post-intervention phase. Our dependent variable is  $i$ 's choice in these two periods, given by  $c_i \in \{0, 1\}$  where 0 represents the status quo behaviour at the end of the pre-intervention phase and 1 represents the alternative behaviour, i.e. the choice targeted by the intervention. We specifically notate the treatments using Iverson brackets  $[\cdot]$  and treatment dummies.  $u_i \in \{0, 1\}$  indicates whether  $i$  was in the identity treatment (dummy variable  $[u_i = 1]$ ), or the neutral treatment (dummy variable  $[u_i = 0]$ ).  $z_i \in \{0, 1\}$  indicates whether  $i$  was targeted by the intervention ( $z_i = 1$ ) or not ( $z_i = 0$ ). The omitted category is the neutral treatment,  $u_i = 0$ , in the pre-intervention phase of the experiment,  $\tau_i = 0$ , for non-targeted participants,  $z_i = 0$ . We estimated the model with robust standard errors clustered at the group level. Note that we re-coded the variables here and in the main study to avoid 3-way and 4-way interactions. These analyses were pre-registered on the Open Science Framework, and are retrievable at [osf.io/6adbx](https://osf.io/6adbx).

## 2.2 Political Labels, Political Preferences & Coordination: Analysis of Individual Behaviour

We extended the linear probability model above to include a measure of intensity of political pref-

erences as a potential moderator of a participant’s willingness to shift to the alternative behaviour in the post-intervention phase. We ran this regression on the subgroup of participants in the identity treatment ( $u_i = 1$ ).

$$c_i = \beta_0 + \beta_1 z_i + \beta_2 \tau_i + \beta_3 z_i \tau_i + \beta_4 p_i + \beta_5 z_i p_i + \beta_6 p_i \tau_i + \beta_7 z_i \tau_i p_i + \epsilon_i \quad (2)$$

Where  $p_i \in [0, 99]$  is a continuous variable indicating intensity of political feelings for each subject in the experiment. We calculated  $p_i$  by taking the differences between an individual’s score on the out-group and in-group partisan feeling thermometers  $FT_{in} - FT_{out}$  (e.g. for Republicans, Thermometer score (Trump) - Thermometer score (Biden)). The feeling thermometer can range from 0 to 100 - see section 7.2 (*therm\_biden* and *therm\_trump*) for more information on these measurements. We restricted our sample to those with a strong preference for their own party and a dislike for the other party. We did this by excluding participants who did not give the out-group candidate a rating of  $< 50$  and the in-group candidate a rating of  $\geq 50$ . Our differences were thus always  $> 0$  and could range from 1 to 100. For ease of interpretation, we re-scaled this difference to range from 0 to 99, and we inverted the scale so that 0 represents the highest intensity of partisan feelings. The empirical values in our sample range from a minimum of 0 to a maximum of 97, with a mean of 28.75.

We also ran these analyses with participants in the neutral treatment. We did not anticipate strong effects of partisan feelings since participants were not primed about their political identities and made choices between options with neutral labels. Participants did not know the political affiliation or political feelings of their group members in either condition. Results of these analyses are shown in Supplementary Table 1.

In the identity treatment, we found that non-targeted individuals with weaker partisan feelings

were more likely to switch to the alternative behaviour after the intervention phase relative to non-targeted participants before the intervention (Supplementary Table 1, Partisan Feeling $\times$ (NT,Post-int)). In the neutral condition, we found that the strength of partisan feelings had no effect on participants' choices, which was expected since they were choosing between neutral labels.



**Supplementary Table 1 | Political Feelings and the Emergence of New Norms.** This table shows the fitted linear probability models for individual choices in the final periods of the pre- and post-intervention phases, separately for the identity and neutral treatments. We included cluster-robust standard errors, clustered at the group level. Composite dummies are defined jointly over (i) whether the participant was targeted (T) or not (NT) by the intervention, and (ii) whether the choices were from the pre- or post-intervention phase. Differences in partisan feelings range from 0 to 99, where 0 is the highest intensity and 99 the lowest. Regressions were pre-registered on OSF.

	Choose alternative behaviour	
	Identity	Neutral
Intercept	-0.01 (0.01)	0.16*** (0.04)
Partisan Feelings	0.001 (0.001)	-0.001 (0.001)
(T,Pre-int)	-0.002 (0.02)	-0.10 (0.06)
(NT, Post-int)	0.08 (0.05)	0.59*** (0.06)
(T, Post-int)	0.62*** (0.08)	0.80*** (0.04)
Partisan Feelings × (T,Pre-int)	0.0001 (0.001)	0.002 (0.002)
Partisan Feelings × (NT,Post-int)	0.004* (0.002)	0.001 (0.002)
Partisan Feelings × (T,Post-int)	0.001 (0.002)	0.0003 (0.001)

\*  $p \in (0.01, 0.05]$     \*\*  $p \in (0.001, 0.01]$     \*\*\*  $p \leq 0.001$

**2.3 Political Labels & Spillovers: Analysis of Group Behaviour** We analysed "long-run" behavioural spillovers following the intervention in both treatments by focusing our analyses on the final period of the post-intervention phase. Spillovers capture the proportion of participants shifting to the choice favoured or promoted by the intervention at the end of the post-intervention phase, accounting for the number of individuals whose incentives were changed by the intervention.

We calculated spillovers using the actual number of targeted group members according to the following equation:

$$\Theta_j = \frac{[\hat{q}_j > \phi_j](\hat{q}_j - \phi_j)}{1 - \phi_j} + \frac{[\phi_j \geq \hat{q}_j](\hat{q}_j - \phi_j)}{\phi_j} \quad (3)$$

where  $[\cdot]$  indicate Iverson brackets. Spillovers are given by  $\Theta_j \in [-1, 1]$  and  $j \in \{1, 2, \dots, I\}$  is a unique index for each group.  $\hat{q}_j$  is the proportion of group  $j$  that chose the alternative behaviour ( $c_i = 1$ ) and  $\phi_j$  is the proportion of group  $j$  that was targeted by the intervention. Positive values indicate that the proportion choosing the alternative behaviour in the post-intervention phase is larger than the number of individuals targeted. Negative values indicate that the proportion choosing the alternative behaviour is smaller than the proportion of individuals targeted.

We examined treatment differences on spillovers with the following regression specification.

$$\Theta_j = \beta_0 + \beta_1[u_j = 1] + \epsilon_j \quad (4)$$

Where treatment is notated again using dummies, by  $u_j \in \{0, 1\}$ , such that  $[u_j = 1]$  indicates that  $j$  is in the identity treatment and  $[u_j = 0]$  in the neutral treatment.

## 2.4 Political Labels, Coordination & a Federal Election: Analysis of Individual Behaviour

In exploratory analyses, we treated the election as a natural experiment and looked at its effects on the emergence of new norms following the intervention in both the neutral treatment and the identity treatment. The result of the 2020 U.S. Federal election was called on November 7th, four days after the election. We treated any observation up to and including November 7th as pre-election results and those after as post-election results. We extended the individual-choice model described above to include an additional pre/post-election dummy,  $e_i \in \{0, 1\}$ , where  $e_i = 0$  indicates observations that took place through November 7th and  $e_i = 1$  indicates observations that took place after.

For these analyses, we were interested in understanding differences by political affiliation. In particular, we were interested in the following contrasts: pre- and post-intervention phase, targeted vs. non-targeted, neutral vs political labels, and pre- and post-election. We ran this analysis on the overall sample, as well as Republican and Democratic subgroups. We ran additional analyses in which we expanded our sample to include not only the final periods of the pre- and post-intervention phases but also the last two periods of each phase and the last four periods of each phase. These additional analyses are indicated by a 2 and a 4 in the columns of Supplementary Table 2. The primary results of interest in the main study still hold.

We found that the election had a circumscribed effect on Republican groups in the neutral treatment. In particular, we found that after the election, non-targeted Republicans in the neutral treatment were much less likely to switch to the alternative behaviour in the post-intervention phase relative to non-targeted Republicans in the pre-intervention phase (Supplementary Table 2, Rep + Rep2 + Rep4, Election $\times$ (Neutral,NT,Post-int)). This finding was unexpected and suggests an avenue for future research.

**Supplementary Table 2 | Emergence of New Norms After an Election.** Linear probability models of individual choices in pre- and post-intervention phases. Models 1-3 show results for all groups, starting with only the final periods of pre-/post- intervention phases (All), the final two periods of each (All2), and the final four periods of each (All4). Models 4-6 and 7-9 show the same results after sub-setting to Republican and Democratic groups, respectively. Cluster-robust standard errors are clustered at the group level. Election is a dummy that splits sessions into those that occurred before or on November 7th 2020 and those that occurred after. Composite dummies are defined jointly over (i) whether the session used neutral or political labels, (ii) whether the participant was targeted (T) or not (NT) by the intervention, and (iii) whether the choices were from the pre- or post-intervention phase. These regressions are exploratory and the extension of the number of periods can be considered a robustness check.

	Choose alternative behaviour								
	All	All2	All4	Rep	Rep2	Rep4	Dem	Dem2	Dem4
Intercept	0.14*** (0.02)	0.20*** (0.04)	0.20*** (0.04)	0.13*** (0.02)	0.22*** (0.04)	0.24*** (0.04)	0.15*** (0.04)	0.20*** (0.05)	0.17*** (0.04)
Election	-0.02 (0.03)	-0.04 (0.04)	-0.001 (0.04)	-0.08 (0.04)	-0.08 (0.05)	-0.06 (0.07)	0.001 (0.05)	-0.02 (0.06)	0.04 (0.05)
(Neutral,T,Pre-int)	-0.07 (0.04)	-0.10** (0.04)	-0.05 (0.02)	0.03 (0.06)	-0.03 (0.04)	-0.04 (0.02)	-0.15*** (0.04)	-0.15** (0.06)	-0.06 (0.04)
(Neutral,NT,Post-int)	0.73*** (0.06)	0.70*** (0.06)	0.69*** (0.06)	0.76*** (0.11)	0.71*** (0.08)	0.66*** (0.09)	0.71*** (0.07)	0.70*** (0.07)	0.72*** (0.07)
(Neutral,T,Post-int)	0.80*** (0.04)	0.74*** (0.05)	0.74*** (0.05)	0.80*** (0.05)	0.71*** (0.06)	0.69*** (0.04)	0.80*** (0.05)	0.76*** (0.06)	0.77*** (0.07)
(Political,NT,Pre-int)	-0.12*** (0.02)	-0.19*** (0.04)	-0.19*** (0.04)	-0.08 (0.06)	-0.16** (0.06)	-0.20*** (0.05)	-0.13** (0.04)	-0.19*** (0.05)	-0.17*** (0.04)
(Political,T,Pre-int)	-0.13*** (0.02)	-0.19*** (0.04)	-0.19*** (0.04)	-0.13*** (0.02)	-0.22*** (0.04)	-0.24*** (0.04)	-0.13** (0.04)	-0.18*** (0.05)	-0.15*** (0.04)
(Political,NT,Post-int)	0.03 (0.07)	-0.03 (0.08)	-0.02 (0.08)	-0.02 (0.04)	-0.05 (0.08)	-0.06 (0.07)	0.04 (0.09)	-0.01 (0.09)	0.004 (0.08)
(Political,T,Post-int)	0.54*** (0.06)	0.47*** (0.07)	0.46*** (0.07)	0.49** (0.18)	0.41*** (0.11)	0.40*** (0.12)	0.54*** (0.08)	0.49*** (0.08)	0.50*** (0.07)
Election×(Neutral,T,Pre-int)	0.06 (0.05)	0.06 (0.05)	0.01 (0.04)	0.001 (0.08)	0.01 (0.08)	0.03 (0.07)	0.11* (0.05)	0.10 (0.07)	0.01 (0.05)
Election×(Neutral,NT,Post-int)	-0.15 (0.09)	-0.13 (0.09)	-0.17 (0.09)	-0.37* (0.16)	-0.35** (0.13)	-0.37** (0.14)	-0.05 (0.09)	-0.04 (0.09)	-0.10 (0.09)
Election×(Neutral,T,Post-int)	0.01 (0.05)	0.03 (0.06)	0.0004 (0.06)	-0.02 (0.11)	-0.04 (0.11)	-0.03 (0.10)	0.02 (0.07)	0.05 (0.07)	-0.003 (0.07)
Election×(Political,NT,Pre-int)	0.001 (0.03)	0.02 (0.04)	-0.001 (0.04)	0.03 (0.07)	0.02 (0.07)	0.03 (0.08)	-0.01 (0.05)	0.02 (0.06)	-0.03 (0.05)
Election×(Political,T,Pre-int)	0.02 (0.03)	0.04 (0.05)	0.002 (0.05)	0.12* (0.06)	0.11 (0.06)	0.10 (0.07)	-0.01 (0.06)	0.01 (0.06)	-0.05 (0.05)
Election×(Political,NT,Post-int)	0.10 (0.12)	0.11 (0.13)	0.08 (0.13)	0.17 (0.20)	0.11 (0.22)	0.06 (0.20)	0.09 (0.13)	0.12 (0.13)	0.08 (0.12)
Election×(Political,T,Post-int)	-0.02 (0.10)	0.02 (0.10)	-0.03 (0.10)	-0.08 (0.19)	-0.05 (0.15)	-0.08 (0.16)	0.03 (0.11)	0.06 (0.11)	-0.01 (0.10)

\*  $p \in (0.01, 0.05]$  \*\*  $p \in (0.001, 0.01]$  \*\*\*  $p \leq 0.001$

### 3 Experimental Procedure and Design Details

#### 3.1 Experimental Design

The experimental procedure had three stages:

1. Participants were selected based on their self-reported partisan identity in a recruitment survey and assigned to politically homogeneous groups composed of either US-Republicans or US-Democrats.
2. Participants played a repeated coordination game, where choice options were labelled with either a neutral symbol or a political image. Groups were randomly assigned to one of two treatments, which remained fixed throughout the session.
3. After 10 to 20 periods of playing the coordination game, where groups converged on one of two choice options, establishing a status quo behaviour, a subset of participants were targeted by an intervention. The intervention changed the payoff structure of a subset of the group to incentivise the alternative behaviour.

