

## Identity and the Self-Reinforcing Effects of Norm Compliance

Mark A. Pickup	Erik O. Kimbrough	Eline A. de Rooij
Department of Political Science	Smith Institute for Political	Department of Political Science
Simon Fraser University	Economy and Philosophy	Simon Fraser University
	Chapman University	
	Department of Economics	
	Simon Fraser University	

**Abstract:** When making social decisions (e.g. voting, donating, etc.), individuals consider the expectations of the groups with which they identify. These expectations are injunctive social norms, shared beliefs about what constitutes appropriate behavior for members of the identity group, and individuals' choices reflect trade-offs between adherence to these norms and their other preferences. Using an incentivized lab experiment, we show that when individuals pay a personal cost as a consequence of avoiding a group norm violation, they subsequently view the norm as stronger and become more willing to punish/reward others with the same identity for violating/complying with the norm than if they did not experience such a cost. Further, such individuals view other norms associated with the identity as stronger. In this way, costly compliance with a norm may have a reinforcing effect on identity based norms.

## Introduction

It is now a commonplace that economists should take into account the ways in which a person's *social identities* influences their behavior (see most prominently, Akerlof and Kranton 2001). Extensive experimental evidence supports the view that identity matters for behavior, showing that even induced minimal group identities, as well as natural identities, can impact generosity, cooperation, coordination, conflict, punishment and a host of other economic decisions (see e.g. Chen and Li 2009, as well as Chen and Mengel 2016 and the papers cited therein). An important channel through which social (group) identities shape behavior is via (injunctive) norms.<sup>1</sup> Members of an identity group have normative expectations about what other group members ought to do (as members of the group), and a desire to adhere to group norms leads individuals to adjust their behavior to be norm-consistent, sometimes incurring costs in order to follow group norms.<sup>2</sup>

Here we explore the possibility that there is a psychological mechanism that may make norm-following behavior *self-reinforcing*. With an incentivized experiment, we test a mechanism through which costly compliance with the norms of one's identity group increases the strength of the identity and its related norms. The mechanism we study is known as "effort justification" and emerges from the psychology literature on cognitive dissonance (Festinger, 1957). The idea is simple – when individuals weigh norm-compliance against self-interest, they will sometimes incur costs in the service of a group norm. Incurring such costs can induce a form of dissonance in which the individual becomes aware of the mutual incompatibility of their personal interests and the dictates of their identity group, yielding psychological discomfort. Aronson and Mills (1959) propose that such dissonance may sometimes be overcome via "effort justification", with individuals persuading themselves that in fact the norm following was even more valuable

---

<sup>1</sup> *Injunctive* norms, which describe shared beliefs about what someone *ought to do*, are contrasted to *descriptive* norms, which describe observed patterns of behavior (Cialdini and Trost 1998). We focus on the former, but to avoid jargon, we simply refer to these as norms from this point on.

<sup>2</sup> This account coheres with the rapidly growing literature on the role of norms in shaping social behavior, in which individuals are modeled as varying in the degree to which they care about following commonly known norms and trading off the desire to follow norms against own interests (e.g. Lopez-Perez 2008, Kessler and Leider 2012, Krupka and Weber 2013, Kimbrough and Vostroknutov 2016, Pickup et al. 2016). Here we simply highlight that norms may be defined at the level of the identity group (as argued by Akerlof and Kranton 2001), rather than globally, as is implicitly assumed in the literature.

than they initially believed. The implication of this mechanism is that those who incur a cost to comply with group norms will subsequently be more likely to incur related costs again.

Our experiment provides a clean test of this theory and generates evidence supporting this account. Our design allows subjects to endogenously choose whether to follow a group norm, even though they know it may be costly to do so, and then randomizes whether norm-following is actually costly for those who elect to comply. By randomly assigning the cost, conditional on the decision to adhere to the norm, we can identify a causal effect of incurring a cost. We show that when individuals (randomly) pay a personal cost as a consequence of norm following, they subsequently view the norm as stronger and become more willing to punish/reward others with the *same* identity for violating/complying with the norm than if they did not (randomly) experience such a cost. Further, such individuals view other norms associated with the identity as stronger. In this way, costly compliance with a norm may have a reinforcing effect on identity based norms.

### **Groups, Identities, Norms and Effort Justification**

A social group is defined as “two or more individuals who share a common social identification of themselves or [...] perceive themselves to be members of the same social category.” (Turner 1982: 15). A social identity is defined as “that part of an individual’s self-concept that arise from his knowledge of his membership in a social group (or groups) together with the value and emotional significance attached to the membership” (Tajfel 1978: 63). From this perspective, behaviors – one’s own or those of others in one’s group – that are not consistent with one’s identity are threatening to one’s self-concept. They can cause psychological distress.

As a consequence of the potential psychological cost of violating one’s social identity, individuals consider how members of the groups with which they identify are expected to behave. These expectations are norms, the rules of behavior that we expect members of our group to follow and believe others in our group will expect us to follow in kind (Bicchieri 2006; Akerlof and Kranton 2000, 2001). It has long been understood that social identities come with norms of behavior (Sherif 1967; Cooper *et al.* 2001; Spencer *et al.* 1998; Turner *et al.* 1987,

Akerlof and Kranton 2000). Norms are important, as they allow groups to coordinate and cooperate. However, norms have this power because they cannot easily be pushed aside when they are inconvenient for the individual, and following the norms of one's identity is not always fully compatible with satisfying one's other preferences. As Sherif and Sherif (1953, pg. 175) note, there is potential for conflict between group-based norms, and both idiosyncratic personal preferences and preferences based on the norms of other group identities. The references in footnote 1 are examples of how economists have begun to model norm-driven behavior; the key implication is that individuals may face tradeoffs between following norms on the one hand and their other interests on the other. If norms are sufficiently important, other interests may sometimes be sacrificed.

