Burning Bridges to Build Bridges: Making yourself disliked by the outgroup can make you seem more trustworthy to the ingroup, and deepen group divides

Scott Danielson, Kumar Yogeeswaran, Andrew Vonasch
Department of Psychology, Speech and Hearing
University of Canterbury

Author Note

Scott W. Danielson https://orcid.org/0000-0002-0637-9723
Correspondence concerning this article should be addressed to Scott Danielson, University of Canterbury, School of Psychology, Speech and Hearing, 20 Kirkwood Avenue, Upper Riccarton, Christchurch, New Zealand 8041
Email: scott.danielson@pg.canterbury.ac.nz

Burning Bridges to Build Bridges: Making yourself disliked by the outgroup can make you seem more trustworthy to the ingroup, and deepen group divides

> When your army has crossed the border [into enemy territory], you should burn your boats and bridges, in order to make it clear to everybody that you have no hankering after home.
>
> —*Sun Tzu, The Art of War*

Political polarization is on the rise not only in the United states (Pew Research Center, 2022) but also globally (Gidron, Adams, & Horne, 2019). This is a problem for democracy when it causes strong centralization of power (Lee, 2015) and makes people less satisfied with the democratic process (Wagner, 2021). Moral polarization also increases perceptions of the breakdown of society and as a result increases support for strong authoritarian leaders (Crimston, Selvanathan, & Jetten, 2022). Compounding these problems is when fringe beliefs like conspiracy theories emerge along political lines. While conspiracy belief exist on both sides of the political aisle, they are strongest at the extremes (Van Prooijen, Krouwel, & Pollet, 2015; Hopp, & Vargo, 2019) and the rise of extreme politicians has brought conspiracy belief into ordinary public discourse, such as with the covid-19 vaccine (e.g. Imhoff & Lamberty, 2020).

A result of political polarization is that partisans of each side do not communicate, and there emerges mutual distrust of either side (McCoy et al., 2018). In fact, dislike of the other side is on the rise (Abramowitz & Webster, 2016), and drives partisanship more reliably than liking one's own side. This is even true world-wide. A study of 25 European nations found that political division is driven less by in group favoritism than by out-group animosity (Hahm et al., 2022). In a climate where political sides are in bitter rivalry, status within one's

own ideological group becomes important, and what the other side thinks of you becomes less important. In this kind of environment, doing or saying things that are distasteful or offensive to those who do not share your political views ceases to have any kind of penalty to one's social status. Being offensive to other partisans may even enhance your social status in your own group because it marks you as a committed member of your own side (Mercier, 2020).

While there are many factors leading to polarization and distrust, this project focuses on one that has been proposed in theory, but to our knowledge not yet tested experimentally: *burning bridges* (Mercier, 2020). This is a method, done consciously or unconsciously, where a person aims to signal commitment to a group by disparaging a rival group, i.e. burning their bridges to that group. We see examples of this on both the left and right, with "cancel culture" and "owning the libs" being ways of ruining the potential for relationships to develop across the political spectrum. Burning bridges may be particularly effective in a polarized climate, where group loyalty is important and outgroup competitors are particularly reviled.

We propose *burning bridges* is a specific example of a *commitment device*, a costly arrangement entered into by an individual with the aim of helping fulfil a difficult goal (Bryan et al., 2010). Examples of commitment devices include anything where an individual self-imposes a cost for a future action, as a way to help achieve a goal. Entering into a savings plan in which interest is forfeited if a monthly deposit is not made, painting one's nails with Control-It Nail Biting Treatment, or publicly declaring a New Year's resolution are all commitment devices. Some individuals opt into savings plans that make withdrawing their own money difficult or impossible at certain times (Ambec and Treich, 2007). A program can help smokers quit by allowing them to make a deposit that is returned to them in 6 months if they quit, but donated to charity if they don't (Giné et al., 2010). Commitment devices aim to

enhance self-control by imposing a cost for failure; in this way they *commit* and individual to their chosen course of action by making failure more punishing. Burning bridges is a specific kind of commitment device in which one intentionally sacrifices status with one social group, e.g. rival partisans, with the goal of improving status within their own group. Burning bridges may be an effective and rational strategy in a highly polarized social climate, but it functions to prevent collaboration across the political aisle.