In the following sections, we start by describing the common set-up, which holds across treatment conditions and phases of the game, unless otherwise stated. We follow this section with a detailed explanation of the treatment condition, and the intervention. Finally, we describe a survey we ran before the coordination game with a separate group of participants to test and select among possible political labels for the political treatment condition.

Assignment to treatment was between-subjects and conditions were clustered at the group level, with 12 participants per group. We constructed homogeneous groups based on self-reported political affiliation (Democrat or Republican). Individuals were not informed about the partisan composition of their group, nor was any direct reference made to U.S. politics during the recruitment or in the instructions.

Politically homogeneous groups began by playing a pure coordination game with uniform monetary incentives associated with two response options, as shown in Supplementary Table 3 with the choice labels of the neutral condition. Groups were randomly assigned in the beginning of a session and remained the same throughout the entire session and participants were aware of this. Participants remained anonymous throughout the game, and there was no communication between participants.

In each period, participants were randomly re-matched with a member of their group into a *participant - counterpart* pair in every period of the game. Once matched, participants simultaneously selected one of the two choice options by clicking on the choice label. Participants could not see their partner's choice when making their own choice, but could earn more in each period of the game if they were able to coordinate on the same choice as their counterpart. Participants tallied points in each period of the game, but were only paid for 5 randomly chosen periods at an exchange rate of 100 points = \$1. All participants also received the equivalent of \$5.5 in completion and show-up payments.

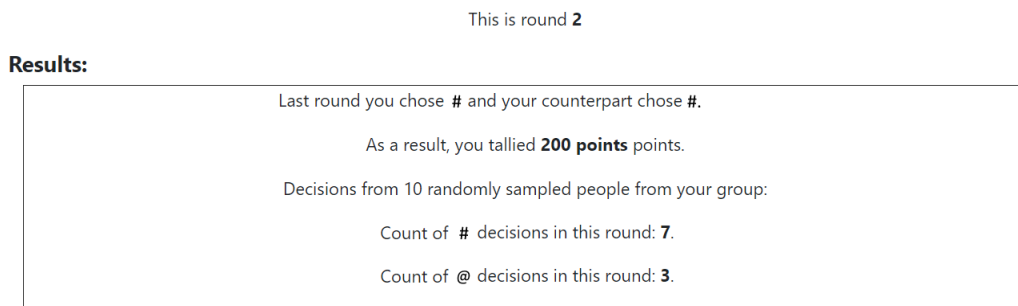
**Supplementary Table 3 | Simplified Symmetric Coordination Game, Neutral Condition.** In a symmetric coordination game, payoffs depend on the combination of a row player's choice and a column player's choice. The payoff  $c > 0$  is the positive payoff earned if both players make the same choice, coordinating on either # or @. As long as they coordinate, the two choices are equally rewarding.

Row	Column	
	#	@
#	$c$	0
@	0	$c$

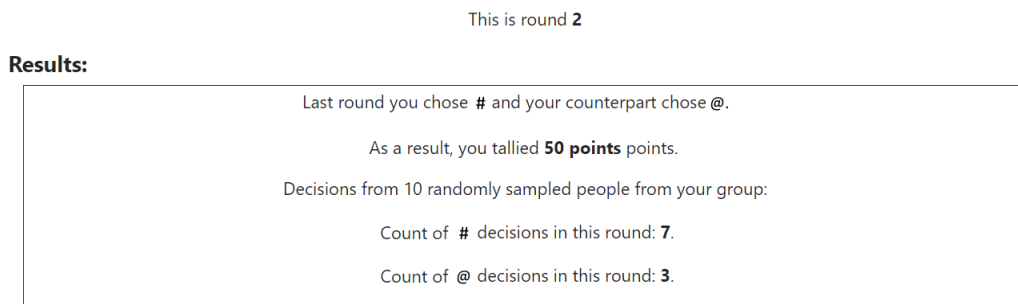
At the end of each period, participants were given feedback about what they earned for that period, the choice option of their randomly assigned partner, and how many players in their group chose each choice option. This count was based on 10 out of 12 randomly selected players in the

group. Supplementary Figure 1 shows an example of this feedback. The first period in the game had no feedback.

Each period of the coordination game had a preset timeout—180 seconds in the first period and 150 seconds in all subsequent periods. A timer and an alert screen were shown for the last 60 seconds of a period.



(a) Feedback for Matched Choices.



(b) Feedback for Mismatched Choices.

**Supplementary Figure 1 | Feedback Following Each Period.** The feedback participants saw in the beginning of each period starting with the second period. The top panel shows feedback for a period on which a participant coordinated with their partner. The bottom panel shows the feedback for a period in which the partners mis-coordinated.

The experiment was implemented using the otree software<sup>1</sup>. Participants gave informed consent and then read the instructions. They could download the instructions to have them available during the study. After reading the instructions, all participants were required to pass a quiz cover-

ing basic questions about the coordination game, group composition, study elements, and payments before they could join the study. If they answered incorrectly, they had to retake the quiz. Like this, we ensured that participants reasoned through the important elements of the instructions.

After reading the instructions and passing the quiz, participants sometimes had to wait up to 10 minutes in a virtual waiting lobby for other participants to arrive. Once at least 12 participants were present, the session started. If the waiting window expired without 12 participants arriving in time, participants could leave and receive a show-up payment or stay on for an additional five minutes, after which they would receive the same show-up payment if the game failed to start.

**3.2 Experimental Treatment Condition** We experimentally manipulated the labels associated with the choice options in the coordination game by assigning groups to either the neutral or the identity treatment condition at the beginning of the game. Labels refer to the images embedded directly in the buttons a participant had to press to indicate their choice in the game. Labels were the same for all members of a group, and did not change between the pre-intervention and post-intervention phase.

In the neutral treatment, the choice options were randomly assigned either an “@” or a “#” symbol. In the identity treatment, the choice options were randomly assigned to political labels. The political labels were either a pro-Republican image (R) or a pro-Democrat image (D). All participants were introduced to either the neutral or political labels before starting the game. Supplementary Figure 2 shows the coordination game with the embedded neutral or political labels as participants encountered them during the game. Further information on the selection of political labels can be found in section 3.4.

We presented the choices in the neutral treatment and in the identity treatment as a matrix. We randomised the order of the labels (top or bottom, left or right) between participants. The order



remained fixed for a participant throughout the course of the experiment.

**3.3 Pre- and Post-Intervention Phase** Regardless of whether groups were assigned to the neutral or identity treatment, the study was divided into two parts. The experiment started with the pre-intervention phase. The instructions for the pre-intervention phase indicated that an intervention would occur after 10 to 20 periods. Once the intervention phase was triggered (see below), participants returned to an instruction screen where they read about the intervention in detail, followed by the post-intervention phase.

The intervention did not vary between the two treatments. The pre-intervention phase was a symmetric coordination game with uniform payoffs, as described above and as shown in Supplementary Figure 2. Participants played the pre-intervention phase for a minimum of 10 periods and a maximum of 20 periods. After 10 periods, the pre-intervention phase ended following convergence on one of the choices. We defined convergence as at least 90% of group members selecting the same choice. If the players did not converge on a choice within the first 20 periods, the pre-intervention phase ended in period 20 and the intervention was introduced.

The post-intervention phase lasted for 25 periods. All participants received the following information in the beginning of Part 2: “[...] for some people in the group (including yourself), the payoffs associated with the different choices may have changed. Thus, some members of the group have different payoffs in Part 2, while others keep the same payoffs as in Part 1.”

The intervention adjusted the payoffs of the targeted participants so that the alternative option now earned more money than the status quo option, regardless of the counterpart’s choice. These modified payoffs lasted the entire duration of the post-intervention phase. Supplementary Figure 3 shows the adjusted payoff matrix for the targeted participants in the post-intervention phase. In this example, targeted individuals were incentivised to choose the option associated with the @

choice label in the neutral condition and the Trump choice label in the identity condition. In the example shown here, these happen to be the choice options in the top row. Which behavior was incentivised in the post-intervention phase was based on the status quo behaviour at the end of the pre-intervention phase. This happened independently of whether the status quo behaviour was visually displayed in the top or bottom row for individual participants. In general, payoffs for targeted participants always incentivised the opposite of the status quo behaviour, subject to the conditions mentioned above. In cases where the group did not meet the 90% criterion in the pre-intervention phase, the intervention targeted the choice option selected by less than 50 % of the group in period 20 as the alternative behaviour. Looking at the data, we find that the final period of the pre-intervention phase never resulted in a tie.



The intervention targeted a random 50% of each group. Participants were assigned to the targeted subset at the start of the session. With dropouts, this means that the effective size of the targeted group can range from 4, to 5, to 6 players.

**3.4 Pre-Testing and Selection of Identity Labels** We piloted several political labels, but to ensure a robust behavioural response, we opted for labels depicting the two political candidates of the 2020 U.S. presidential election, namely Joe Biden and Donald Trump. A growing body of evidence suggests that negative political expression is especially strong for the figureheads of the Democratic and Republican parties.<sup>2,3</sup>

We hired a U.S.-based artist (Max Alnutt) to design several images of Joe Biden and Donald Trump that would be likely to trigger strong positive and negative reactions based on political identity. Max produced neutral and positive images of Trump and Biden, as well as a positive image of Biden with Kamala Harris and of Trump with Mike Pence (Supplementary Figure 4).

		Your counterpart's choice	
		@	#
Your Choice	@	200 points	50 points
	#	50 points	200 points

(a) Neutral Label





		Your counterpart's choice	
			
Your Choice		200 points	50 points
		50 points	200 points

(b) Political Label

**Supplementary Figure 2 | Pre-Intervention Phase Payoff Matrix.** The pre-intervention payoffs were symmetrical, encouraging coordination but not favouring either choice option. The payoff matrices were identical for all participants in both treatments, save for the labels associated with the choice options.

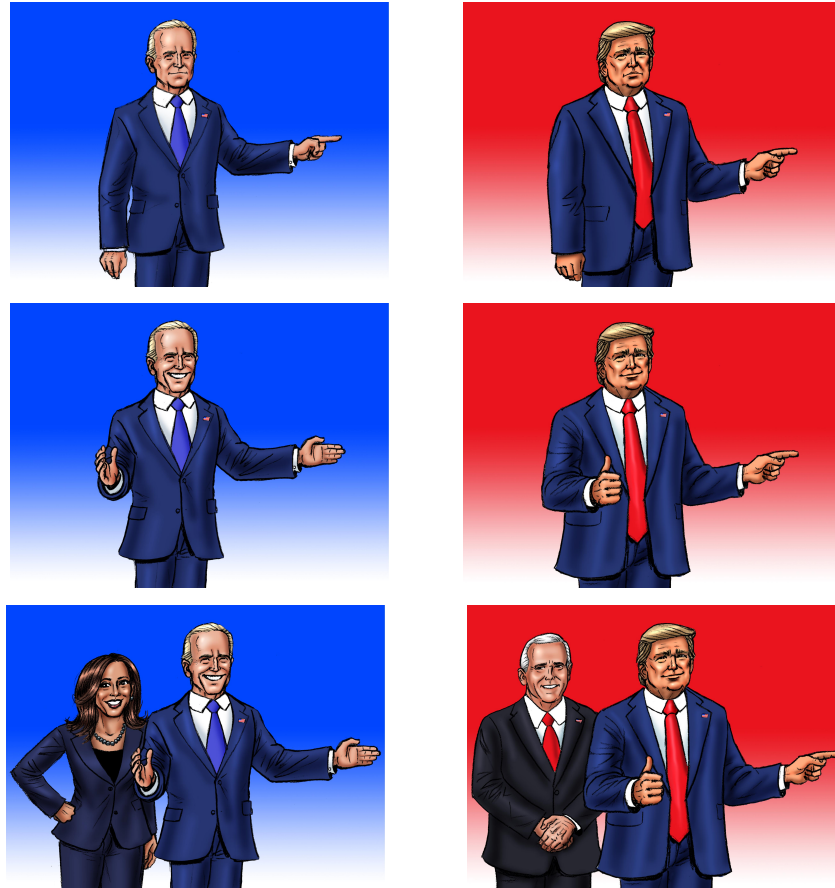
		Your counterpart's choice	
		@	#
Your Choice	@	350 points	350 points
	#	50 points	200 points

(a) Neutral Label

		Your counterpart's choice	
			
Your Choice		350 points	350 points
		50 points	200 points

(b) Political Label

**Supplementary Figure 3 | Post-Intervention Phase Payoff Matrix.** Post-intervention payoffs for targeted individuals, where the incentivised alternative behaviour was either @ or *Trump*. Post-intervention payoffs were again identical in both treatments but differed for the targeted participants and the non-targeted participants. Non-targeted participants kept the same payoff matrix from the pre-intervention phase.