Our experiment focuses on the influence of political identities. Political identities (partisan, ideological, etc.) are an important kind of social identity (Green et al. 2002; Huddy *et al.* 2015; Achen and Bartels 2016; Malka and Lelkes 2010), and political identity groups have normative expectations like any other group. Thus norms help to shape political decisions such as donations and voting via the mechanism described above. For example, the identity "Republican" in the US is normatively associated with increased restrictions on abortion. Nevertheless, many individuals who identify as Republican may personally prefer restrictions not be increased. As another example, an individual may believe that as a "conservative" they should oppose funding for global sexual and reproductive rights programs but believe that as a woman they should support these rights for women across the globe. When such conflicts with the individual's political identity arise, voting behavior reflects this tension, with some individuals putting sufficient weight on group norms to vote against their other preferences. Individuals' vote choices thus reflect trade-offs between adherence to political norms on one hand and personal preferences and other group norms on the other (Pickup *et al.* 2016).

Under this view, when an individual incurs a cost on behalf of an identity (McConnell *et al.* 2017), this is the consequence of privileging the norm in the trade-off between norm compliance and other preferences. The consequence of incurring such a cost may be that the individual suffers from a form of cognitive dissonance (Festinger, 1957). The individual finds herself in a position of having to justify to herself why she chose to incur a cost. One way she

can do this is to engage in “effort justification”. Aronson and Mills (1959) suggested that there are two possible psychological processes by which an individual can reduce the cognitive dissonance associated with experiencing a cost due to a choice made. The first is to downplay the cost incurred. When this cannot be done, the individual increases the perceived value of the choice made in order to justify the cost. This is effort justification. It can explain why shared experiences of suffering or of struggling toward a common goal can induce strong feelings of group identity and has even been used to explain how hazing conducted by some Greek Letter Organizations, athletics teams, military organizations, etc. is effective at increasing loyalty to the group. The mechanism of effort justification predicts that experiencing a cost (e.g., in the form of effort and/or humiliation) in order to join or remain part of a group will increase the individual’s attachment to the group and its norms. Thus when individuals incur costs in order to comply with the norms of their political identity, effort justification predicts they will magnify the importance of the group and the associated norms in subsequent decisions. They persuade themselves that their decision truly was the right and necessary one in light of the importance of the group and in the process strengthen the norm. The result is that costly compliance with a norm can increase the strength of the group norm and increase (future) adherence to the norm itself. Group norm compliance is self-reinforcing.

Our research question then is: *What are the effects of experiencing a cost due to following a group norm?* Our hypotheses are that: 1) *complying with a group norm at a cost reinforces that group norm;* and 2) *complying with a group norm at a cost reinforces other norms of the group.*

### **Experimental Design**

Our experiment is designed to test whether subjects who incur costs in order to follow an identity based norm will be more likely than those who incur no cost to support and enforce that (and other group) norms. Our experiment uses subjects’ self-reported partisan and ideological identities. It also uses our foreknowledge of the norms regarding support for (opposition toward) the Affordable Care Act and restrictions on immigration that are associated with those identities. This allows us to construct choices in which subjects face a trade-off

between norm-following and their (expected) monetary earnings. Specifically, in our experiment a subject can either receive a certain payoff of \$6 for taking a norm-inconsistent action or receive an uncertain lottery that is equally likely to pay \$6 or \$2 for taking a norm-consistent action.

Our identification strategy relies on comparisons of subjects who took norm-consistent actions and differ only in whether they received a favourable lottery outcome and/or an unfavourable lottery outcome. Those whose lottery outcomes were unfavourable suffered a cost in order to follow the norm, and thus we would expect that they are more willing to punish norm-violators, reward norm-followers and report stronger normative expectations of others.

The experiment unfolds as follows (see Table 1 for a summary). Prior to the experiment, subjects completed a questionnaire in which they indicated their partisan (or ideological identity). These experiments were conducted in the US, so we know whether subjects identify as Democrats or Republicans in the Partisanship treatment and whether they identify as liberals or conservatives in the Ideology treatment. On the basis of these self-reports, we recruited subjects to the lab and assigned them to different tasks where they answered questions and made decisions related to either the Affordable Care Act (Obamacare treatment) or immigration policy (Immigration treatment). We chose these policies because we believe they are the focus of a salient divide between Democrats and Republicans (and between liberals and conservatives) in contemporary American politics.

Upon arrival to the lab, subjects confirm their self-reported identity. The experiment consists of 5 tasks, two of which are incentivized. Task 1 asks for subjects' beliefs about the extent to which other members of their own identity group would approve or disapprove if they chose to advocate for norm-inconsistent policies. We also ask how much the subject would approve or disapprove if they learned that others in their identity group advocated norm-inconsistent policies. This gives us a measure of subjects' beliefs about the norms of their identity group. For instance, in our Immigration-Partisanship treatment a self-identified Democrat would be asked how much they believe other Democrats would approve if they advocated in favour of reducing immigration. We anticipate that such advocacy will be perceived as a norm violation and thus

will result in disapproval. This task also serves to prime the individual’s political identity and the norm of the identity regarding a specific policy (a policy norm).

**Table 1: Structure of the Experiment**

<b>Pre-Experiment Questionnaire</b>	<b>Task 1: Policy Norm Elicitation</b>	<b>Task 2: Letter-writing Decision</b>	<b>Task 3: Vote Norm Elicitation</b>	<b>Task 4: Incentivized Policy Norm Elicitation</b>	<b>Task 5: Beliefs about Others’ Norm-Following</b>
Subjects report their partisan (ideological) identity and this is used to assign them to treatment. This occurs days before the experiment.	Subjects confirm their self-reported identity. Subjects report beliefs about other identity group members’ and own approval for norm-inconsistent behavior.	Subjects choose their expected payoff by deciding whether to send norm-consistent (\$2 or \$6 lottery) or norm-inconsistent letters (\$6). They must sign the letters, and address and stamp the envelopes.	Subjects learn the payoff of their decision (which is uncertain only if they chose norm-consistent letters). Then we ask them their beliefs about other identity group members’ and own approval of voting for a candidate of the opposite identity group.	Subjects receive an add’l \$6. Using the strategy method, subjects indicate their willingness to punish/reward norm-consistent and norm-inconsistent letter writers from both their own and the other group (up to \$1.5, each \$1 translates to reward/punishment of \$3). They are then matched with one other subject in the experiment.	Subjects report how likely others of the same group were to write norm-consistent and norm-inconsistent letters. Outcomes are revealed. Subjects are paid and dismissed.