Prior work has shown how commitment devices are used by individuals to influence how they are perceived by others. Researchers have argued that emotions like shame may have evolved as a commitment device to show deference to a group norm (De Hooge et al., 2008). Kurzban and Christner (2011) argue that supernatural beliefs, often in the context of religion, act as commitment devices. The authors argue that because supernatural beliefs are counterfactual and often bizarre they reliably signal group membership with those that shares those beliefs. There would be no practical benefit to professing belief in invisible or impossible things other than to show you are a member of the group who believes those things. A related concept is *costly signaling*: doing behaviors that would be costly to the self if one weren't strongly committed to the group (Hall et al., 2015; McAndrew, 2021). Examples include undergoing difficult or lengthy initiations to join a religion or being hazed to join a fraternity. The new initiate signals loyalty to the religion or the frat by being willing to undergo such a difficult process for group membership. Additionally, strongly and openly affiliating with one religion or group can be costly in that it reduces your other options: it makes you unappealing to other groups that do not share the same beliefs, or are otherwise competitors. In the same way, we argue burning bridges serves as a costly signal to group membership. Our approach is novel in that "burning bridges" specifically sacrifices *inter*group status to enhance *intra*group status. Although burning bridges may be harmful to

democracy and society more broadly, we propose that people do it because it can provide reputational benefits within a narrower group.

We propose to test the idea that burning bridges is a costly signal that can enhance one's reputation in the ingroup by intentionally ruining one's reputation in other groups (Mercier, 2020). One way to convincingly signal commitment to a group is by making yourself so unappealing to rival groups that you have no choice but to be loyal to your target group. This creates an incentive structure where joining or remaining in that group is now the only good option. Thus, people observing someone burning bridges behaviors should believe and trust their loyalty to that group—regardless of whether the observer loves or despises that group—because the bridge burner's signal of commitment is credible.

We propose that burning brides signals two kinds of commitment. It signals *internal* commitment, which we define as strong motivation to accomplish the goal or course of action: it makes the bridge burner's intentions clear to everyone. Second, it signals *external* commitment, which we define as an imposed incentive structure: it rallies external forces to bind the bridge burner to the action. Deciding to marry someone is making an internal commitment, but actually going through with it creates an external commitment, i.e. legal ramifications for breaking that commitment. Burning bridges to a rival political party, for instance by tweeting insulting things at members of that party, signals both and internal and external commitment: it shows you are determined not to join or work with their party but also makes joining or working with their party very hard, because members will remember what you said and be unlikely to accept you. It is an open question how much these two types of commitment influence perceptions of bridge burners: either or both may drive perceptions of commitment.

To test this model we present five empirical studies: two already conducted which support the model and we propose three additional ones. Study 1 empirically tests simple

commitment devices in an online sample and shows they reliably signal commitment, Study 2a extends this finding to include intergroup dynamics and our definition of *burning bridges*. Proposed Study 2b would add the dimension of perspective taking to study 2a, i.e. test perceptions of commitment when the judge either supports or opposes the group to which a person is burning bridges. Study 3 would test the burning bridges account in the domain of partisan politics and study 5 in the domain of fringe extremist groups.

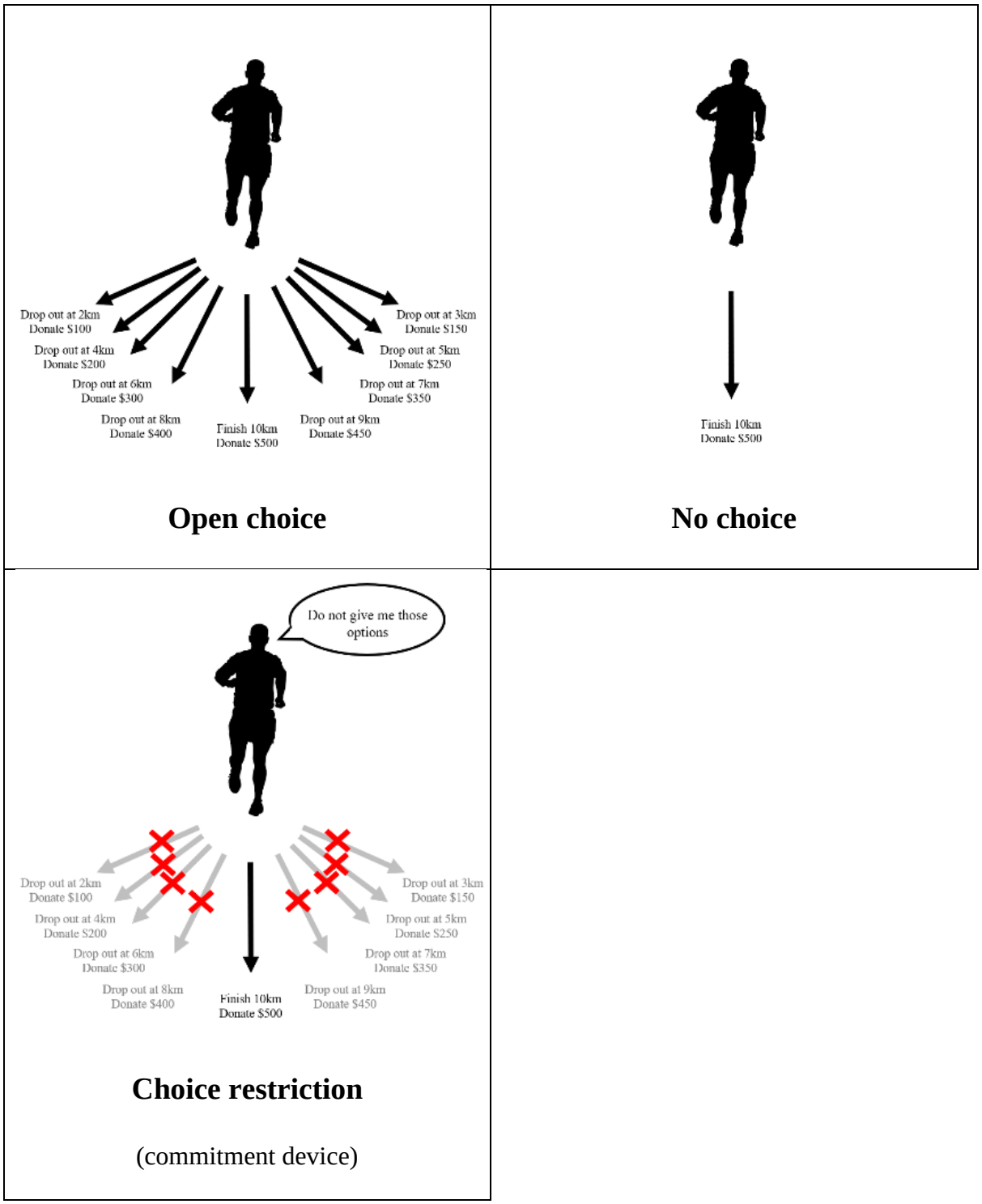
Study 1: Commitment devices as social signals