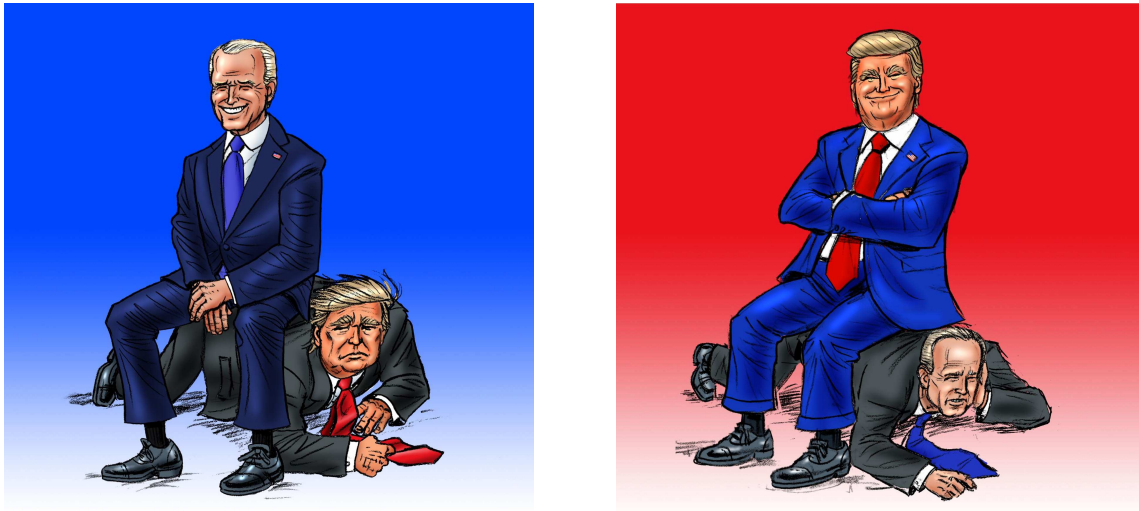


**Supplementary Figure 4 | Non-Selected Images for the Political Label Condition.** Images of Trump and Biden alone and with Harris and Pence. First row shows neutral images, second row shows positive images with expression, third row shows positive images with both the presidential and vice-presidential candidates.

To increase aversive reactions based on identity and out-group dislike, Max produced additional images that show Biden smiling in a victory pose over Trump and Trump smiling in a victory pose over Biden (Supplementary Figure 5). We expected that these images would trigger even stronger group identity reactions by increasing the dislike for the out-group. Background colours in all images are red or blue, matching the Republican and Democratic party colours.

We pre-tested people's reactions to these images with a small survey in October (13/10/2020 to 20/10/2020, N=198). We ran the same survey after the election as well to see whether preferences

had shifted with the change in political power in the U.S. (25/01/2021 to 27/01/2021). In the present study, we report only the data from the pre-election survey.



**Supplementary Figure 5 | Selected Images for the Political Label Condition.** Images of a smiling Trump or a smiling Biden in a “victory” position.

In this survey, we first asked respondents to report their political affiliation. We then assessed whether respondents were able to identify images of Biden and Trump (84% and 98% of respondents accurately identified Biden and Trump respectively). Next, we measured their willingness to pay for different Trump and Biden images with a short vignette study. The goal was to estimate the approximate monetary incentive needed to induce a partisan participant to choose the political label of the opposite party. This approximate estimate was then used to scale the intervention incentives in the coordination game.

The vignette described a situation in which one faces an urgent need to buy a hoodie at a gas station to stay warm (see pre-test questionnaire in section 7.2, item *image-hoodie*). We elicited each participants’ willingness to pay for a neutral hoodie. We also created eight hoodies, each with a different image of Trump or Biden on it, as shown in Supplementary Figures 4 and

5. Supplementary Figure 6 shows an example of hoodies with the “victory” and “vice-president” images. Participants indicated an amount between \$0 and \$50 that they would be willing to pay for each hoodie. We made sure to highlight in the vignette that this was a private decision—they would not be seen wearing the hoodie as they drove home alone in the car—to ensure that concerns about reputation did not affect the willingness to pay. We measured each respondent’s willingness to pay for each image.

We calculated the difference between the willingness to pay for each hoodie with an image and the hoodie without an image. We report the mean differences for Republicans and Democrats respectively in Supplementary Table 4.



**Supplementary Figure 6 | Images in the Willingness to Pay Vignette.** Example of a hoodie showing Biden triumphing over Trump on the left and one depicting Trump and Pence on the right.

The mean willingness to pay for the plain hoodie was \$30.3. Democrats reported an average willingness to pay difference of -\$1.6 for the Biden images and -\$14.7 for the Trump images relative to the plain hoodie. In both cases, Democrats were less willing to pay for a political hoodie than for a plain hoodie. Yet, in monetary values, Democrats reported an average net benefit of \$13.1 for Biden over Trump. Republicans reported an average willingness to pay difference of -\$10.5 for the Biden images and \$0.9 for the Trump images—a net benefit of \$11.4 for Trump relative to Biden. Thus the average net benefit, across the two groups, of choosing one’s favoured label and avoiding the disfavoured label was \$12.3. This is the magnitude required to make the average partisan buyer

**Supplementary Table 4 | Mean Willingness to Pay Differences.** Summary of the mean willingness to pay difference (in US-Dollars) for the hoodies with different political images relative to the plain hoodie without a political image. Negative values indicate that respondents were willing to pay more for the plain hoodie than for the hoodie with a political image. Grouped by participants who self-identify as Democrats and Republicans.

	Biden Images				Trump Images			
	Neutral	Positive	Victory	Vice-Presidents	Neutral	Positive	Victory	Vice-Presidents
Democrat	-2.08	-1.11	-1.52	-1.53	-14.48	-14.39	-15.14	-12.73
Republican	-10.41	-9.69	-11.61	-10.44	1.02	1.84	0.63	0.06

indifferent between Trump and Biden hoodies in our sample, all else equal. The Victory images yielded the largest difference in willingness to pay for Trump vs. Biden hoodies of all the images we considered. This was true for both Democrats (\$13.6) and Republicans (\$12.2).

#### 4 Coordination Game Payoff and Parameter Selections

In the following section, we briefly present a formal description of a coordination game with an exogenous intervention, and the role of political labels. We designed the experiment such that spillovers were strongly incentivised. Additionally, we provided strong incentives for the targeted participants to choose the alternative behaviour in the post-intervention phase. This in turn incentivises spillovers from the status quo behavior to the alternative behavior in the post-intervention phase.

Starting with the general setup, let  $N$  be the group size,  $N_T$  the number of group members targeted by the intervention, and  $\gamma \geq 0$  a constant payoff. The parameter  $c$  is the amount the row player receives if she successfully coordinates with her partner. The intervention is represented by a monetary incentive  $\delta > 0$ . For exposition, we arbitrarily take choice @ to be the status quo behaviour and # the policy maker's preferred behaviour. Post-intervention, participants face the following payoffs, as shown in Supplementary table 5:

**Supplementary Table 5 | Coordination Game with Payoffs for Targeted Individuals in the Post-Intervention Phase.** Payoffs for the row player conditional on the choices of the column player. For exposition, assume @ is the status quo behavior, and # the alternative behavior targeted by the policy maker.  $\gamma \geq 0$  is a constant payoff value for all outcomes,  $c > 0$  is the positive payoff earned if both players made the same choice in the pre-intervention phase, and  $\delta$  is an additional payoff from choosing the targeted behavior.

Row	Column	
	@	#
@	$\gamma + c$	$\gamma$
#	$\gamma + c + \delta$	$\gamma + c + \delta$

It can be seen from Supplementary Table 5, that as long as  $c \geq 0$  and  $\delta > 0$ , the targeted player would strictly prefer the alternative behaviour #. Note that the intervention has two components though, the parameter  $\delta$  and an additional payoff  $c$  for one of the non-coordination outcomes. This additional payoff ensures that targeted players earn more by choosing the alternative behaviour regardless of what the counterpart plays, removing any incentive to coordinate with the counterpart in this game.

**4.1 Introducing the Targeted Intervention** We designed our intervention to be maximally favourable for tipping and rapid transitions from one norm to another. This entails two goals for selecting our parameters  $\delta$ —the intervention amount—and  $\phi$ —the fraction of the group that is targeted.

1. Targeted players are sufficiently incentivised to change their behaviours.
2. A sizable fraction of the group is targeted to create conditions strongly favourable for tipping.

It is straightforward to derive the conditions under which non-targeted participants are expected to change behaviour given a number  $N_T$  of targeted participants. Expected payoffs for a



non-targeted player playing @ or # take the form,

$$\begin{aligned} E[\Pi(@)] &= \frac{N - N_T - 1}{N - 1}c \\ E[\Pi(\#)] &= \frac{N_T}{N - 1}c \end{aligned} \tag{5}$$

For the situation of interest to lead to *no* spillovers, we require that  $E[\Pi(\#)] < E[\Pi(@)]$ . Putting these conditions together and rearranging we get

$$\frac{1}{2} > \frac{N_T}{N - 1}. \tag{6}$$

The decision to target 50% of players,  $N_T = N/2$ , or  $\phi = 0.5$ , thus created a robust incentive for spillovers to emerge in the neutral treatment. Beyond considerations based on simple expected value calculations, prior research<sup>4</sup> suggests that interventions of the size we implemented are large enough to tip the group.

The experimental treatment relabelled choice options with images designed to activate partisan political identities. But apart from this difference, experimental manipulations for both political groups targeted the same fraction of the group  $\phi = 0.5$  and had the same intervention amounts.

**4.2 Setting the Intervention Amounts** In order to determine the unique intervention amount, we incorporated the additional consideration that the identity treatment contained favoured and disfavoured choice labels. A label was favoured if it corresponded with a participant's in-group and was disfavoured if it corresponded with the out-group. We assume that a participant who chooses the option associated with her favoured label experiences a non-monetary utility,  $\alpha$ . Analogously, we assume that a participant experiences a non-monetary disutility,  $-\beta$ , from choosing the option associated with her disfavoured label. The total utility—the net benefit for choosing one's favoured vs disfavoured label—is  $\alpha + \beta^a$ . Groups were politically homogeneous so all members were likely to have similar preferences in terms of partisan commitments (see section 5.2 for details).

---

<sup>a</sup>Note that we assume for the neutral treatment,  $\alpha = \beta = 0$ .

We use Biden (B) and Trump (T) as political labels. To illustrate, we arbitrarily assign the row player's favoured label to B and their disfavoured label to T. We also assume that the status quo behaviour in the player's group is B. By extension, we take T to be the policy maker's preferred behaviour. Post-intervention, targeted row players in this set-up would have the following payoffs as in Supplementary Table 6:

**Supplementary Table 6 | Coordination Game with Payoffs for the Targeted Individuals in the Post-Intervention Phase, Identity treatment.** Payoffs for the row player conditional on the choices of the column player in the post-intervention phase with political labels. For exposition, assume  $B$  is the status quo and favoured behavior, and  $T$  the alternative but disfavoured behavior targeted by the policy maker.  $\gamma \geq 0$  is a constant payoff value for all outcomes,  $c > 0$  is the positive payoff earned if both players made the same choice in the pre-intervention phase, and  $\delta$  is an additional payoff from choosing the targeted behavior.  $\alpha > 0$  and  $\beta > 0$  are non-monetary (dis-) utilities incurred for choosing the favoured and disfavoured choices, respectively.

Row	Column	
	B	T
B(favoured)	$\gamma + c + \alpha$	$\gamma + \alpha$
T(Disfavoured)	$\gamma + c + \delta - \beta$	$\gamma + c + \delta - \beta$

The intervention transforms the payoffs for targeted players such that they receive a *monetary* payoff of  $c + \delta + \gamma$  for playing T, regardless of what their partners chose. However, they also receive a dis-utility of  $\beta$  at the same time from having chosen their disfavoured label.