Task 2 provides subjects with an incentivized choice between a norm-inconsistent and a norm-consistent action. Specifically, subjects are told that they will have to sign, address, stamp and fill ten envelopes with advocacy letters, which will be mailed to ten different members of the US Congress after the experiment. These letters either advocate for or against a policy (the Affordable Care Act or restrictions on immigration). Subjects are offered \$6 for sure if they choose to send letters that are norm-inconsistent. Subjects are told they will receive a lottery payment that yields \$2 or \$6 with equal probability if they choose to send letters that are norm-consistent. Subjects do not learn the outcome of the lottery until they have made their decision and completed the task. Keeping with the example above, this means that a self-

identified Democrat in the Immigration-Partisanship treatment is offered the choice between sending letters advocating for increasing restrictions on immigration and receiving \$6 for sure or else sending letters advocating for reducing restrictions on immigration and receiving the lottery payment. After making their choices, subjects were given time to complete the task (typically 15 minutes). Once the monitor has determined that the task is complete and the letters have been correctly addressed and stamped, the subjects are given a password that allows them to go on to Task 3.<sup>3</sup>

At the outset of Task 3, subjects are told their payoff based on their choice (and when applicable the outcome of the lottery) in Task 2. Task 3 then asks subjects another question intended to measure the strength of identity-driven norms. Specifically, subjects are asked how much they believe other members of their identity group would approve or disapprove if they voted for a candidate from the other party (or ideology) in the next federal election. We also ask subjects how much they would approve or disapprove if they learned that another member of their identity group voted for the opposite group. Again, following the example, this means that a Democrat would be asked how much they believe other Democrats would approve or disapprove if they voted for a Republican in the next federal election. Then they would be asked how much they would approve if they learned that another Democrat voted for a Republican in the next federal election. These questions measure an identity related norm – specifically, the norm against voting for the other side (a vote norm).

Task 4 provides an incentivized measure of norm-strength. This is the norm regarding the policy they were asked to support or oppose in their letters. Specifically, we once again remind subjects of their decision and earnings from Task 2, and then we use the strategy method to elicit willingness to punish/reward others who chose to send norm-consistent or norm-inconsistent letters in Task 2. All subjects are told that they have been allocated an additional \$6 at the start of Task 4. They are then told that they can spend money to punish and reward others. Specifically, for each \$0.01 that they spend on punishment they can reduce the payoff

---

<sup>3</sup> All but one participant completed the task. One subject asked to have their data withdrawn after making their choice in Task 2 because, to paraphrase, they didn't realize that we would actually have them send letters to Congress.

of another subject by \$0.03, or for each \$0.01 that they spend on reward, they can increase the payoff of another subject by \$0.03. Subjects are told that they will be matched with one other person in the room at the end of the experiment at random. This means that their punishment/reward decision will impact that person's payoff and that person's punishment/reward decision will impact their own payoff. They are asked how much they would like to punish or reward that individual depending on that individual's choice to send norm-consistent or -inconsistent letters.

In order to avoid deceiving subjects while also collecting data on our primary quantities of interest (punishment/reward of norm-inconsistent members of their own group), we employ the strategy method to get punishment/reward decisions for all possible subject types with whom they might be matched. That is, we ask subjects to make punish/reward decisions for each of the four possible cases:

1. a Democrat (or liberal) who sent norm-consistent letters (i.e. pro-ACA or pro-immigration);
2. a Democrat (or liberal) who sent norm-inconsistent letters (i.e. anti-ACA or anti-immigration);
3. a Republican (or conservative) who sent norm-consistent letters (i.e. anti-ACA or anti-immigration); and
4. a Republican (or conservative) who sent norm-inconsistent letters (i.e. pro-ACA or pro-immigration).

Participants can spend a maximum of \$1.50 since this ensures that their worst-case earnings from this Task are \$0 (i.e. if they spend \$1.50 punishing or rewarding others and are punished maximally for \$4.50). With this task we seek to measure whether subjects who have incurred a cost by following the norm, i.e. have sent norm-consistent letters and received the lower lottery payoff of \$2, are more likely to punish members of their group who were norm-inconsistent and reward members of their group who were norm-consistent. The comparison group is those that also selected the norm-consistent letters but did not experience a cost.

Task 5 asks subjects how likely it is that a member of their own group chose to send each kind of letter. This allows us to address the possible concern that the relevant punishment/reward decisions were made with the expectation that they would not actually have consequences. In other words, if a subject believes that others in their group are extremely unlikely to send norm-inconsistent letters, then their punishment/reward decision might not be informative.

After Task 5, subjects are matched at random with another subject in the room, their final payoffs are calculated and displayed on their screen, and then they complete a short demographic questionnaire and are called over one-by-one for payment.

In total 206 subjects participated in 10 total experimental sessions, containing between 10 and 30 subjects each, depending on sign-up and show-up rates. Overall, we found it easier to recruit Democrats and liberals than to recruit Republicans and conservatives, and as a consequence less than 40% of the sample consists of members of the latter two identity groups. Of the 206 subjects, 136 were liberals or Democrats and 70 were conservatives or Republicans (see Table 2). During the experimental sessions, 87 of the liberals/Democrats had their ideological identity primed and 49 had their partisan identity primed. Meanwhile, 31 of the conservative/Republicans had their ideological identity primed and 39 had their partisan identity primed. Sessions lasted on average 45 minutes, and subjects were paid an average of 17.90 dollars, including a \$7 payment for arriving on time. Instructions were delivered on-screen via zTree and were self-paced, since the content was personalized by the subject’s self-reported partisan or ideological identity. A monitor answered any questions that arose privately. Due to a software error, we were unable to collect demographic questionnaire data on the 23 subjects in our first session. We collected no additional unreported data and no pilot sessions.