Before examining the effect of burning one's bridges on others' perception of commitment, we first test the broader phenomenon: the effect of commitment devices on others' perceptions of commitment. Since burning bridges is only one example of a commitment device, it stands to reason that if commitment devices fail to convince others that one is more committed to a course of action, we can consider our hypothesis disproved. On the other hand our hypothesis would be supported if using a commitment device, reducing one's option in an effort to commit, effectively convinces others of one's commitment.

We did this by surveying online participants and presenting them one of three versions of a story. In the story someone promises something costly (finishing a 10k charity race) and either 1) restricts choice using a commitment device (demands donations only if he finishes the full race) or 2) leaves his options open (allowing for donations for every kilometer completed). In a third version 3) we separate internal commitment from external commitment by having a commitment device already in place with no choice (the race is structured to award donations only if runners finish the full race). We predicted that participants would judge the runner employing a commitment device as more likely to finish the race, as more committed, and would make it further in the race on average compared to the character who leaves his options open and the one who already had a commitment device in place.

Study was pre-registered at https://aspredicted.org/GFT_ZTF.

Figure 1. Summary of conditions in Study 1.



Method

We recruited 300 participants (100 per condition) from Prolific. We had no exclusions. Participants read a story about a runner who wants to finish a difficult race in order to secure donations to a charity. Participants read one of three versions:

1) The runner will gain donations for every kilometer of the race he completes even if he does not finish the full race (open choice)
2) The runner will only gain donations for completing the full race, and nothing if he does not (No choice)
3) The runner is offered the first option above, but requests the second option (Choice restriction commitment device)

Following this, participants answered three questions each on a Likert scale: 1) How committed is George to finishing the full 10k distance? (1-7, not at all committed – very committed), 2) How likely is it that George will complete the full race (10k)? (1-7, extremely unlikely – extremely likely), and 3) How far do you think George [the runner] will make it in the race? (1km-10km). Our model predicts higher estimates for all three measures for the choice restriction condition compared to the open choice condition, with the no choice condition between.

Results

Supporting the our hypothesis, people perceived the runner who employs a commitment device to be more committed.

Estimated Distance Completed. Participants were also asked to predict in kilometers how far the runner would make it in the 10km race. In line with our hypothesis we saw a significant impact of condition, according to a Kruskal-Wallis one-way ANOVA, $\chi^2(2)=19.10$, $p<.001$, $\varepsilon^2=.06$ (Figure 1, left). Pairwise comparisons showed participants in the open choice condition gave significantly lower estimates of kilometers completed (M=6.16 SD=1.71) compared to both the choice restriction (M=6.95 SD=2.06, W=-3.95, p=.014) and