It was our goal to incentivise targeted players to choose the alternative behaviour regardless of what other players do, meaning that they would choose T even if they knew they were matched with someone choosing B. For a targeted player, the expected payoff from choosing the status quo behaviour B is, assuming all members of the group choose B,  $E[\Pi(A)] = \gamma + c + \alpha$ . The expected payoff from choosing T is  $E[\Pi(B)] = \gamma + c + \delta - \beta$ . Behaviour T is strictly preferred iff

$$\frac{\delta}{c} > \frac{\alpha + \beta}{c}. \quad (7)$$

To fulfil condition (7) it suffices that  $\delta > \alpha + \beta$  for targeted players. While  $\alpha + \beta$  is an unknown

quantity, we introduced a strong intervention aimed at changing the behaviour of the *targeted* participants. Their shift to the alternative choice would in turn create the spillover incentives for the non-targeted participants to change their behaviours as well, analogous to the neutral label condition. To achieve this goal, we used the data from the pre-test to approximate the monetary utility of choosing a label that conforms with one's own political identity group while avoiding the other group. This should be approximately equal to the value of  $\alpha + \beta$ . For each of the eight images, we calculated the average difference of the mean willingness to pay for each of the images  $WTP_{diff}$  and the willingness to pay for the plain hoodie  $WTP_{neutral}$ . We used the willingness to pay differences relative to the plain hoodie to scale the parameters  $\alpha + \beta$ , as illustrated in equation (8).

$$\frac{WTP_{diff}}{WTP_{neutral}} \propto \frac{\alpha + \beta}{c} \quad (8)$$

Given that, on average,  $WTP_{diff} = \alpha + \beta = \$12.6$  and  $WTP_{neutral} = \$30.3$ , the condition above implies  $(\alpha + \beta) \approx \frac{1}{3}c$ . Based on the pre-test results, we thus selected  $\delta$  conservatively, and set  $\delta = c$ . By doing this, we made sure that the inequality above was fulfilled. This choice ensured that individuals targeted ex-ante in the identity treatment would be likely to change their choices to the alternative behaviour, putting pressure on the non-targeted to change their status quo behaviors as well.

## 5 Data Collection

We used a virtual lab setup<sup>5-7</sup>. The setup consisted of two steps. We first built a participant panel which we used to specify the sampling frame for the main experiment. We then recruited individuals from this panel into pre-scheduled sessions to participate in an interactive virtual lab experiment.

**5.1 Pilot Sessions** We conducted pilot sessions of the main experiment to assess 1) the overall duration of the study, 2) to anticipate technical issues, and 3) to evaluate participant dropout rates. Specifically, we assessed recruitment and the on-boarding procedure, i.e. reading instructions, answering quiz questions and time spent in the online waiting lobby. We proceeded in two steps. The first set of pilot sessions was conducted from 23/09/2020 to 02/10/20 with a total of 8 groups, recruited from Amazon Mechanical Turk. These sessions included only the neutral condition and the pre-intervention phase of the study—thus, participants played a multi-period coordination game with neutral labels.

We ran a second set of pilot sessions on 27/10/20. These sessions included both the pre-intervention and post-intervention phases. We recruited participants from both Amazon Mechanical Turk and Prolific Academic. We assigned Amazon Mechanical Turk participants to the neutral treatment and the Prolific Academic participants to the identity treatment. We collected four groups, two on Amazon Mechanical Turk and two on Prolific Academic. For the identity treatment sessions, we used the “positive” single candidate images for homogeneous groups of Democratic participants.

**5.2 Participant Panel and Sample Selection** Participants were recruited via Prolific Academic, an online recruitment company<sup>8,9</sup>. We targeted respondents located in the United States and who use large screen devices (tablet device or laptop/desktop PC). Sampling for the participant panel involved filling six recruitment cells, based on interlocking and equally-sized demographic strata: self-identified partisanship (we included only self-reported Republicans and Democrats), age (two groups: 18 to 38 years, and 39 years and older) and sex (male and female). To qualify, participants had to be U.S. nationals residing in the U.S. at the time of the study. We only included participants with a Prolific Academic study approval rate of 95% or higher. This means that at least 95% of the studies completed prior to our study were deemed sufficiently complete by other researchers

on Prolific. We recruited participants between 13/10/20 and 1/12/20. In order to register for the study, participants had to fill out an extensive background questionnaire on their political attitudes and partisanship, and other demographic questions (see section 7.2 for details), resulting in a total initial database of 4,244 unique participants.

Using the background questionnaire, we screened our participants for duplicated IP addresses, IP addresses that indicate any location other than the U.S., self-identified records indicating a location other than the U.S., survey responses that took less than 3 minutes (the average duration of the survey was 10 minutes), and incomplete responses. In order to establish control of group composition in terms of opposing political self-identification and preferences, we further restricted the remaining database. We selected participants who were likely to react to our political labels with either sympathy or aversion due to their own political identities. Concretely, this means that subjects had to fulfil two conditions regarding their political preferences to be invited to participate in the study: 1. self-identify as either Republican or Democrat, and 2. demonstrate feelings of like (for the in-group) and dislike (for the out-group) with respect to their political identities.

On survey-based measures, political affect is often recorded via feeling thermometers<sup>3,10,11</sup>. These are survey scales where a respondent can indicate the degree to which he/she feels positively (warm) or negatively (cool) about a group, candidate or general category. The scale ranges from 0 to 100, and a score of 50 implies neutrality. To be eligible, self-identified Democrats had to have scores of 50 or higher for Biden and 49 or lower for Trump. These requirements were reversed for Republican eligibility. The recruitment survey questions used to filter participants can be found in section 7.2 (we used items *dem\_live*, *id\_party*, *therm\_biden* and *therm\_trump*).

This sample selection criteria means that our subjects did not have sympathies for both candidates, did not dislike the candidate of their own party, and preferred, even if weakly, the candidate of their own party over the candidate of the other party. These criteria and the initial screening com-

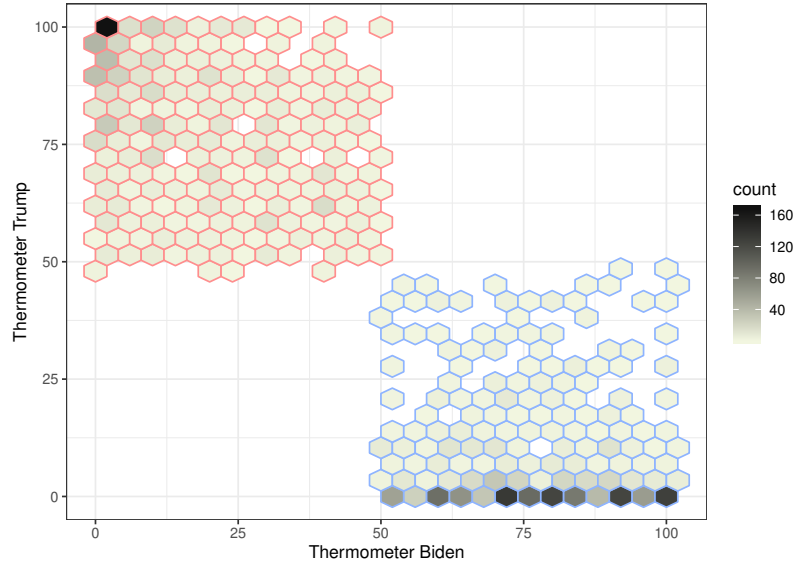
bined resulted in 1,253 Republicans and 1,535 Democrats. Supplementary Table 7 summarises the demographics of this final participant panel which served as the sampling frame for the experiment (N=2,788). Participants who completed the study, or who dropped out of the study at any point on or after the first period of group interaction, were excluded from future sessions.

**Supplementary Table 7 | Demographics, Participant Panel.** Key demographic statistics of the recruitment panel after subsetting on eligible participants: Democrat or Republican, feeling thermometer score  $\geq 50$  for the in-group, and  $< 50$  for the out-group presidential candidate. For question wording and answer scales, consult section 7.2.

variable	min	max	mean	sd	N
Age	18	89	39.63	14.05	2788
Education	1	4	2.86	0.77	2788
Income	1	8	4.12	1.85	2788
Sex	1	3	1.55	0.51	2788
Party (1:R, 2:D)	1	2	1.55	0.50	2788
Prolific approval rating	95	100	99.59	0.94	2788
Thermometer Biden	0	100	49.11	34.32	2788
Thermometer Trump	0	100	38.82	40.21	2788

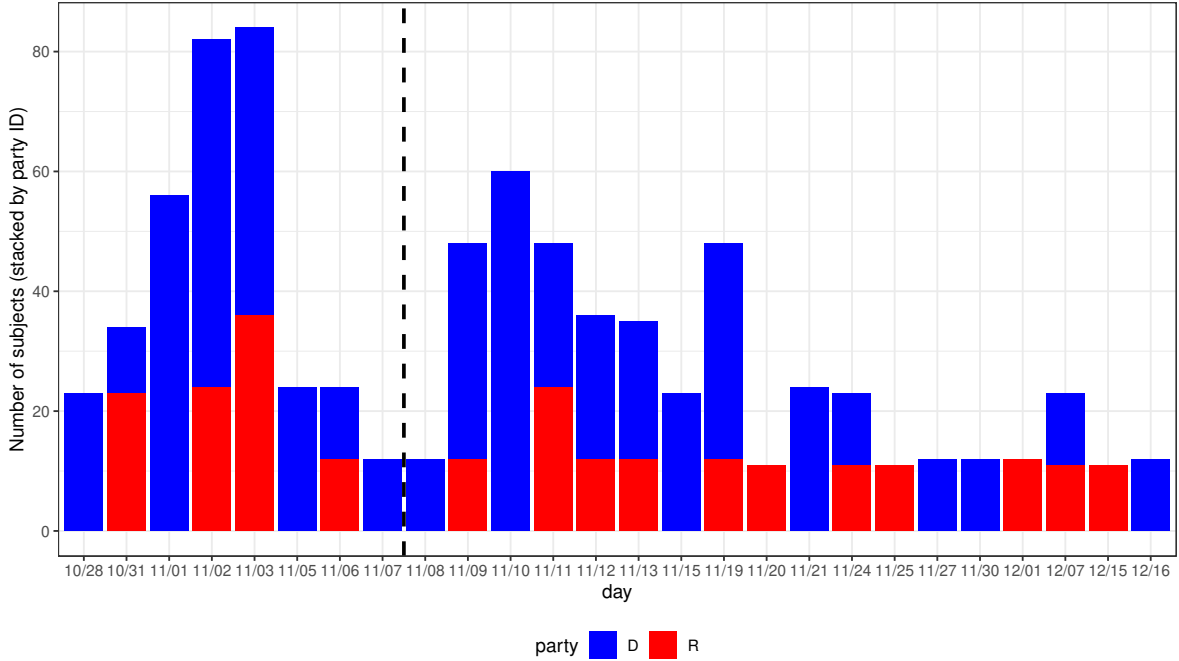
Though our selection procedure would have permitted moderate levels of polarization, the resultant panel of participants was highly polarised between the two candidates in the 2020 election season as can be discerned in Supplementary Figure 7. The panel shows a clear tendency of mutual opposition against the other party’s candidate. The experimental virtual lab sample consisting of the subjects drawn from this panel exhibits an analogous degree of mutual opposition.

**5.3 Study Scheduling** We conducted the study between 28/10/2020 and 16/12/2020, and collected group data on 26 days. Each week, participants from the panel were invited to fill out a short compensated scheduling survey, which allowed respondents to choose among five time slots per day in the following week. Our earliest sessions started at 9 am EST and the latest ones at 7 pm EST. Using this survey data, we pre-scheduled late morning, afternoon and evening sessions. For each of these sessions, we invited participants from two sources: 1) prior indicated availability and



**Supplementary Figure 7 | Feeling Thermometer Scores in the Participant Panel.** Bivariate scatter plots of feeling thermometer scores for Joe Biden and Donald Trump, shown for the final participant panel after sub-setting. The plot shows the count of cases for the combination of Trump and Biden feeling thermometer scores. Points in the top left quadrant are Republicans who disfavour Biden and favour Trump, while points in the bottom right quadrant are Democrats who disfavour Trump and favour Biden.

time zone, and 2) the general panel, irrespective of the time zone. Participants received a maximum of four invitations per week, and we counterbalanced Democratic and Republican sessions across days and times. Sessions lasted an average of 46 minutes. Supplementary Figure 8 shows the detailed participant scheduling outcomes—Democratic and Republican sessions were evenly spread out over time, though Republicans had worse turnout than Democrats (179 sessions were scheduled: 76 for Democrats and 103 for Republicans).



**Supplementary Figure 8 | Number of Subjects by Party and Day.** Total number of sampled participants in the coordination game by day. The x-axis shows the days on which we collected data. The y-axis shows the number of subjects joining the study on that day. Bars are stacked and additive, with blue indicating Democratic participants and red indicating Republican participants.

**5.4 Experimental Sample: Cleaning and Overview of the Group Data** The raw data-set contained 77 groups and 908 players. In a first step, we removed groups that, due to dropouts, did not have at least one post-intervention phase period. This was necessary because our comparison of interest is between the pre-intervention and post-intervention phase behaviours in the individual level analyses, and spillovers in the post-intervention phase for the group-level analyses. This resulted in the removal of 9 groups, leaving 68 unique groups and 805 players. This is the data-set used for all of the final analyses in the paper.

Our sample was split into 48 Democratic (570 subjects) and 20 Republican groups (235 participants). These groups were randomised into 35 neutral label groups (415 subjects) and 33 political label groups (390 subjects). We summarise the data—by treatment, party, groups and



participants—both pre and post-election in Supplementary Table 8. In total, 29 groups (N=344; 12 neutral and 17 political) played the game before the election, while 39 played after the election (N=461; 23 neutral and 16 political).