**Table 2: Identity Distribution of Participants**

<b>Treatment</b>	<b>Participant Identity</b>	
	liberal/Democrat	conservative/Republican
Ideology identity	87	31
Partisan identity	49	39
Total	136	70

## Results

Our theory applies to individuals that experience a cost because they chose to comply with the norms of their identity. This requires individuals to perceive such norms. *Do individuals perceive group norms?* The first thing we explore is whether individuals do in fact perceive a group norm for the issue on which they will be asked to experience a cost. This norm is measured as: 1) the (dis)approval that individuals expect others with the same identity to express towards them if they chose to violate the norm; and 2) the (dis)approval that individuals would express to others with the same identity that chose to violate the norm.

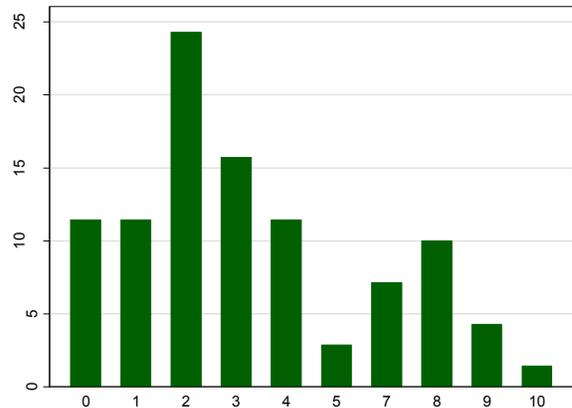
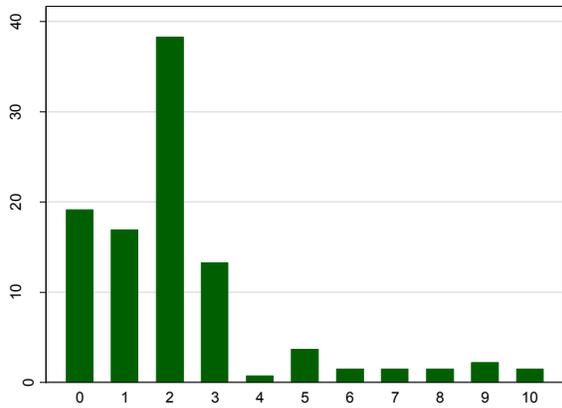
Figure 1 gives the distribution of expected approval (from 0 to 10) by others for violating the norm. We see that for liberals/Democrats that the vast majority (88.2%) would expect disapproval. For conservatives/Republicans, again a majority would expect disapproval but there is a significant minority (22.9%) that would not. Figure 2 gives the distribution of approval of others for violating the norm. Again, the vast majority of liberals/Democrats (77.21) would disapprove. Among conservatives/Republicans those that would approve are roughly equal with those that would disapprove, although the magnitude of disapproval is stronger.

It is clear that liberals and Democrats perceive a norm against supporting the issues for which they are asked to experience a cost (ACA and immigration). There is some but weak evidence that conservatives and Republicans also perceive a norm.

**Figure 1: (Dis)approval by others for violating the norm**

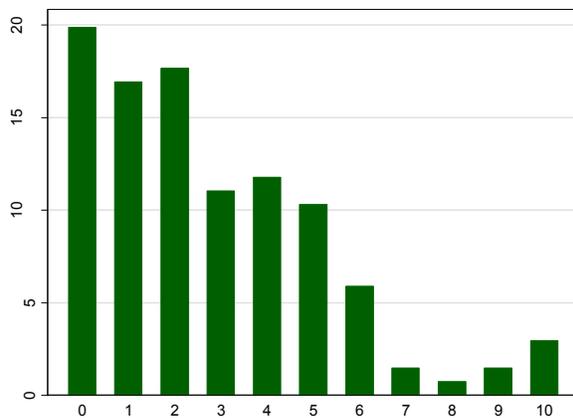
Other Liberals/Democrats Approval of  
You Advocating Norm Inconsistent Position

Other Conservative/Republicans Approval of  
You Advocating Norm Inconsistent Position

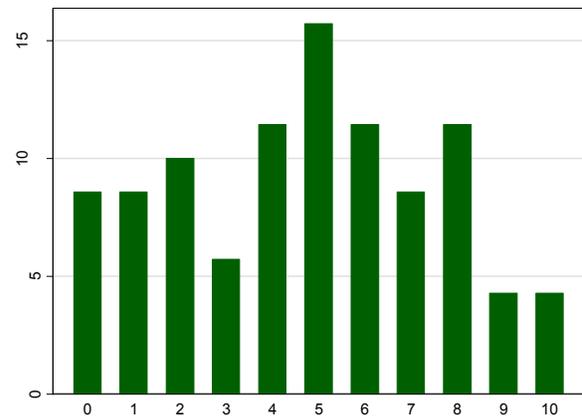


**Figure 2: (Dis)approval of others for violating the norm**

Your Approval of Other Liberals/Democrats  
Advocating Norm Inconsistent Position



Your Approval of Other  
Conservative/Republicans Advocating Norm  
Inconsistent Position



Again, our theory applies to individuals that experience a cost due to group norm compliance. *How many individuals followed the norm despite the potential cost of doing so?* Overall, 76.5 percent of liberals/Democrats (74.7/79.6) chose the norm consistent task, while only 37.14 percent of conservatives/Republicans (32.3/41.0) did the same. And, of those that perceived a norm (gave a score less than five on both of the two norm questions), 87.21 percent of liberals/Democrats chose the norm consistent task, while 64.71 percent of conservatives/Republicans did the same. This means that of the original 136

liberals/Democrats, 85 (63%) perceived a norm and followed it at a cost. Meanwhile, of the original 70 conservative/Republicans, 14 (20%) perceived a norm and followed it at a cost.