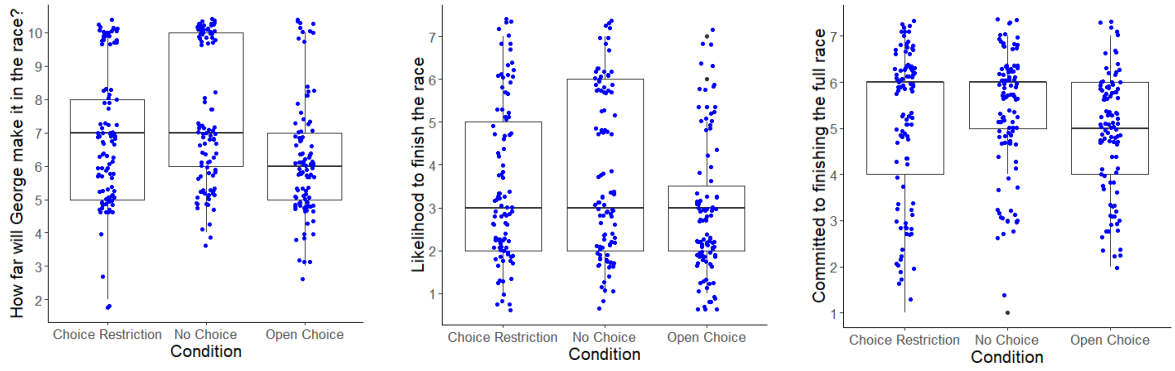
no choice (M=7.47 SD=2.09, W=-6.07, p<.001) conditions. There was no difference between the burning bridges and no bridges condition (W=2.20, p=.265). This supports the hypothesis that participants would judge the character who restricted their choices with a commitment device as likely to run further than the participant who left his options open, but that the participant with no choices from the start would match the pattern of judgment as when the runner restricted their own choice.

Likelihood to finish the race. Participants rated George as being more likely to complete the full race when he restricted his choices compared to when he left them open. There was an overall difference between conditions in a Kruskal-Wallis one-way ANOVA, $\chi^2(2)=9.98$, p=.007, ε^2=.03 (Figure 1, center). In line with our hypothesis pairwise comparisons showed participants thought the runner in the choice restriction condition was more likely to finish the race (M=3.54 SD=1.82) compared to the open choice condition (M=2.93 SD=1.58), W= -3.39, p=.044. Similarly the no choice condition was also rated higher (M=3.75 SD=1.92) than the open choice condition (M=2.93 SD=1.58), W=-4.20, p=.008. There was no difference between the choice restriction and the no choice condition, W=0.93, p=.786. The results support the hypothesis that commitment devices effectively signal that one will accomplish the goal committed to; however, it seems not to matter whether one employs a commitment device oneself, as in the choice restriction condition, or whether the commitment device is already in place, as in the no choice condition.

Perceptions of Commitment. Participants differed in their perceptions of commitment to finishing the race between conditions, although the effect size was small according to a Kruskal-Wallis non-parametric one-way ANOVA, $\chi^2(2)=6.38$, p=.041, ε^2=.02 (Figure 1, right). Pairwise comparisons showed the runner in the no choice condition was rated as more likely to finish the race (M=5.34 SD=1.29) than in the open choice condition (M=4.89 SD=1.32), W=-3.69, p=.025. There was no significant difference between the choice

restriction and the open choice condition, W=-2.19, p=.268. These results support our hypothesis that external commitments matter to perceptions of commitment, but does not support the hypothesis that internal commitments matter. There was no significant increase in commitment from the no choice condition to the restricted choice condition, suggesting that the commitment signal in this kind of commitment device may be due more to the external commitment, the change in incentive structure, than to perceptions of internal commitment, the act of employing a commitment device.

Figure 1. Boxplots of ratings for items in Study 1. Dots are individual participants with applied jitter. There are no significant differences between conditions.



Discussion

Can restricting one's options down to a single goal signal to others that one is committed to accomplishing that goal? Our findings give support to the hypothesis that the presence of a commitment device does actually increase perceptions of commitment, as well as belief that one will achieve the goal to which one is committing. Being in a "do or die" situation increased judgment that runner is committed to his goal, is more likely to complete it, and will make it further to completing it on average compared to a runner who had multiple opportunities to give up.

Our method distinguishes two types of commitment: internal and external. Someone who restricts his own choices deliberately to achieve his goal we propose signals both

internal commitment, i.e. motivation, as well as external commitment, i.e. a strong outside incentive. On the other hand when someone has the same incentive structure through no choice of their own, we propose that signals external commitment but not internal commitment. Our findings suggest that only one of these, external commitment, is necessary for increased perceptions of commitment and goal accomplishment. Across our three measures, a runner who signaled both internal and external commitment saw equal benefits to others' judgment as one who only signaled external commitment. This implies that commitment devices need not be employed by the individual for them to be effective signals: the incentive structure of an external commitment is enough for others to perceive commitment and believe one will achieve one's goal.

Altogether, this provides support for our hypothesis that commitment devices effectively signal commitment because they changes one's options and incentives. It matters less so whether one intentionally brought about that incentive structure and more the incentives in place when predicting behavior. This means that restricting your own options by using a commitment device makes you no different than someone who had no options to begin with. This does not rule out restricting choice as an effective signaling strategy however. Willfully changing your own incentives does seem to engender in others the belief you will accomplish the thing you are incentivizing yourself for. Overall, these results are consistent with the hypothesis that commitment devices can act as a credible signal of commitment and of future behavior. Next, we use an analogous paradigm to test whether a specific kind of commitment device, *burning bridges*, has similar effects in a social intergroup context.