**Supplementary Table 8 | Counts, Experimental Sample, by Treatments and Party ID.** Counts of groups and participants joining the experiment a) through November 7, 2020 or b) after November 7, 2020

(a) Before and On Nov 7				(b) After Nov 7			
treatment	party	groups	players	treatment	party	groups	players
Neutral	D	7	84	Neutral	D	16	191
Neutral	R	5	60	Neutral	R	7	80
Political	D	14	164	Political	D	11	131
Political	R	3	36	Political	R	5	59

In Supplementary Table 9, we show a summary of the demographic composition of the final experimental sample used in all of our analyses. The sample’s demographic features largely correspond to the U.S. CENSUS (2019) in terms of age (M age US: 38.4 years), education (mode: High School followed by a few years of college), and household income (M: \$68,000).

**Supplementary Table 9 | Demographics, Experimental Sample.** Demographics statistics of the experimental sample (805 players). For response wording, consult section 7.2

variable	min	max	mean	sd
Age	18	77	40.74	13.42
Education	2	5	2.90	0.87
Income	2	9	5.35	2.01
Sex	0	1	0.45	0.50
Thermometer Biden	0	100	59.13	31.82
Thermometer Trump	0	100	26.51	36.60

**5.5 Communication with Subjects During the Experiment** Participants were contacted via the internal Prolific Academic messaging system. All messages were delivered anonymously. Furthermore, a virtual lab support chat system was in place during the study and monitored by research assistants in case of technical or logistical issues, or if a participant became non-responsive. All

communication was strictly related to technical or logistical issues during and after the game. The existence of the chat system was highlighted on the first page, to signal the possibility to chat with the researcher if assistance were needed. We did not use artificial interactive chat bots (besides an initial “hold the line” message). Instead, we used manually typed responses to preserve the credibility of the study in the eyes of participants. Anecdotal evidence from the chat record indicates that quick chat follow ups on logistical issues, e.g. relating to processing the payments or inquiries about waiting times in the waiting room, received positive evaluations by the participants.

**5.6 Randomization and Sample Composition by Treatment Group** The computer randomly assigned groups to the different treatments. Randomization was sequentially applied on the session level. In this section, we look at whether there were sizeable differences in background characteristics between our treatment groups. As shown in Supplementary Table 10, we do not find systematic differences between the treatment groups based on observable demographic characteristics, including age, gender or income conditional on the party affiliation of the group. Furthermore, we also assess the differences between Democrats and Republicans for the same demographics. We find no evidence of significant differences between the groups.

In general, online convenience samples (e.g. Mturk samples) report significant demographic differences for recruited participants across party lines, introducing uncontrolled imbalances into experiments dealing with partisanship<sup>12,13</sup>. The sample we recruited, while a convenience sample, is balanced across demographic characteristics for both treatment conditions and identity groups.

**Supplementary Table 10 | Demographics, Experimental Sample, by Treatment and Party ID.** Demographic variables by treatment and political groups. Numbers indicate the mean of each variable.

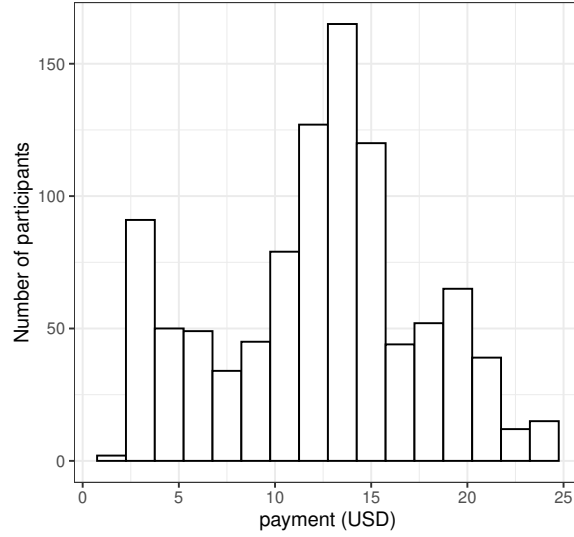
Treatment	Party	Age	Gender	Income	Payoff	Thermometer	
						Biden	Trump
Neutral	Republican	39.63	0.49	5.31	194.05	14.89	80.10
Political	Republican	38.61	0.47	5.40	199.30	14.42	80.74
Neutral	Democrat	41.37	0.43	5.35	207.83	78.84	4.15
Political	Democrat	41.35	0.43	5.39	206.74	76.03	3.74

**5.7 Participant Payments** The average payment per participant was \$12.8. Since the experiment was hosted on Prolific Academic, a U.K. based platform, show-up payments were denominated in British Pounds, £1.25. Subjects received an additional \$ 4 for completing the study, plus incentive payments. The incentive payments were calculated based on the outcomes in five randomly drawn periods (minimum possible was \$2.5, maximum possible was \$17.5). At the end of the session, dollar payments were converted into British Pounds at a prior stated rate of \$1.3 per GBP and payments were facilitated via Prolific Academic. The average market exchange rate was 1.33, implying that we slightly overpaid participants. Supplementary Figure 9 shows a histogram of payments made. Participants were fully informed about the entire payment procedure at the start of the game, including the random draw of five periods and currency conversion.

**5.8 Ethics and Preregistration** The study was approved by the ethics boards of the University of Lausanne ( “HEDGE 2”), the University of Bern (162020), and Princeton University (IRB 12733). Digital consent was obtained for the recruitment survey and the main experiment. The study was pre-registered on OSF, and is retrievable at [osf.io/6adbx](https://osf.io/6adbx).

## 6 Robustness Checks

**6.1 Dropouts** While all groups started with 12 participants, feedback about the group’s choices made in the prior period was provided always about a random sample of only 10 group members.



**Supplementary Figure 9 | Histogram of Payments to Participants, Experimental Sample.** This plot contains payments made to full and partial participants (dropouts).  $M = \$12.8$ ,  $sd = \$5.45$ .

If a participant left the game before it ended, and the player could not or would not reconnect, the current counterpart of the player who dropped out was also removed from the game for subsequent periods. If such a dropout happened with a group of 12 participants, the remaining group of 10 players was able to finish the session. Out of our 68 groups, 27 groups were affected by dropouts, out of which 12 groups ended prematurely. Overall, we did not find a statistically significant difference in the number of individuals who dropped out by treatment condition ( $\chi^2 = 0.10$ ,  $df = 1$ ,  $p = 0.68$ ).

**6.2 Participant Authentication** It is common—though often under-reported for incentivised on-line studies—that participants attempt to circumvent authentication barriers<sup>14,15</sup>. In our case, we were able to identify participants where authentication may have been circumvented based on the I.P. address and other information. To assess the potential impact of these individuals on our results, we conducted a robustness analysis. We removed all groups with a “suspect” player (24 groups) resulting in 44 remaining groups out of the initial 68 groups, and ran the group level spillover estimation again. The results are shown in Supplementary Table 11, first column. We then also

removed all groups with a suspect player and additionally, where at least one dropout happened. This leads to the removal 40 groups, resulting in 28 remaining groups used for analysis. While this is a very strict test, our results hold in both cases—spillovers are far less prevalent in the group identity than the neutral treatment.

**Supplementary Table 11 | Group Level Analysis: Excluding Groups.** We exclude selected groups in two categories: 1) groups which contain participants with invalid IP address (remove a total of 24 groups), 2) groups of the first category, plus those where at least one dropout occurred (remove a total of 40 groups). Results are from an OLS regression that models spillovers as a function of treatment. Robust standard errors are in parentheses. Spillovers were highly significantly positive in the neutral treatment (Intercept), and relabelling choice options in the identity treatment resulted in a large and highly significant reduction in beneficial spillovers.

	<i>Dependent variable: Spillovers</i>	
	suspect	suspect or dropout
Intercept	0.65*** (0.10)	0.74*** (0.10)
Identity	-0.82*** (0.15)	-0.76*** (0.18)

\*  $p \in (0.01, 0.05]$     \*\*  $p \in (0.001, 0.01]$     \*\*\*  $p \leq 0.001$

**6.3 Additional Robustness Analyses** To ensure the robustness of our main results (Main Study, Tables 2 and 3, Model 1) to model specification, we ran additional analyses and find that our primary results remain unchanged.

First, we included an additional robustness check for groups with dropouts. In case of a dropout, groups were reduced to a size of ten and it became then possible that less or more than 50% of that group was targeted in the post-intervention phase. The exact number depended on whether the dropouts included two targeted, one targeted and one non-targeted, or two-non-targeted participants. Due to the low number of cases, we pooled groups that deviated from the 50% targeting (i.e. groups with 4 or 6 targeted players out of 10). Supplementary Table 12 shows estimates from a model including only groups of size ten in the intervention phase, where the number targeted was

either 4 or 6 (left column), as compared to groups of size ten with 5 targeted participants (right column). Again, the main results hold.

Second, in Supplementary Table 13, we analyse the full individual model, with the same specification as in Table 2 in the Main Study, but subset to only those cases where all targeted individuals choose the alternative behaviour at least once. This “compliance” model directly subsets the sample to cases where the behaviour of the targeted individual matches the theoretical prediction at least in one period in the post-intervention phase, that is, where a targeted individual chooses the alternative behaviour. Again, the main results hold: non-targeted individuals pick up the alternative behaviour in the neutral treatment, yet in the identity treatment, this does not happen.

Third, to control for potential variation across specific days where data was collected, we included dummy variables for each of these days. As it can be seen in Table 14, the results also do not change, if we include day fixed effects. In the pre-intervention phase, the identity treatment leads to strong status quo norm (see also Main Study, Fig. 3). In the post-intervention phase identity treatment, however, non-targeted individuals do not pick up the alternative behaviour, while targeted individuals do. In the neutral treatment both targeted and non-targeted individuals pick up the alternative behaviour, however.

**Supplementary Table 12 | Individual Choice: 10 Player Groups.** Linear probability models for individual choices in the final period of pre-intervention and post-intervention phases. Cluster-robust standard errors are clustered at the group level. These analyses include only groups with ten players in the post-intervention period, instead of the twelve usual players. We further split these observations by groups where either 4 or 6 players were targeted after the dropout occurred, or where 5 players were targeted.

	<i>Choose Alternative Behaviour</i>	
	4 or 6 targeted	5 targeted
Intercept	0.08 (0.08)	0.08 (0.05)
(Neutral,T,Pre-int)	0.04 (0.21)	0.04 (0.07)
(Neutral,NT,Post-int)	0.72*** (0.09)	0.58*** (0.12)
(Neutral,T,Post-int)	0.88*** (0.06)	0.78*** (0.10)
(Political,NT,Pre-int)	-0.06 (0.09)	-0.08 (0.05)
(Political,T,Pre-int)	-0.08 (0.08)	-0.08 (0.05)
(Political,NT,Post-int)	0.13 (0.16)	0.16 (0.20)
(Political,T,Post-int)	0.52** (0.17)	0.56*** (0.13)
* $p \in (0.01, 0.05]$ ** $p \in (0.001, 0.01]$ *** $p \leq 0.001$		

**Supplementary Table 13 | Individual Choice: Compliance.** Linear probability models for individual choices in the final period of pre-intervention and post-intervention phases. Cluster-robust standard errors are clustered at the group level. This analysis includes all targeted participants  $P$  who chose  $c_P = 0$  at least once in any period of the post-intervention phase. In other words, here we exclude all players who did not choose the targeted alternative behaviour at all.

	<i>Choose Alternative Behaviour</i>
Intercept	0.13*** (0.02)
(Neutral,T,Pre-int)	-0.03 (0.02)
(Neutral,NT,Post-int)	0.63*** (0.05)
(Neutral,T,Post-int)	0.87*** (0.02)
(Political,NT,Pre-int)	-0.12*** (0.02)
(Political,T,Pre-int)	-0.12*** (0.02)
(Political,NT,Post-int)	0.09 (0.06)
(Political,T,Post-int)	0.87*** (0.02)
* $p \in (0.01, 0.05]$ ** $p \in (0.001, 0.01]$ *** $p \leq 0.001$	



**Supplementary Table 14 | Individual Choice: Day Fixed Effects.** Linear probability models for individual choices in the final period of pre-intervention and post-intervention phases. Cluster-robust standard errors are clustered at the group level. This analysis includes fixed effect dummy variables for the 26 days of the study (not reported here).

	<i>Choose Alternative Behaviour</i>
Intercept	0.18*** (0.04)
(Neutral,T,Pre-int)	-0.03 (0.02)
(Neutral,NT,Post-int)	0.63*** (0.05)
(Neutral,T,Post-int)	0.81*** (0.03)
(Political,NT,Pre-int)	-0.13*** (0.03)
(Political,T,Pre-int)	-0.13*** (0.03)
(Political,NT,Post-int)	0.07 (0.07)
(Political,T,Post-int)	0.52*** (0.06)
* $p \in (0.01, 0.05]$ ** $p \in (0.001, 0.01]$ *** $p \leq 0.001$	

## 7 Instructions and Questionnaires

In this section, we include the instructions for the coordination game as participants saw them, as well as the image pre-test and the recruitment survey questionnaires. We included the image pre-test and recruitment surveys directly downloaded from Qualtrics so that our variable coding and skip logic is visible.