It is not clear if we have just chosen the wrong issues for conservatives/Republicans or conservatives/Republicans are less likely to perceive and follow group norms (at a cost) or if the individuals that identified as conservative/Republican in our sample are not typical of the group with which they identify. Whatever the explanation, we do not have enough conservatives and Republicans perceiving and choosing the norm consistent task to test our theory.<sup>4</sup> We do, however have plenty of liberals/Democrats that both perceive the norm and choose to experience a cost in order to comply with the norm (63%).

We are now prepared to ask: *What are the effects of experiencing a cost due to choosing the norm consistent option?* We are interested in the effect of experiencing a cost due to choosing a norm and so we compare those that chose the norm-consistent task and experienced a cost to those that chose the norm-consistent task and did not experience a cost. We would also like to restrict our sample to those that perceived a norm and we would like to see how the effect varies by the strength of this norm. We begin by looking at the effect on how much individuals are willing (at a cost to themselves) to reward others with the same identity for complying with group norms (i.e., choose the norm-consistent task). Figure 3 plots the effect at varying strengths of the norm. The norm strengths are based on those that gave a score <5 (weak), <3 (moderate), and <1 (strong) on both the expected approval by others and the approval of others for violating the policy norm from Task 1.<sup>5</sup> The groups are mutually exclusive, so that if an individual qualified for a norm strength category they are excluded from all weaker norm strength categories.

We see that those that experienced the cost are more likely to reward liberals/Democrats like themselves for complying with the norm. The effect magnitude increases with the strength of

---

<sup>4</sup> In a related study conducted after this one we have used an incentivized task to measure normative beliefs of a representative sample of US conservatives as well as the extent to which norms associated with conservative identity vary across the age distribution (de Rooij, Kimbrough and Pickup 2018). We find that although the chosen issues are normatively important for conservatives in general, this is less true among millennials, who make up the bulk of the subjects in this study.

<sup>5</sup> We estimate effects varying by policy norm strength with the following linear regression:  $\$Amount_i = \beta_0 + \beta_1 Cost_i + \beta_2 Norm\_Strength_i + \beta_3 Cost_i \times Norm\_Strength_i + \varepsilon_i$

the perceived policy norm. Using a 0.05 one-tailed test, these effects are significant in two of the three levels of norm strength. When the strength of the policy norm is moderate, the magnitude of the effect is \$0.47 and when the strength of the policy norm is strong, the magnitude of the effect is \$1.12. This increase in costly rewarding of norm consistent behavior is 25 percent of the maximum possible reward.

**Figure 3. Rewarding of a Norm Consistent Democrat/Liberal**

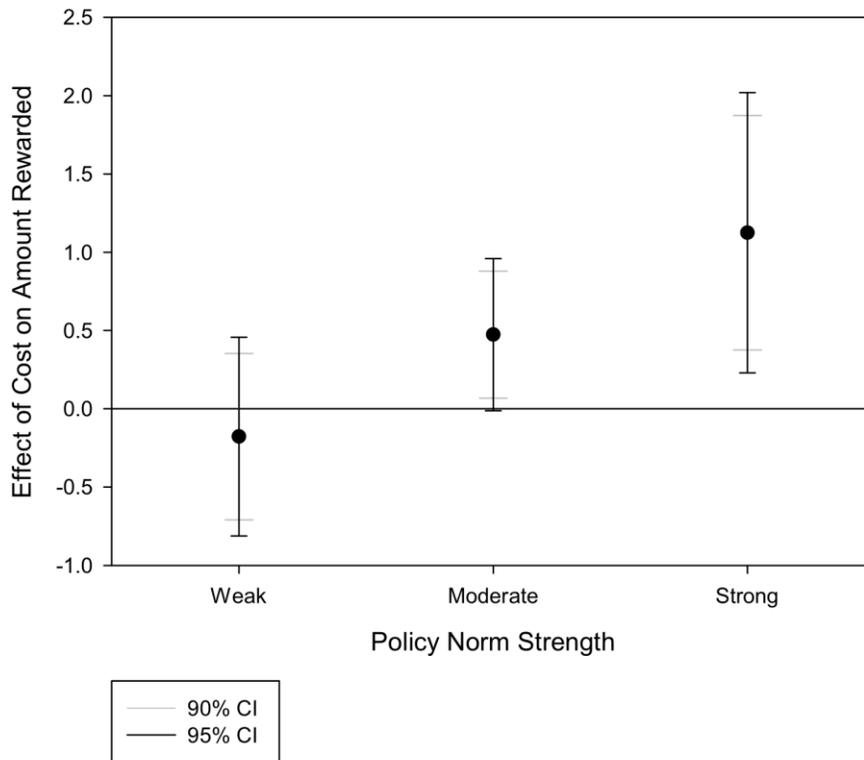
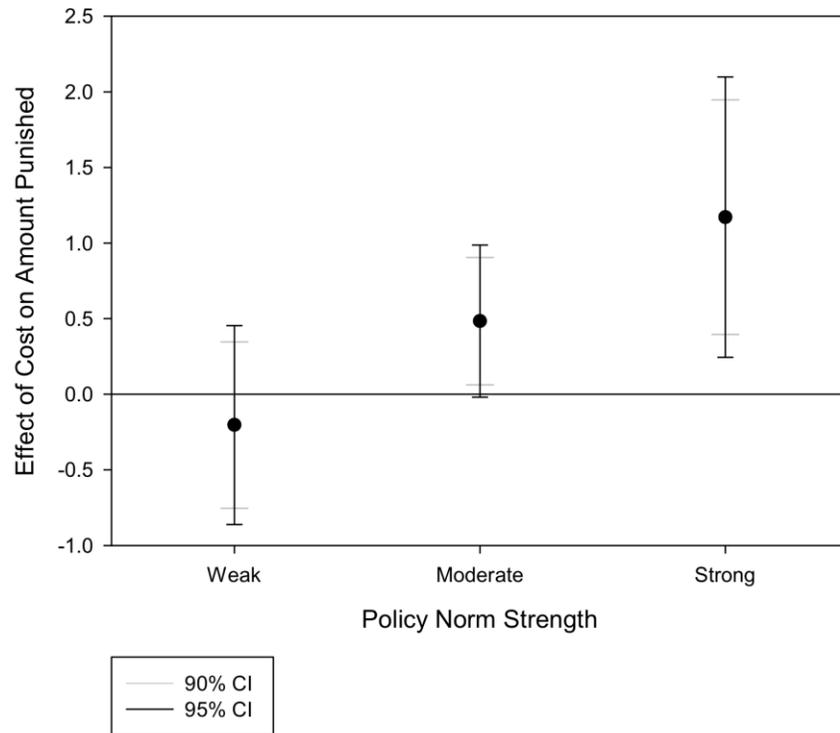