Study 2a- Gang member signaling allegiance to police

Study 1 established that using a commitment device and restricting your options can convince others you are committed. We extend this to an intergroup context with a novel vignette and additional measures. Study 1 showed that an individual can credibly signal to others a goal by restricting their own choices that lead away from that goal. In the same way, a person can signal that they will be a loyal member of a group by reducing their choices of other groups to join, i.e. by burning their bridges.

Burning bridges is a type of commitment device in which one signals loyalty to one group by making themselves unappealing to a rival group. Thus we created a vignette that describes two rival groups, a police organization and a criminal gang, and one individual aiming to signal loyalty to one over the other. Similar to study 1, we create three versions of the story in which the character either 1) restricts his own options, burning bridges to the gang, 2) leaves his options open, trying to stay loyal to both groups, or 3) has no bridges back to the gang though no choice of his own.

We hypothesize that *burning bridges* is a *credible signal of commitment* to a course of action and engenders belief that *the actor will perform that action*. Thus, items test each step in this chain: 1) Do participants believe the individual's bridges are burned with his gang? 2) Do participants think the individual is credibly signaling commitment to leaving the gang? 3) Do participants think the individual *will* leave the gang? Perceived commitment we measure across 2 dimensions: intentions and attitudes. We predict that participants in the burning bridges condition will rate all items higher than participants in the open bridges condition, with the no bridges condition rated in between these two. We also include seven exploratory items for which we have no strong predictions.

Study was pre-registered at https://aspredicted.org/R93_ZQ4

Method

We recruited 300 participants over the age of 18 from the United States through the website Prolific. After excluding participants who gave incomplete data and those who failed a competence check to identify manipulation-relevant information in the vignette, we were left with a final sample of 242. This exclusion rate was higher than expected and upon further inspection we found that the burning bridges condition had the highest rate of exclusions (40%) compared to the no bridges (13%) and open bridges (7%) conditions. It could be that the specific wording of that version of the story tended to confuse more participants about the content. However, the analysis was run in parallel with the full dataset and no differences in results emerged.

Three conditions are presented between subjects which manipulate the situation around the former gang member. The gang member, George, is nearing the end of his prison sentence for assault and expressed to police the desire to leave the gang and work with the police to bring them to justice. Participants are randomly presented one scenario and asked to make judgments about George, his intentions, and his future behavior. The three conditions were:

Open Bridges

In the vignette the former gang member returns to his old life when he is released from prison, moving into his old neighborhood where his friends and family are part of the gang, and maintains contact with old gang members.

No Bridges

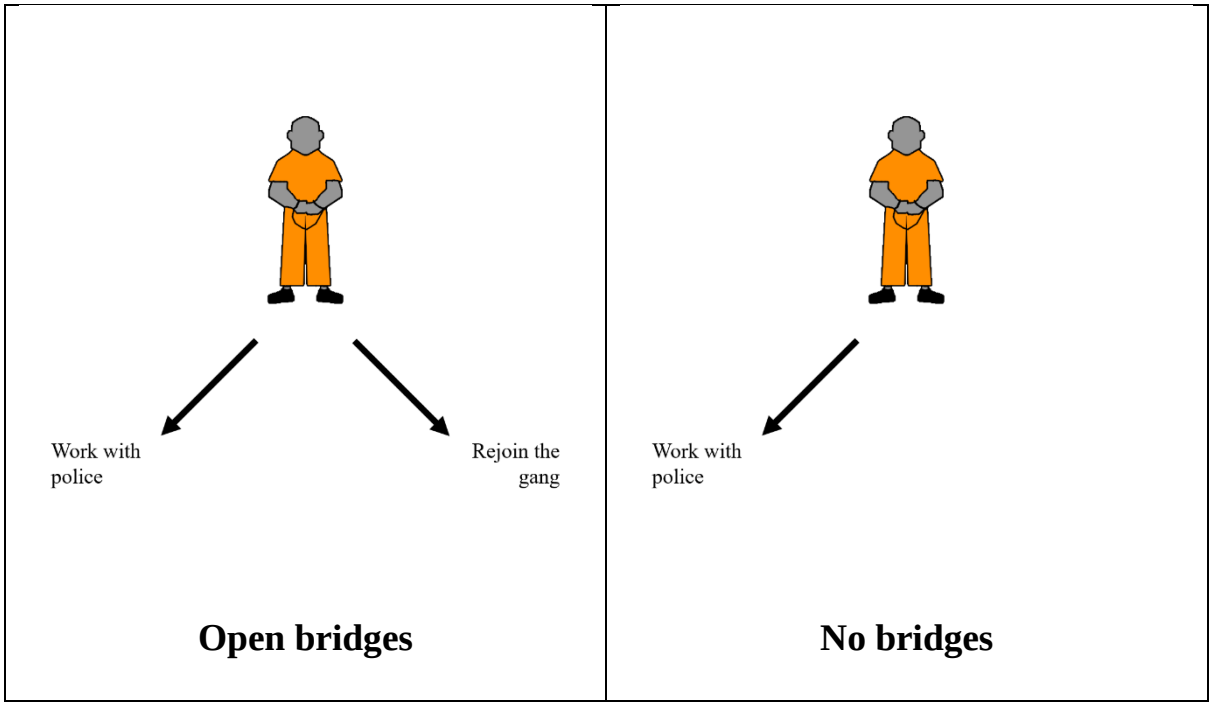
In the vignette the former gang member is accidentally revealed to the gang to be working with the police. For his safety, he is forced to move to a distant city and cut off all contact with his gang contacts.