**7.1 Experimental Instructions for the Coordination Game** We include here the full instructions as participants saw them. They were able to download this document and make reference to it throughout the game. Note that both the initial instructions and the intervention phase instructions are included here. The break between the two is signalled by the phrase "Detailed Instructions Part 2". Participants were able to download the instructions, though when they went through them before the game, the instructions were split into manageable chunks per page.

## 1/3 - Introduction

This document is also available as pdf.

[\\*\\*\\* OPEN PDF IN NEW TAB \\*\\*\\*](#)

### Welcome!

You are about to participate in a study. In this study, you will earn money that will be directly transferred to your payment account.

You are likely to earn more money if you:

- read the instructions carefully,
- follow these instructions to the letter,
- and think hard about your decisions.

If you have questions while reading the instructions or during the study, please do not hesitate to contact the researcher, using the information provided in the invitation email.

Your earnings will be in part based on your decisions during the study, and they will be calculated in points. At the end of the study, the points you earned will be converted into US Dollars at the following exchange rate:

**100 points = \$1**

**This means one point is 1 cent**

In addition to these earnings, you will also receive a fixed amount of \$4 for completing the study plus \$1.5 (£1.25 on prolific) for waiting up to 10 minutes at the start of the study for other participants to arrive. The study consists of several rounds. At the end of the session, your total payment will be \$4 + \$1.5 + the sum of your earnings on five randomly selected rounds.

You will never know who the other participants in this study are, and the other participants will never know who you are. Your identity and decisions will remain anonymous.

### General Set-up

The study is made up of two parts: Part 1 and Part 2. We will make it very clear when you are in Part 1 and when you are in Part 2.

- Part 1 will last between 10 and 20 rounds.
- Part 2 will last 25 rounds.

The computer will randomly assign you to a group of 12 players. Thus, together with 11 others, you will form a group of 12 participants. Please note: **The group will remain the same through parts 1 and 2 of the study.**

In every round, the computer will randomly match you with a participant, or “counterpart” from your group. It is unlikely you will be matched with the same person back to back. You will not know with what

other person you are matched. Thus, you will not know if it is a person you have been matched with before or a person you have never been matched with.

Here's an EXAMPLE of how this might happen. In this example you are "Participant 2". Please note that in the actual study you cannot see who the other participant you are matched with is.

Example sequence, rounds 1 - 4.		
Round	You	Your counterpart
Round 1	Participant 2	Participant 9
Round 2	Participant 2	Participant 11
Round 3	Participant 2	Participant 1
Round 4	Participant 2	Participant 9
...	...	...

This is just an example to show you how the random rematching works. **In the study, you will not know who you have been matched with in each round.**

In every round you and your counterpart for that round will each make a decision. We will now explain in detail how the game works.

Next

[Page Break]

\*\*\*\*\*

2/3

Detailed Instructions Part 1

On a given round, you and your counterpart will each simultaneously and privately choose between two options. These options will be represented by a choice label and will determine your payoffs on any given round. Your payoff on a given round will also depend on the choice your counterpart made on that round.

The labels may be neutral symbols or images that carry meaning. You should click on the choice label to indicate the option you are choosing in the game.

As an example, if the labels were # (“pound”) and @ (“at”), your choice table would look like this:

Your payoff in each round will depend on both **your choice and your counterpart’s choice** as follows:

		Choice of your counterpart	
		#	@
Your choice	#	200	50
	@	50	200

You should read this “payoff table” like this:

- if you choose # and your counterpart also chooses #, you will earn 200 points
- if you choose # and your counterpart chooses @, you will earn 50 points
- if you choose @ and your counterpart chooses #, you will earn 50 points
- if you choose @ and your counterpart also chooses @, you will earn 200 points

Your payoff table will be the same throughout Part 1. Everyone will see the same payoff table. In each round, you and your counterpart will each choose between one of the two choice labels. In this example, # (“Pound”) or @ (“At”).

You will not be able to see what your counterpart has chosen before you decide.

Remember, this is just an **example to familiarize you with the game!**

Importantly, each counterpart you are paired with will see the same payoff table as you. Also, everyone in your group will be randomly re-matched with a counterpart in every new round.

[Back](#)

[Next](#)

[Page Break]

\*\*\*\*\*

### 3/3 (Examples)

Now we will show you an example of a payoff table from your perspective and from that of your counterpart. Note that in the actual game, you will not see your counterpart’s payoff table.

#### EXAMPLE 1

Example 1							
Your payoff table				The counterpart’s payoff table			
		Choice of <b>Counterpart</b>				Your Choice	
		#	@			#	@
Your Choice	#	200	50	Choice of <b>Counterpart</b>	#	200	50
	@	50	200		@	50	200

- If you choose # and your counterpart chooses #, both you and your counterpart will receive 200 points
- If you choose # and your counterpart chooses @, both you and your counterpart will receive 50 points
- If you choose @ and your counterpart chooses #, both you and your counterpart will receive 50 points
- If you choose @ and your counterpart chooses @, both you and your counterpart will receive 200 points

## Payments

As mentioned above, you will play up to 45 rounds across the entire session today. As you play, you will tally points in each round. **When all rounds have been played, the computer will randomly select 5 rounds for payment, and you will be paid based on the total points you earned in those 5 rounds.**

For example, if the computer draws rounds 2,7,19, 30, 33, and you earned the following points in these rounds.

Randomly selected round	Points
2	200
7	50
19	200
30	200
33	50
<b>TOTAL POINTS</b>	<b>700</b>

The sum of points in the five randomly selected rounds is 700 points. Because 100 Points = \$1, you would earn 700 points = \$7. Your total earnings in the game would be **\$7 for your decisions + \$4 for participation + \$1.5 or waiting up to 10 minutes. Thus, you would make \$12.5 total.**

Also for technical assistance, you can find our contact information in the invitation email, or contact us directly at [pacelab@unil.ch](mailto:pacelab@unil.ch)

All payments made in £ GBP via prolific, converted from USD at announced rate.

## Questionnaire

At the end of the game, we will ask you some brief questions about your experience.

**\*\*\*End of Instructions\*\*\***

**[\\*\\*\\* OPEN PDF VERSION OF INSTRUCTIONS IN NEW WINDOW \\*\\*\\*](#)**

Quiz time! By clicking "Next", you will start with a quiz first, followed by the interaction rounds.

**Back**

**Next**

[Page Break]

\*\*\*\*\*

[Directly after instructions:]

## Quiz

\*\*\* OPEN Instructions (PDF) IN NEW TAB \*\*\*

1. How many rounds will you play in this study in Part 1?
  - a) 10 rounds
  - b) 20 rounds
  - c) Between 10 and 20
  
2. How many rounds will you play in Part 2?
  - a) 25 rounds
  - b) Between 30 and 40 rounds
  - c) 20 rounds
  
3. How many rounds will you be paid for?
  - a) I will be paid for all the rounds I played.
  - b) I will be paid for 1 randomly selected round.
  - c) I will be paid for 5 randomly selected rounds.
  
4. How many people will be in one group?
  - a) 18
  - b) 12
  - c) 9
  
5. Will the members of the group change between Part 1 and Part 2?
  - a) Yes
  - b) No
  - c) Depends on the group

6. In Part 1, do the other members of your group see the same payoff table as you?

- a) Yes
- b) No
- c) Not sure

7. Within Part 1, do you keep the same payoff table from one round to the next.

- a) Yes
- b) No
- c) Not sure

8. Using the following payoff table, how many points do you get when you choose # and your counterpart chooses @?

		Choice of the <b>Counterpart</b>	
		#	@
Your Choice	#	350	50
	@	50	350

- a) 350
- b) 50
- c) 100

**Next**

[Page Break]

\*\*\*\*\*



## Quiz Results

[\\*\\*\\* OPEN Instructions \(PDF\) IN NEW TAB \\*\\*\\*](#)

Your answers to the quiz questions:

1. How many rounds will you play in Part 1?

Correct answer: **10 to 20 rounds**

2. How many rounds will you play in Part 2?

Correct answer: **25 rounds**

3. How many rounds will you be paid for?

Correct answer: **I will be paid for 5 randomly selected rounds**

4. How many people will be in one group?

Correct answer: **12**

5. Will the members of the group change between Part 1 and Part 2?

Correct answer: **No**

6. In Part 1, do the other members of your group see the same payoff table as you?

Correct answer: **Yes**

7. In Part 1, do you keep the same payoff table from one round to the next?

Correct answer: **Yes**

8. Using the following payoff table, how many points do you get when you choose # and your counterpart chooses @?

Correct answer: **50**

**Congratulations! You passed all questions. You are ready to start the game!**

**Click "NEXT" to start !**

[Page Break]

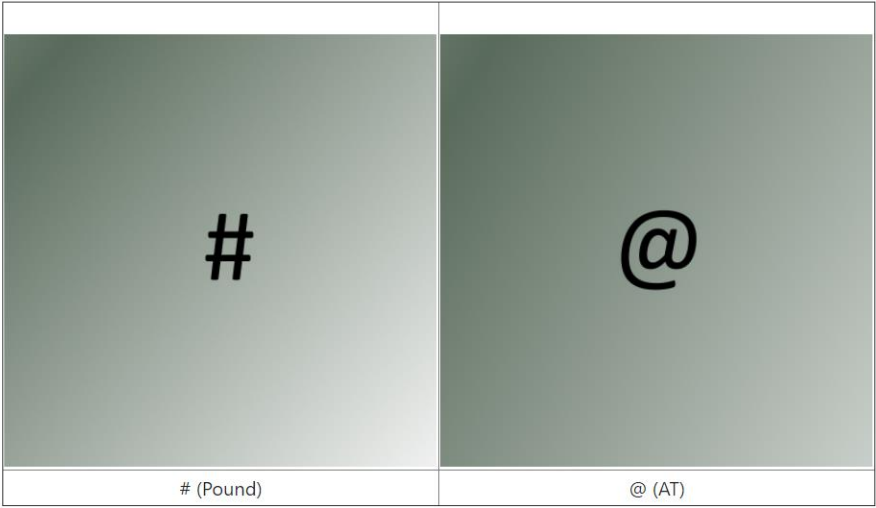
\*\*\*\*\*

[Start, Part 1]

[Part 1, first page before start of decision making]

[For Neutral Treatment]

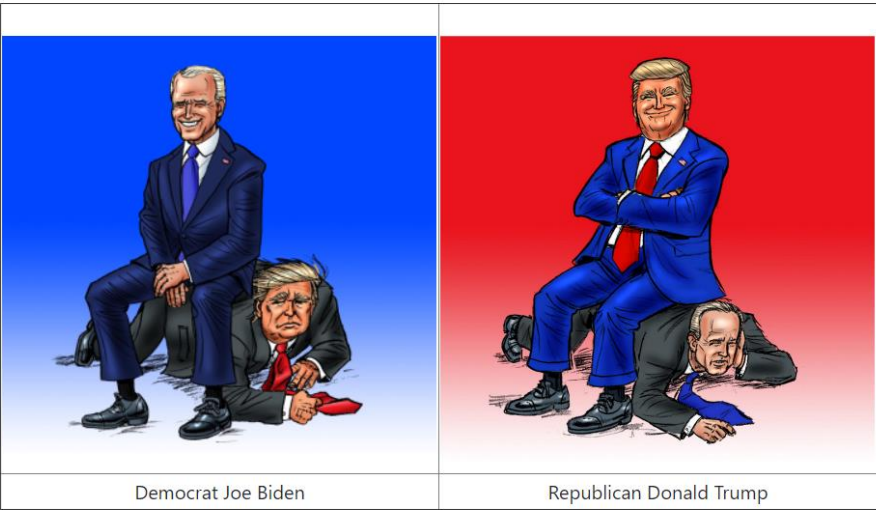
On the next pages, the choice options are represented by "#" and "@" buttons - next page, you should press directly on the image corresponding to your preferred choice.



Next

[For Identity Treatment]

On the next pages, the choice options are represented by "Joe Biden" and "Donald Trump" buttons - next page, you should press directly on the image corresponding to your preferred choice.



Next

[Page Break]

\*\*\*\*\*

[Start Game, Part 1]

[Shown only in the first round of Part 2]

## You are now entering Part 2 of the game!

Please pay attention to the following information. It is important that you understand what has changed.

- In part 2, you will play for 25 more rounds.
- Your group is the same as in Part 1 and will stay the same for the duration of Part 2.
- You will still be randomly paired with counterparts from one round to the next.

### Detailed Instructions Part 2

The payoff table is in principle the same as in Part 1. However, for some people in the group (including yourself), the payoffs associated with the different choices may have changed. Thus, some members of the group have different payoffs in Part 2, while others keep the same payoffs as in Part 1.

Now we will show you some examples.

- Some members of your group have new payoff tables that differ from those in Part 1.
- Other members of the group have the same payoff table from Part 1.
- Each person's payoff table will remain fixed for all of Part 2, or the remaining 25 rounds.