Figure 4 plots how much individuals are willing (at a cost to themselves) to punish others with the same identity for violating group norms (i.e., choose the norm-inconsistent task) at varying strengths of the policy norm. We see that those that experienced the cost are more likely to punish Liberals/Democrats like themselves for violating the norm. Again, the magnitude of the effect increases with the strength of the perceived norm. The effect is significant at two of the three norm strengths. When the strength of the policy norm is moderate, the magnitude of the effect is \$0.48 and when the strength of the policy norm is strong, the magnitude of the effect is \$1.17 (26 percent of the maximum possible reward).

**Figure 4. Punishment of a Norm Inconsistent Democrat/Liberal**

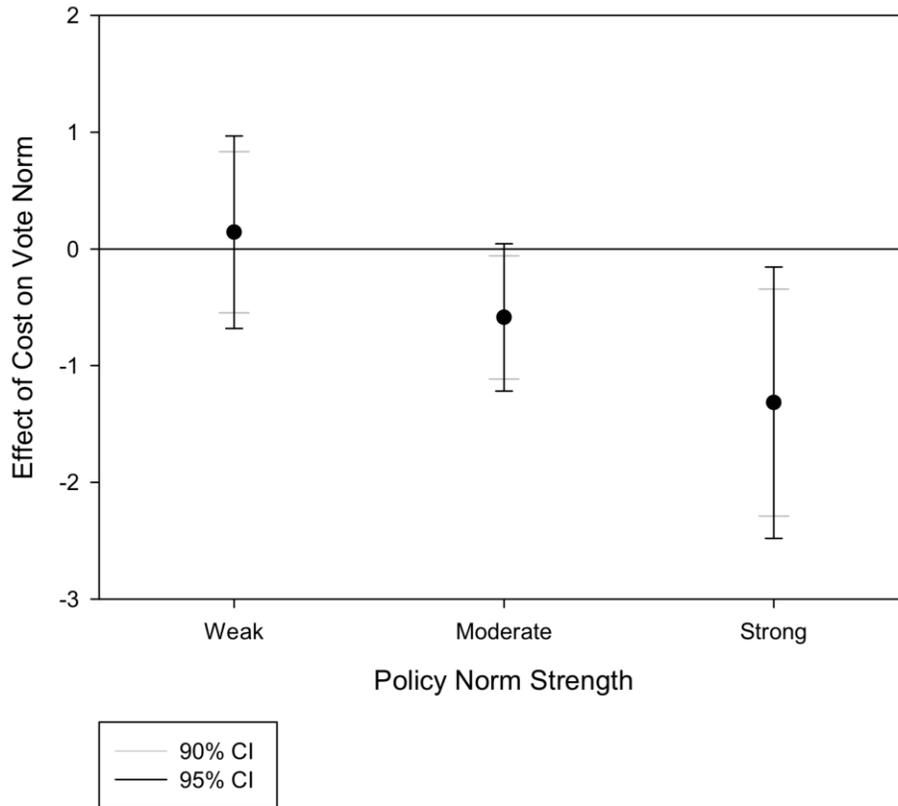


Based on these results, there is evidence that *complying with a group norm at a cost reinforces that group norm (H1)*. Given that the individual's choosing to pay a cost to reward norm compliance and punish norm violations are the same individuals that already experienced a cost due to their own choice to comply with the group norm, the evidence is compelling.

Next, we look at the effect of expressing a cost due to complying with a norm on the perceived strength of a different norm. This is the norm: do not vote for the other side. We call this the vote norm and it is measured as the expected approval of others (on a 0 to 10 scale) for voting for a candidate of the opposite identity (a conservative candidate if the individual identifies as liberal and a Republican candidate if the individual identified as Democrat). Figure 5 plots the effect of experiencing a cost on the perception of this vote norm (measured in Task 3, as beliefs about others' approval) at varying levels of strength of the original policy norm regarding the ACA/immigration (Task 1). The magnitude of the effect again varies by the strength of the policy norm and is statistically significant at two of the three levels. When the strength of the policy norm is moderate, experiencing a cost decreases the expected approval of others for voting for

a candidate of the opposite identity by 0.58 (on a 0 to 10 scale). When the strength of the policy norm is strong, the magnitude of the effect is 1.59. A similar, but less pronounced effect is observed for first-order normative beliefs (see Appendix A).

**Figure 5. Effect of Cost on Norm against Voting for ID Inconsistent Candidate (Beliefs about Others' Approval)**



There is moderate evidence that *individuals that experience a cost because they chose to comply with a norm of their political identity are more likely to perceive other group norms more strongly (H2)*. Overall, the experiment provides evidence that experiencing a known cost as a result of complying with the norms of one's political identity increases the strength of that identity and the norms of that identity. In fact, the cost increases the individual's willingness to experience an additional cost in order to punish and reward others in their group for violating and complying with group norms.

As a robustness check, we asked respondents how likely it was that someone with the same identity would choose the norm consistent and inconsistent options. This is done to check that individuals saw the choices regarding punishing and rewarding to be meaningful. For example, if the individual indicated they would pay \$0.50 to punish a norm-inconsistent liberal/Democrat, they did believe there was a possibility they would have to pay this amount. On choosing the norm consistent option, liberals and Democrats gave a score of 7.9 on a 1 to 10 scale. On choosing the norm inconsistent option, liberals and Democrats gave a score of 3.3. The norm-consistent option was seen as more likely than the norm-inconsistent option. However, only 2.9 percent (3) individuals felt that there was no chance of another liberal/Democrat choosing the norm-inconsistent option.