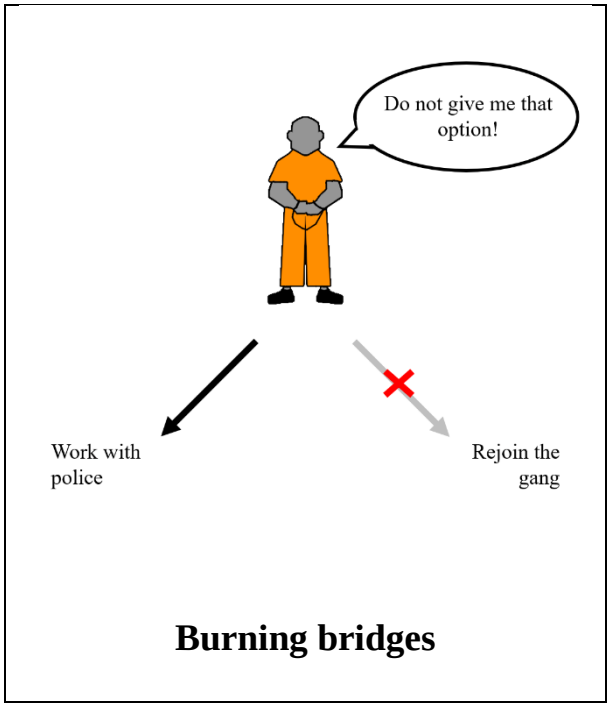
Burning Bridges

In the vignette the former gang member deliberately reveals to the gang that he is working with the police so that the gang knows his intentions. He decides to move to a distant city and cut off all contact with his gang contacts.

We predicted that participants will judge the gang member who burned his bridges to more strongly intend to leave the gang, and will be more likely to follow through on leaving compared to the other conditions. To measure this, we organized a 12-item questionnaire into three steps (Appendix 1 for full questionnaire). All items were measured on an 11-point (0 to 10) Likert scale.

Figure 2. Summary of conditions of Study 2a.





First, we measure whether participants perceive that George burned his bridges with the gang (items 1-3), by asking what participants think the gang thinks of him, and how hard it would be to re-join the gang. These items we predicted would be higher in the burning bridges condition compared to open bridges.

Second, we test whether participants perceive George as believably signaling a commitment to leave the gang (items 4-8). These include ratings of how committed George is to leaving the gang, how loyal he is to the gang (reverse coded) and police. We predict participants in the burning bridges condition will rate George higher in these items compared to participants in the open bridges condition.

Third, we test perceptions of George's likely future behavior (items 9-12). These items asked how much the police should trust George, and how likely it is that George will never re-join the gang.

These three steps form the basis of our model. They measure the intended manipulation, the extent to which George's behavior act as a signal of commitment, and participants' predictions of George's future behavior as a result. As an exploratory analysis,

seven additional items were included to assess George's character. These included whether he was friendly, moral, or trustworthy. Because our model does not have strong predictions of character judgement we did not pre-register predictions for these items and they are not reported in this proposal, but will be in the final publication.

Results

Step 1: Do participants think that George's bridges are burned with his gang?

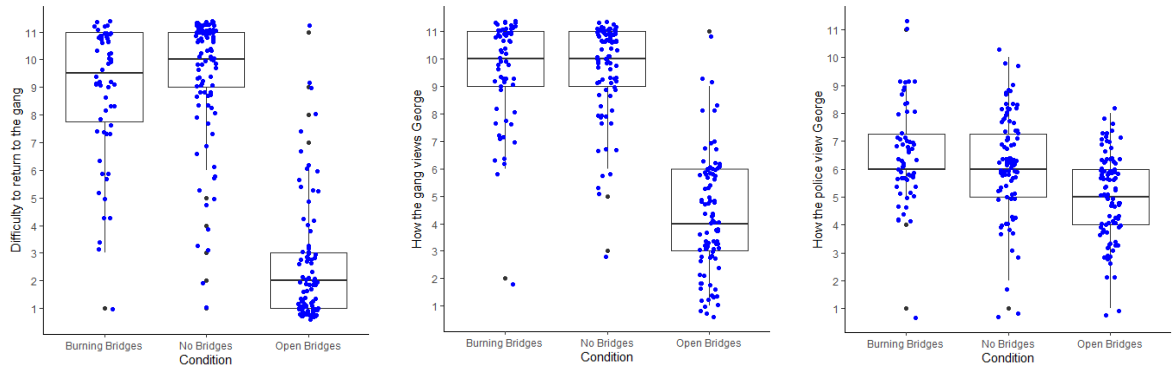
Medians for all measures are reported in Table 1. All measures were significantly non-normal in their distribution according to a Shapiro-Wilk test, $p<.05$, so medians are reported as a best measure. Additionally, all analyses of variance used the non-parametric Kruskal-Wallis test.