A person with a <b>new</b> payoff table			
		Counterpart	
		@	#
This person's Choice	@	200	50
	#	350	350

A person with the <b>same</b> payoff table			
		Counterpart	
		@	#
This person's choice	@	200	50
	#	50	200

\*Note the order of rows and columns could be different for you

[Shown to the targeted subjects]

*You have a new payoff table!  
This will be your payoff table for the remaining 25 rounds!*

[Show to the non-targeted subjects]

*You have the same payoff table!  
This will be your payoff table for the remaining 25 rounds!*

We are ready to begin Part 2!

**Next**

[Start Game, Part 2]

**7.2 Questionnaire Items: Recruitment and Image Pre-test** The following two instruments are the 1) questionnaire used for recruiting individuals on Prolific for the purpose of building a participant panel that met our participant criteria (“Recruitment Questionnaire”), and 2) the questionnaire used for pretesting the political images (“Pretest Questionnaire”), which was run on a separate sample at a separate time. Both documents are direct downloads from Qualtrics and contain the raw *variable names* and the coding scheme used for the variables, *Comment* in grey, and display and conditional logics.

# Recruitment Questionnaire

---

*dem\_live*

Are you currently living in the U.S.?

☐ Yes (1)

☐ No (2)

---

*dem\_sex*

What is your sex?

☐ Male (1)

☐ Female (2)

☐ Other (3) \_\_\_\_\_

---

*dem\_age*

How old are you?

\_\_\_\_\_

---

*dem\_edu*

What is the highest level of education you have completed?

☐ No High School (1)

☐ High School (2)

☐ Bachelor degree (3)

☐ Graduate degree or higher (4)

*dem\_income*

What is the typical yearly income of your household?

- ☐ Less than \$25,000 (1)
- ☐ \$25,000–\$35,000 (2)
- ☐ \$35,001–\$50,000 (3)
- ☐ \$50,001–\$75,000 (4)
- ☐ \$75,001–\$100,000 (5)
- ☐ \$100,001–\$150,000 (6)
- ☐ \$150,001–\$250,000 (7)
- ☐ More than \$250,000 (8)

*id\_ideal*

When it comes to politics would you describe yourself as liberal, conservative, or neither liberal nor conservative?

- ☐ Very liberal (1)
- ☐ Somewhat liberal (2)
- ☐ Closer to liberals (3)
- ☐ Neither liberal nor conservative (4)
- ☐ Closer to conservatives (5)
- ☐ Somewhat conservative (6)
- ☐ Very conservative (7)

*id\_party*

Generally speaking, do you usually think of yourself as a Republican, a Democrat, an Independent, or something else?

- ☐ Republican (1)
  - ☐ Democrat (2)
  - ☐ Independent (3)
  - ☐ Other (4)
- 

**Comment**

*Display This Question:*

*If Generally speaking, do you usually think of yourself as a Republican, a Democrat, an Independent,... = Republican*

*id\_rep*

Would you consider yourself a strong Republican or a not very strong Republican?

- ☐ Strong (1)
  - ☐ Not very strong (2)
- 

**Comment**

*Display This Question:*

*If Generally speaking, do you usually think of yourself as a Republican, a Democrat, an Independent,... = Democrat*

*id\_dem*

Would you consider yourself a strong Democrat or a not very strong Democrat?

- ☐ Strong (1)
- ☐ Not very strong (2)



**Comment**

*Display This Question:*

*If Generally speaking, do you usually think of yourself as a Republican, a Democrat, an Independent,... = Independent*

*id\_ind*

Do you think of yourself as closer to the Republican Party or to the Democratic party?

☐ Lean Republican (1)

☐ Lean Democratic (2)

---

**Comment**

*Display This Question:*

*If Generally speaking, do you usually think of yourself as a Republican, a Democrat, an Independent,... = Other*

*id\_other*

Do you think of yourself as closer to the Republican Party or to the Democratic party?

☐ Lean Republican (1)

☐ Lean Democratic (2)

---

*id\_self*

How big a part does being a/an \${id\_party/ChoiceGroup/SelectedChoices} play in how you see yourself?

- ☐ None (1)
  - ☐ Small (2)
  - ☐ Moderate (3)
  - ☐ Large (4)
  - ☐ Very Large (5)
- 

*elec\_pct*

Thinking about the votes cast for the two major parties, what percentage of the vote do you think Biden and Trump will each receive in the national vote? The total of your responses should not exceed 100.

Biden percentage : \_\_\_\_\_ (1)  
Trump percentage : \_\_\_\_\_ (2)  
Total : \_\_\_\_\_

---

*elec\_pred*

Who do you think will be elected President in November?

- ☐ Biden (2)
- ☐ Trump (1)
- ☐ Other (3) \_\_\_\_\_

*elec\_care*

How much do you care who wins the presidential election this fall?

- ☐ A great deal (1)
  - ☐ A lot (2)
  - ☐ A moderate amount (3)
  - ☐ A little (4)
  - ☐ None at all (5)
- 

*elec\_vote*

In the 2020 presidential election between Donald Trump for the Republican Party and Joe Biden for the Democratic Party, will you vote for Donald Trump, Joe Biden, someone else, or probably not vote?

- ☐ Donald Trump (1)
  - ☐ Joe Biden (2)
  - ☐ Someone else (3)
  - ☐ Probably not vote (4)
  - ☐ I am not eligible to vote (5)
-

*elec\_count*

In the November 2020 general election, how accurately do you think the votes will be counted?

- ☐ Not at all accurately (1)
  - ☐ Not very accurately (2)
  - ☐ Moderately accurately (3)
  - ☐ Very accurately (4)
  - ☐ Completely accurately (5)
- 

*elec\_covid*

Do you think that the COVID-19 pandemic will impact voter turnout and bias or distort election results?

- ☐ Yes, bias in favor of Republicans (1)
- ☐ Yes, bias in favor of Democrats (2)
- ☐ No (3)
- ☐ Unsure (4)

*psyc\_char*

Rate the extent to which you feel each of the following descriptive adjectives is characteristic or uncharacteristic of your personality and behavior.

	Very uncharacteristic (5)	Moderately uncharacteristic (4)	Neutral (3)	Moderately characteristic (2)	Very characteristic (1)
Rebellious (psyc_char_reb)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Unorthodox (psyc_char_unor)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Conforming (psyc_char_conf)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Conventional (psyc_char_conv)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Old-fashioned (psyc_char_trad)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Free-living (psyc_char_free)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Non-conforming (psyc_char_nconf)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Moralistic (psyc_char_moral)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Obedient (psyc_char_obed)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Unconventional (psyc_char_unconv)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Unpredictable (psyc_char_unpred)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Erratic (psyc_char_err)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Respectful (psyc_char_resp)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Predictable (psyc_char_pred)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

*psyc\_plan*

Please indicate how much you agree or disagree with the following questions.

	Strongly disagree (1)	Somewhat disagree (2)	Neither agree nor disagree (3)	Somewhat agree (4)	Strongly agree (5)
I do something I want to do even if no one else wants to do it. (psyc_plan_1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I never keep at an idea (or plan) when I know I am wrong. (psyc_plan_2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

*psyc\_self*

Please rate the extent to which you agree/disagree with the following statements, using a 5-point scale ranging from strongly disagree to strongly agree.

	Strongly disagree (1)	Somewhat disagree (2)	Neutral (3)	Somewhat agree (4)	Strongly agree (5)
I rely on myself most of the time; I rarely rely on others (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My personal identity, independent of others, is very important to me (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I feel good when I cooperate with others (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It is important to me that I respect the decisions made by my groups (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I generally consider changes to be a negative thing (6)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My views are very consistent over time (8)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I will sacrifice my self-interest for the benefit of the group I am in (10)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

psyc\_sdo7

Show how much you favor or oppose each idea below. You can work quickly; your first feeling is generally best.

	Strongly oppose (1)	Somewhat oppose (2)	Neutral (3)	Somewhat favor (4)	Strongly favor (5)
An ideal society requires some groups to be on top and others to be on the bottom. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Groups at the bottom are just as deserving as groups at the top. (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
No one group should dominate in society. (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
We should not push for group equality. (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Group equality should not be our primary goal. (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
We should do what we can to equalize conditions for different groups. (6)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
No matter how much effort it takes, we ought to strive to ensure that all groups have the same chance in life. (7)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

*psyc\_tipi*

We're interested in how you see yourself. Please mark how well the following pair of words describes you, even if one word describes you better than the other.

	Extremely poorly (1)	Somewhat poorly (2)	Neither poorly nor well (3)	Somewhat well (4)	Extremely well (5)
'extraverted, enthusiastic' (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
'critical, quarrelsome' (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
'dependable, self- disciplined' (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
'anxious, easily upset' (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
'open to new experiences, complex' (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
'reserved, quiet' (6)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
'sympathetic, warm' (7)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
'disorganized, careless' (8)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
'calm, emotionally stable' (9)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
'conventional, uncreative' (10)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



To what extent do you agree or disagree with the following statement?

	Strongly agree (5)	Somewhat agree (4)	Neither agree nor disagree (3)	Somewhat disagree (2)	Strongly disagree (1)
Differences in people's standards of living should be small	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

We now ask for your willingness to act in a certain way. Please again indicate your answer on a scale from 0 to 10, where 0 means you are “completely unwilling to do so” and a 10 means you are “very willing to do so”. You can also use any numbers between 0 and 10 to indicate where you fall on the scale, like 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.

[illegible]

**COMMENT**

*Display This Question:*

*If Generally speaking, do you usually think of yourself as a Republican, a Democrat, an Independent,... = Democrat*

*Or If Generally speaking, do you usually think of yourself as a Republican, a Democrat, an Independent,... = Independent*

*And Do you think of yourself as closer to the Republican Party or to the Democratic party? = Lean Democratic*

*Or Generally speaking, do you usually think of yourself as a Republican, a Democrat, an Independent,... = Other*

*And Do you think of yourself as closer to the Republican Party or to the Democratic party? = Lean Democratic*

*aff\_dem*

Would you say that you are a Democrat because you are for what the Democratic party represents, or are you more against what the Republican party represents?

- ☐ For what the Democratic party represents (1)
- ☐ Against what the Republican party represents (2)
- ☐ Not sure (3)

---

**COMMENT**

*Display This Question:*

*If Generally speaking, do you usually think of yourself as a Republican, a Democrat, an Independent,... = Republican*

*Or If Generally speaking, do you usually think of yourself as a Republican, a Democrat, an Independent,... = Independent*

*And Do you think of yourself as closer to the Republican Party or to the Democratic party? = Lean Republican*

*Or Do you think of yourself as closer to the Republican Party or to the Democratic party? = Lean Republican*

*aff\_rep*

Would you say that you are a Republican because you are for what the Republican party represents, or are you more against what the Democratic party represents?

- ☐ For what the Republican party represents (1)
- ☐ Against what the Democratic party represents (2)
- ☐ Not sure (3)

*therm\_biden*

The next four questions will ask you to rate your feelings on a scale of 0 to 100, using a "feeling thermometer". On this feeling thermometer scale, ratings between 0 and 49 degrees mean that you feel unfavorable and cold (with 0 being the most unfavorable/coldest). Ratings between 51 and 100 degrees mean that you feel favorable and warm (with 100 being the most favorable/warmest). A rating of 50 means you have no feelings one way or the other.

How would you rate your feeling towards **Joe Biden** on this feeling thermometer? 0 is the most unfavorable/coldest, 50 is neutral, and 100 is the most favorable/warmest.

	0	10	20	30	40	50	60	70	80	90	100
Rating											

*therm\_dem*

The next four questions will ask you to rate your feelings on a scale of 0 to 100, using a "feeling thermometer". On this feeling thermometer scale, ratings between 0 and 49 degrees mean that you feel unfavorable and cold (with 0 being the most unfavorable/coldest). Ratings between 51 and 100 degrees mean that you feel favorable and warm (with 100 being the most favorable/warmest). A rating of 50 means you have no feelings one way or the other.

How would you rate your feeling towards the **Democratic Party** on this feeling thermometer? 0 is the most unfavorable/coldest, 50 is neutral, and 100 is the most favorable/warmest.

	0	10	20	30	40	50	60	70	80	90	100
Rating											

### *therm\_trump*

The next four questions will ask you to rate your feelings on a scale of 0 to 100, using a "feeling thermometer". On this feeling thermometer scale, ratings between 0 and 49 degrees mean that you feel unfavorable and cold (with 0 being the most unfavorable/coldest). Ratings between 51 and 100 degrees mean that you feel favorable and warm (with 100 being the most favorable/warmest). A rating of 50 means you have no feelings one way or the other.

How would you rate your feeling towards **Donald Trump** on this feeling thermometer?  
0 is the most unfavorable/coldest, 50 is neutral, and 100 is the most favorable/warmest.

	0	10	20	30	40	50	60	70	80	90	100
Rating											

### *therm\_rep*

The next four questions will ask you to rate your feelings on a scale of 0 to 100, using a "feeling thermometer". On this feeling thermometer scale, ratings between 0 and 49 degrees mean that you feel unfavorable and cold (with 0 being the most unfavorable/coldest). Ratings between 51 and 100 degrees mean that you feel favorable and warm (with 100 being the most favorable/warmest). A rating of 50 means you have no feelings one way or the other.