As a second robustness check, we tested if experiencing a cost due to choosing the norm-consistent option had an effect on liberals/Democrats punishing or rewarding conservatives/Republicans. While we expect liberals/Democrats to punish or reward conservatives/Republicans for their choice, our theory does not predict any effects from experiencing a cost on these choices. We found no evidence of such effects.<sup>6</sup> Note that if punishment/reward decisions were driven by generalized inequality aversion (or some preference defined over payoff distributions), we would not expect differential treatment of in- and out-group members.<sup>7</sup>

As a sanity check, we looked at how perceptions of the policy norm predicted whether an individual would choose the norm-consistent or inconsistent letter option. Among liberals/Democrats, those that would expect disapproval by others (score <5) for violating the norm were 23 percent more likely to choose the norm-consistent option. Those liberals/Democrats that indicated they would disapprove of others for violating the norm were 36 percent more likely to choose the norm-consistent option. In other words, perceiving the

---

<sup>6</sup> At the 0.05 (one-tailed) level, the only significant effect was that liberals/Democrats with the weakest perception of a policy norm were more likely to reward conservatives/Republicans for being consistent with their own group norms, as a result of experiencing a cost.

<sup>7</sup> Models in which shared identity induces differential treatment of in- and out-group members via a change in preferences (e.g. Chen and Li 2009) can potentially account for some of our observations. However, we know of no model that predicts all of the following: 1) in-group (but not out-group) inequality aversion is triggered by norm-breaking; 2) in-group (but not out-group) altruism is triggered by norm-following; and 3) these effects are triggered by having incurred a cost for following the norm (and not just having followed the norm).

norm increases the probability of selecting the norm consistent option. Interestingly among conservatives/Republicans the corresponding numbers are 2.4 percent less likely (not statistically significant) and 32 percent more likely. It would appear that amongst this group, expected (dis)approval by others has little effect on choice. This underlines the problem we have with our conservative/ Republican sample.

### **Conclusion**

Social identities shape behavior in part because of the normative expectations shared by group members, and to the extent that individuals care about satisfying these expectations (and following group norms) they may sometimes incur costs as a result of doing so. We provide experimental evidence supporting the idea of “effort justification” as a mechanism by which such costly norm-following can be self-reinforcing. By randomly assigning the cost, conditional on the decision to follow the norm, we are able to identify a clear causal effect of costly norm following on subsequent punishment/reward decisions and normative beliefs. When individuals incur costs to follow a norm associated with an identity group, they subsequently become more willing to punish/reward other members of their own identity group. This happens despite the fact that, if there are income effects in the decision to punish/reward, we would expect those to work in the opposite direction of the treatment effect. Thus we provide evidence for an important channel by which identity groups strengthen and persist over time: by encouraging group members to follow norms (at personal cost), and thereby reinforcing the underlying normative commitments.

## References

- Aronson, E. and Mills, J. 1959. The effect of severity of initiation on liking for a group. *The Journal of Abnormal and Social Psychology*, 59(2):177.
- Achen, C.H. and L.M. Bartels. 2016. *Democracy for Realists: Why Elections Do Not Produce Responsive Government*. Princeton, NJ: Princeton University Press.
- Akerlof, G.A. and R.E. Kranton. 2000. "Economics and Identity." *Quarterly Journal of Economics* 115:715-53.
- Akerlof, G.A. and R.E. Kranton. 2001. "Identity and the Economics of Organizations." *Journal of Economic Perspectives* 19:9-32.
- Aronson, Elliot and Judson Mills. 1959. "The Effect of Severity of Initiation on Liking for a Group." *Journal of Abnormal and Social Psychology* 59:177-181.
- Bicchieri, C. 2006. *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge, UK: Cambridge University Press.
- Chen, Yan and Friederike Mengel. 2016. "Social Identity and Discrimination: Introduction to the Special Issue." *European Economic Review*, 90:1-3.
- Chen, Yan and Sherry Li. "Group Identity and Social Preference." *American Economic Review*, 99(1):431-457.
- Cialdini, Robert B. and Melanie R. Trost (1998). "Social Influence: Social Norms, Conformity, and Compliance." In *The Handbook of Social Psychology*. edited by D. T. Gilbert, S. T. Fiske, and G. Lindzey. McGraw-Hill, Vol. 1–2:151–192.

- Cooper, J., K.A. Kelly and K. Weaver. 2001. "Attitudes, Norms, and Social Groups." In *Blackwell Handbook of Social Psychology: Group Processes*. Editors, M.A. Hogg and S. Tyndale. Oxford: Blackwell Publishers Inc.
- de Rooij, E., E.O. Kimbrough, M.A. Pickup. 2018. "Mapping Political Norms in the US Electorate." *Working Paper*.
- Downs, A. 1957. *An Economic Theory of Democracy*. New York: Harper.
- Festinger, L. 1957. *A Theory of Cognitive Dissonance*. Stanford, CA: Stanford University Press.
- Green, D.P., B. Palmquist and E. Schickler (2002). *Partisan Hearts and Minds*. New Haven: Yale University Press.
- Huddy, L., L. Mason and L. Aarøe. 2015. "Expressive Partisanship: Campaign Involvement, Political Emotion, and Partisan Identity." *American Political Science Review* 109(1):1-17.
- Kessler, Judd B. and Steven Leider (2012). "Norms and Contracting." *Management Science*, 58:62–77.
- Kimbrough, Erik O. and Alexander Vostroknutov. "Norms Make Preferences Social." *Journal of the European Economic Association*, 14(3):608-638.
- Krupka, Erin L. and Roberto A. Weber (2013). "Identifying Social Norms using Coordination Games: Why Does Dictator Game Sharing Vary?" *Journal of the European Economic Association*, 11:495–524.
- Lau, R.R. and D.P. Redlawsk. 2006. *How Voters Decide: Information Processing in Election Campaigns*. Cambridge University Press.
- Lopez-Perez, Raul. 2008. "Aversion to Norm-Breaking: A Model." *Games and Economic Behavior*, 64:237-267.