Participants shown the burning bridges vignette rated it as more difficult for George to re-join the Six Group compared to the open bridges condition ($W(2)=-13.41$, $p<.001$) and no different than the no bridges condition ($W(2)=1.90$, $p=.373$) ($\chi^2(2)= 145.09$, $p<.001$, $\varepsilon^2=0.60$). Participants shown the burning bridges vignette thought the Six Group viewed George more negatively than the open bridges condition and no different than the no bridges condition ($\chi^2(2)=150.17$, $p<.001$, $\varepsilon^2=0.62$). Lastly, those in the burning bridges condition rated the police as having a better view of George compared to the open bridges condition and no different from the no bridges condition ($\chi^2(2)= 33.13$, $p<.001$, $\varepsilon^2=0.14$). Together, these results suggest that participants judge the open bridges and burning bridges conditions in line with our intended manipulation: when George burns bridges with the gang he is rated as having higher favor with the police and lower with the gang, and rate it as harder to return to the gang than when George left his options open. This suggests participants are sensitive to our manipulation. Further, it supports our finding from Study 1 that external commitment is all that is necessary to influence perception: the no bridges condition (external commitment) is rated similarly as the burning bridge condition (internal and external commitment) (See Figure 4 for summary).

Table 1. Medians table for items in Step 1 that assess the degree to which participants view George as burning his bridges. Items were ranked 0-10 (here coded as 1-11) with higher scores representing more perception of bridge burning. All items show ratings in the burning bridges condition to be higher than in the open bridges condition and similar to the no bridges condition.

Item	Burning Bridges	No Bridges	Open Bridges
1. Difficulty to return to the gang	9.5	10	2
2. How the gang views George (reversed)	10	10	4
3. How the police view George	6	6	5

Figure 4. Boxplots for three items in step 1. Blue dots are individual participant ratings with applied jitter. All items show participants in the burning bridges condition to give higher ratings to the open bridges condition and similar to the no bridges condition.



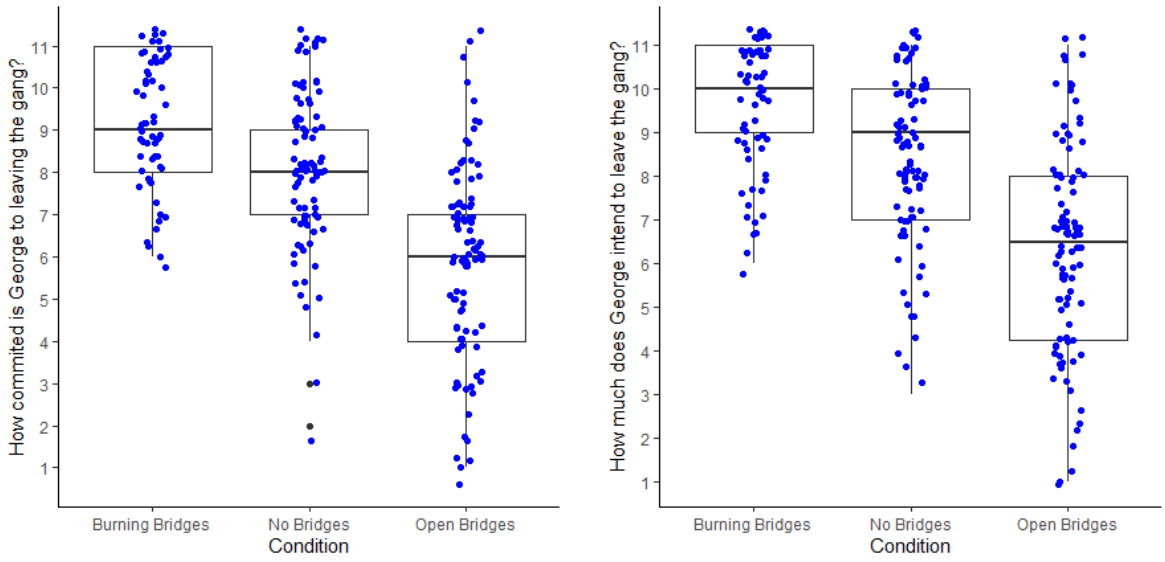
Step 2: Do participants think George is credibly signaling commitment to leave the gang? *Signal of Intention.* Medians are reported in Table 2a. Condition had a significant impact on both perceptions of commitment, $\chi^2(2)= 77.96$, $p<.001$, $\varepsilon^2=0.32$, and intention, $\chi^2(2)= 68.22$, $p<.001$, $\varepsilon^2=0.28$. Participants in the burning bridges condition rated George as significantly more committed ($W(2)=-11.22$, $p<.001$) and having stronger intentions ($W(2)=-10.28$, $p<.001$) compared to the open bridges condition. The no bridges condition fell