How would you rate your feeling towards the Republican Party on this feeling thermometer?  
0 is the most unfavorable/coldest, 50 is neutral, and 100 is the most favorable/warmest.

	0	10	20	30	40	50	60	70	80	90	100
Rating											

### *feedback*

Do you have any feedback for us regarding this part?

---

---

---

---

# Pretest Questionnaire

dem\_age

What is your age in years?

---

dem\_sex

What is your gender?

☐ Male (1)

☐ Female (2)

☐ Other (3) \_\_\_\_\_

dem\_ideo

When it comes to politics, would you describe yourself as liberal, conservative, or neither liberal nor conservative? Select one answer.

☐ Very liberal (1)

☐ Somewhat liberal (2)

☐ Closer to liberal (3)

☐ Neither liberal nor conservative (4)

☐ Closer to conservative (5)

☐ Somewhat conservative (6)

☐ Very conservative (7)

dem\_pol

Generally speaking, do you usually think of yourself as a Republican, a Democrat, an Independent, or something else?

- ☐ Republican (1)
- ☐ Democrat (2)
- ☐ Independent (3)
- ☐ Other (4)

---

**COMMENT**

Display This Question:

If Generally speaking, do you usually think of yourself as a Republican, a Democrat, an Independent,... = Republican

dem\_rep

Would you consider yourself a strong Republican or not a very strong Republican?

- ☐ Strong Republican (1)
- ☐ Not Very Strong Republican (2)

---

**COMMENT**

Display This Question:

If Generally speaking, do you usually think of yourself as a Republican, a Democrat, an Independent,... = Democrat

dem\_dem

Would you consider yourself a strong Democrat or not a very strong Democrat?

- ☐ Strong Democrat (1)
- ☐ Not Very Strong Democrat (2)

## COMMENT

Display This Question:

If Generally speaking, do you usually think of yourself as a Republican, a Democrat, an Independent,... = Independent

dem\_ind

Do you think of yourself as closer to the Republican Party or to the Democratic Party?

☐ Independent Leaning Republican (1)

☐ Independent Leaning Democratic (2)

---

## Display This Question:

If Generally speaking, do you usually think of yourself as a Republican, a Democrat, an Independent,... = Other

dem\_other

Do you think of yourself as closer to the Republican Party or to the Democratic Party?

☐ Other Leaning Republican (1)

☐ Other Leaning Democratic (2)

who\_b-im



who\_biden  
Who is this?

---

who\_b-info



This is Joe Biden, the democratic presidential candidate.



who\_t-im



who\_trump  
Who is this?

---

who\_t-info



This is Donald Trump, the current Republican President.



im-hoodie1

Please use the bar below to indicate the maximum price in \$ you would be willing to pay for this hoodie.

[Political image hoodie item here]

	Price in \$										
	0	5	10	15	20	25	30	35	40	45	50
Max \$ you're willing to pay for this hoodie? (1)											

### COMMENT

Display This Question:

If Generally speaking, do you usually think of yourself as a Republican, a Democrat, an Independent,... = Republican

Or If Generally speaking, do you usually think of yourself as a Republican, a Democrat, an Independent,... = Independent

And Do you think of yourself as closer to the Republican Party or to the Democratic Party? = Independent Leaning Republican

Or If Generally speaking, do you usually think of yourself as a Republican, a Democrat, an Independent,... = Other

And Do you think of yourself as closer to the Republican Party or to the Democratic Party? = Other Leaning Republican

Im-mugRep

We need your advice! We've been hired to design mugs in the lead up to the election as a way to raise money for the Republican campaign. We designed four prototypes and we need to pick one of them. They are similar but all have slightly different images. The money from the sale of the selected mug will be donated to the Republican party.

Which mug should we pick in order to make the most money for the campaign? When answering, please try to guess which mug most Republicans would prefer.



## COMMENT

*Display This Question:*

*If Generally speaking, do you usually think of yourself as a Republican, a Democrat, an Independent,... = Democrat*

*Or If Generally speaking, do you usually think of yourself as a Republican, a Democrat, an Independent,... = Independent*

*And Do you think of yourself as closer to the Republican Party or to the Democratic Party? = Independent Leaning Democratic*

*Or If Generally speaking, do you usually think of yourself as a Republican, a Democrat, an Independent,... = Other*

*And Do you think of yourself as closer to the Republican Party or to the Democratic Party? = Other Leaning Democratic*

Im-mugDem

We need your advice! We've been hired to design mugs in the lead up to the election as a way to raise money for the Democratic campaign. We designed four prototypes and we need to pick one of them. They are similar but all have slightly different images. The money from the sale of the selected mug will be donated to the Democratic party.

Which mug should we pick in order to make the most money for the campaign? When answering, please try to guess which mug most Democrats would prefer.



im-instr

Next, we will ask you to look at pairs of images and indicate how much you prefer one of the images over the other one.

Im-tb\_neutral

Please use the slider bar to indicate your preference between the two images. If you place the bar all the way to the left, this indicates that you prefer the image on the left much more than the one on the right. If you place the bar all the way to the right, this indicates that you prefer the image on the right much more than the one on the left. If you place the slider bar in the middle this indicates that you are indifferent between the two images.

Which one of these two images do you prefer?



100 80 60 40 20 0 20 40 60 80 100

Im-bt\_neutral

Please use the slider bar to indicate your preference between the two images. If you place the bar all the way to the left, this indicates that you prefer the image on the left much more than the one on the right. If you place the bar all the way to the right, this indicates that you prefer the image on the right much more than the one on the left. If you place the slider bar in the middle this indicates that you are indifferent between the two images.

Which one of these two images do you prefer?



100 80 60 40 20 0 20 40 60 80 100

Im-tb\_positive

Please use the slider bar to indicate your preference between the two images. If you place the bar all the way to the left, this indicates that you prefer the image on the left much more than the one on the right. If you place the bar all the way to the right, this indicates that you prefer the image on the right much more than the one on the left. If you place the slider bar in the middle this indicates that you are indifferent between the two images.

Which one of these two images do you prefer?



100 80 60 40 20 0 20 40 60 80 100

Im-bt\_positive

Please use the slider bar to indicate your preference between the two images. If you place the bar all the way to the left, this indicates that you prefer the image on the left much more than the one on the right. If you place the bar all the way to the right, this indicates that you prefer the image on the right much more than the one on the left. If you place the slider bar in the middle this indicates that you are indifferent between the two images.

Which one of these two images do you prefer?



100 80 60 40 20 0 20 40 60 80 100

Im-tb\_vp

Please use the slider bar to indicate your preference between the two images. If you place the bar all the way to the left, this indicates that you prefer the image on the left much more than the one on the right. If you place the bar all the way to the right, this indicates that you prefer the image on the right much more than the one on the left. If you place the slider bar in the middle this indicates that you are indifferent between the two images.

Which one of these two images do you prefer?



100 80 60 40 20 0 20 40 60 80 100

Im-bt\_vp

Please use the slider bar to indicate your preference between the two images. If you place the bar all the way to the left, this indicates that you prefer the image on the left much more than the one on the right. If you place the bar all the way to the right, this indicates that you prefer the image on the right much more than the one on the left. If you place the slider bar in the middle this indicates that you are indifferent between the two images.

Which one of these two images do you prefer?



100 80 60 40 20 0 20 40 60 80 100

Im-tb\_victory

Please use the slider bar to indicate your preference between the two images. If you place the bar all the way to the left, this indicates that you prefer the image on the left much more than the one on the right. If you place the bar all the way to the right, this indicates that you prefer the image on the right much more than the one on the left. If you place the slider bar in the middle this indicates that you are indifferent between the two images.

Which one of these two images do you prefer?



100 80 60 40 20 0 20 40 60 80 100

Im-bt\_victory

Please use the slider bar to indicate your preference between the two images. If you place the bar all the way to the left, this indicates that you prefer the image on the left much more than the one on the right. If you place the bar all the way to the right, this indicates that you prefer the image on the right much more than the one on the left. If you place the slider bar in the middle this indicates that you are indifferent between the two images.

Which one of these two images do you prefer?



100 80 60 40 20 0 20 40 60 80 100



Please take a good look at the image below. The following questions will all pertain to this image.

[Political image item here]

im-partyrep

Do you think this image is pro- or anti- Republican? Please use the scale: 0 means very anti-Republican, 100 means very pro-Republican, and 50 means neither pro nor anti Republican.

	Very Anti-Republicans					Very Pro-Republicans					
	0	10	20	30	40	50	60	70	80	90	100
Rating											

im-partydem

Do you think this image is pro- or anti- Democrat? Please use the scale: 0 means very anti-Democrat, 100 means very pro-Democrat, and 50 means neither pro nor anti Republican.

	Very Anti-Republicans					Very Pro-Republicans					
	0	10	20	30	40	50	60	70	80	90	100
Rating											

im-valence

How positive, negative or neutral is your feeling about this image?

- ☐ Very positive (1)
- ☐ Moderately positive (2)
- ☐ Neither positive nor negative (3)
- ☐ Moderately negative (4)
- ☐ Very negative (5)

im-feel

What feeling does this image evoke for you? Please select all that apply

- ☐ Anger (10)
- ☐ Fear (11)
- ☐ Disgust (12)
- ☐ Happiness (13)
- ☐ Sadness (14)
- ☐ Surprise (15)
- ☐ Pain (16)
- ☐ Pleasure (17)
- ☐ Other (18) \_\_\_\_\_

## 8 Supplementary References

1. Chen, D. L., Schonger, M. & Wickens, C. oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance* **9**, 88–97 (2016). URL <https://www.sciencedirect.com/science/article/pii/S2214635016000101>.
2. Morisi, D., Jost, J. T. & Singh, V. An Asymmetrical “President-in-Power” Effect. *American Political Science Review* **113**, 614–620 (2019). URL <https://www.cambridge.org/core/journals/american-political-science-review/article/an-asymmetrical-presidentinpower-effect/569413D40D79A79C3F7CA6F2183743B9>. Publisher: Cambridge University Press.
3. Druckman, J. N. & Levendusky, M. S. What Do We Measure When We Measure Affective Polarization? *Public Opinion Quarterly* **83**, 114–122 (2019). URL <https://doi.org/10.1093/poq/nfz003>.
4. Centola, D., Becker, J., Brackbill, D. & Baronchelli, A. Experimental evidence for tipping points in social convention. *Science* **360**, 1116–1119 (2018). URL <https://science.sciencemag.org/content/360/6393/1116>. Publisher: American Association for the Advancement of Science Section: Report.
5. Suri, S. & Watts, D. J. Cooperation and Contagion in Web-Based, Networked Public Goods Experiments. *PLOS ONE* **6**, e16836 (2011). URL <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0016836>. Publisher: Public Library of Science.
6. Mason, W. & Suri, S. Conducting behavioral research on Amazon’s Mechanical Turk. *Behavior Research Methods* **44**, 1–23 (2012). URL <https://doi.org/10.3758/s13428-011-0124-6>.

7. Arechar, A. A., Gächter, S. & Molleman, L. Conducting interactive experiments online. *Experimental Economics* **21**, 99–131 (2018). URL <https://doi.org/10.1007/s10683-017-9527-2>.
8. Peer, E., Brandimarte, L., Samat, S. & Acquisti, A. Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology* **70**, 153–163 (2017). URL <https://www.sciencedirect.com/science/article/pii/S0022103116303201>.
9. Palan, S. & Schitter, C. Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance* **17**, 22–27 (2018). URL <https://www.sciencedirect.com/science/article/pii/S2214635017300989>.
10. Iyengar, S., Sood, G. & Lelkes, Y. Affect, Not Ideology: A Social Identity Perspective on Polarization. *Public Opinion Quarterly* **76**, 405–431 (2012). URL <https://doi.org/10.1093/poq/nfs038>.
11. ANES. ANES: The American National Election Studies ([www.electionstudies.org](http://www.electionstudies.org)) (2016).
12. Berinsky, A. J., Huber, G. A. & Lenz, G. S. Evaluating Online Labor Markets for Experimental Research: Amazon.com’s Mechanical Turk. *Political Analysis* **20**, 351–368 (2012). URL [https://www.cambridge.org/core/product/identifier/S1047198700013875/type/journal\\_article](https://www.cambridge.org/core/product/identifier/S1047198700013875/type/journal_article).
13. Krupnikov, Y. & Levine, A. S. Cross-Sample Comparisons and External Validity. *Journal of Experimental Political Science* **1**, 59–80 (2014). URL [https://www.cambridge.org/core/product/identifier/S2052263014000074/type/journal\\_article](https://www.cambridge.org/core/product/identifier/S2052263014000074/type/journal_article).
14. Wood, D., Harms, P. D., Lowman, G. H. & DeSimone, J. A. Response Speed and Response Consistency as Mutually Validating Indicators of Data Quality in Online Samples. *So-*

*cial Psychological and Personality Science* **8**, 454–464 (2017). URL <http://journals.sagepub.com/doi/10.1177/1948550617703168>.

15. Chmielewski, M. & Kucker, S. C. An MTurk Crisis? Shifts in Data Quality and the Impact on Study Results. *Social Psychological and Personality Science* **11**, 464–473 (2020). URL <http://journals.sagepub.com/doi/10.1177/1948550619875149>.