- Malka, A. and Y. Lelkess. 2010. "More than Ideology: Conservative–Liberal Identity and Receptivity to Political Cues." *Social Justice Research* 23:156–88.
- McConnell, Christopher, Yotam Margalit, Neil Malhotra, and Matthew Levendusky. 2017. "The Economic Consequences of Partisanship in a Polarized Era." *American Journal of Political Science*. In press.
- Pickup, Mark A., Erik O. Kimbrough and Eline A. de Rooij. 2016. "Group Identity and Voting: Norms and Personal Interest." Presented at the 2016 Midwest Political Science Association meeting (April 2016) and Toronto Behaviour Workshop (November 2016).
- Popkin, S.L. 1991 *The Reasoning Voter: Communication and Persuasion in Presidential Campaigns*. Chicago: University Of Chicago Press.
- Redlawsk, D. 2002. "Hot Cognition or Cool Consideration." *The Journal of Politics* 64(4):1021–1044.
- Sherif, M. 1967. *Group conflict and cooperation: Their Social Psychology*. London: Routledge and Kegan Paul Ltd.
- Sherif, M. and Sherif, C.W. 1953. *Groups in harmony and tension; an integration of studies of intergroup relations*. Harper & Brothers.
- Spencer, S.J., S. Fein, C.T. Wolfe, C.F. and M.A. Duinn. 1998. "Automatic Activation of Stereotypes: The Role of Self-Image Threat." *Personality and Social Psychology Bulletin* 24 (11):1139-1152.
- Suhay, E. 2015. "Explaining Group Influence: The Role of Identity and Emotion in Political Conformity and Polarization." *Political Behavior* 37:221-251.

Taber, C.S. and M. Lodge. 2006. "Motivated Skepticism in the Evaluation of Political Beliefs."

*American Journal of Political Science* 50(3):755–769.

Tajfel, H. 1978. "Social Categorization, Social Identity, and Social Comparisons." In Henri Tajfel,

ed., *Differentiation Between Social Groups*. London: Academic Press.

Turner, John C. 1982. "Towards a cognitive redefinition of the social group." *Social identity and*

*intergroup relations* : 15-40.

Turner, J.C., M.A. Hogg, P.J. Oakes, S.D. Reicher and M.S. Wetherell. 1987. *Rediscovering the*

*Social Group: A Self-Categorization Theory*. Cambridge, MA: Basil Blackwell.

Appendix A.

Figure A1. Effect of Cost on Norm against Voting for ID Inconsistent Candidate (Own Approval)

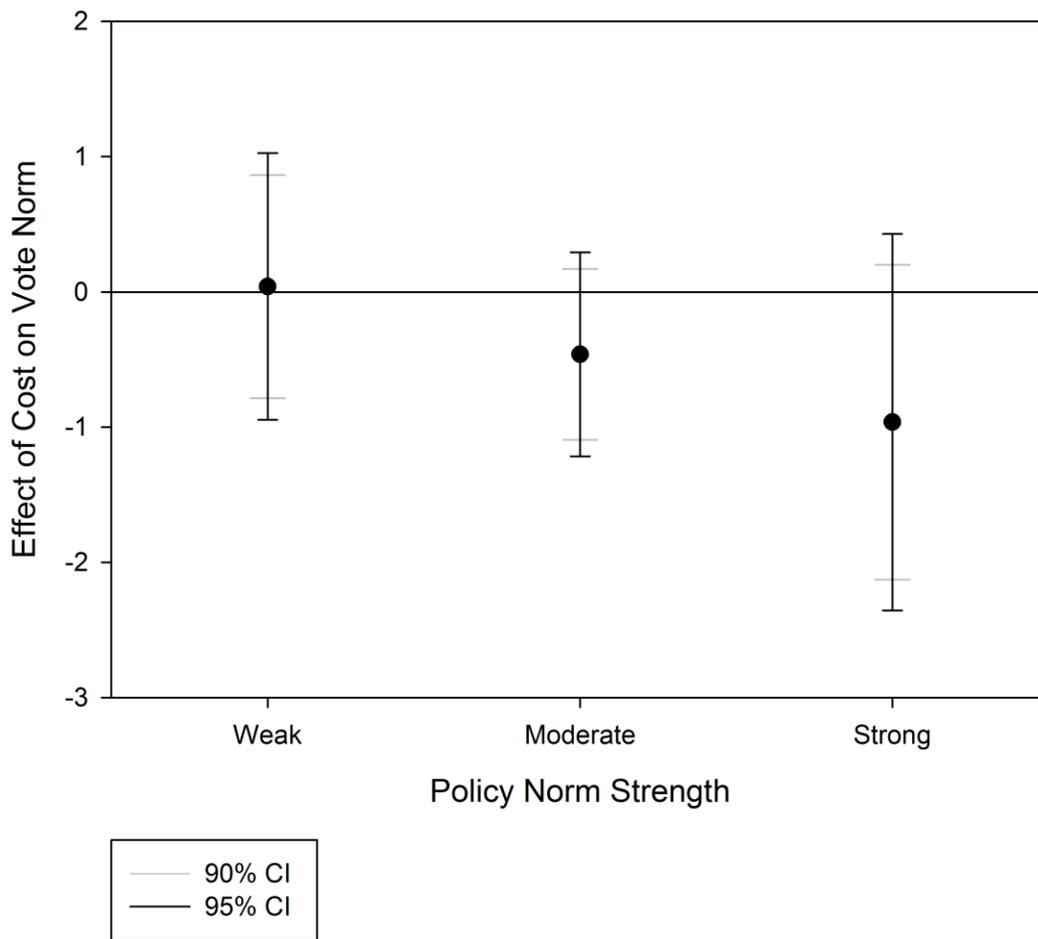


Figure A1 reports the effect of incurring the cost on own normative beliefs about the appropriateness of voting for a candidate from the other party/ideology. As can be seen in the figure, the effect of the cost is not statistically significant on own normative beliefs, but the directional effect matches that for second order normative beliefs reported in Figure 5.