significantly between these two (Figure 5). This finding supports our hypothesis that burning bridges is a signal of commitment and intentionality.

Table 2a. Medians table for items in Step 2a that assess the degree to which participants view George signaling commitment to leave the gang. Items were ranked 0-10 (here coded as 1-11) with higher scores representing more perception of bridge burning. Both items show ratings in the burning bridges condition to be higher than in the open bridges condition with the no bridges condition rated in between.

Item	Burning Bridges	No Bridges	Open Bridges
4. How committed is George to leaving the gang?	9	8	6
5. How much does George intend to leave the gang?	10	9	6

Figure 5. Boxplots of signals of intention items in step 2a. Blue dots are individual participant ratings with applied jitter. Both items show participants in the burning bridges condition to give higher ratings to the open bridges condition with the no bridges condition rated significantly in between.



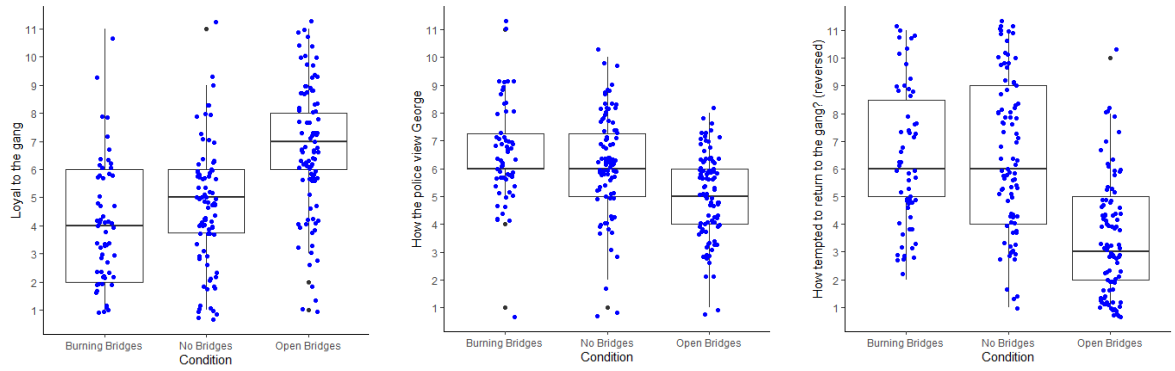
Signal of Attitude. Medians are reported in Table 2b. Condition had a significant impact on ratings of gang loyalty, $\chi^2(2)= 55.06$, $p<.001$, $\varepsilon^2=0.23$, police loyalty, $\chi^2(2)= 33.03$, $p<.001$, $\varepsilon^2=0.14$, and temptation to return to the gang, $\chi^2(2)= 66.42$, $p<.001$, $\varepsilon^2=0.28$. Participants in the Burning Bridges condition rated George as less loyal to the gang, more loyal to the police, and less tempted to return to the gang (reverse coded) compared to participants in the Open Bridges condition. Across all items, there was no difference between the Burning Bridges and No Bridges condition (see Figure 6). This supports our alternative hypothesis that participants would rate the bridge burner higher in attitudes that aligned with his professed commitment (i.e. to leave the gang) compared to when bridges are left open, but no different from when the character has no other options from the start.

Table 2b. Medians table for items in Step 2b that assess the degree to which participants view George's attitudes towards leaving the gang. Items were ranked 0-10 (here coded as 1-11) with higher scores representing more perception of bridge burning. Both items show ratings in the burning bridges condition to be higher than in the open bridges condition with the no bridges condition rated in between.

Item	Burning Bridges	No Bridges	Open Bridges
6. Loyal to gang	4	5	7
7. Loyal to police	7	6	5
8. Tempted to return to gang (reversed)	6	6	3

Figure 6. Boxplots for three items in step 2b. Blue dots are individual participant ratings with applied jitter. When asked to rate George's loyalty to the gang (left) participants in the open bridges condition gave significantly higher ratings than the burning bridges condition. When rating George's loyalty to the police (center) and how tempted George would be to return to the gang (reverse coded) (right) participants in the burning bridges condition gave higher

ratings than those in the open bridges condition. Across items, there were no significant differences between the no bridges and the burning bridges condition.



Step 3: Do participants think George *will* leave the gang?

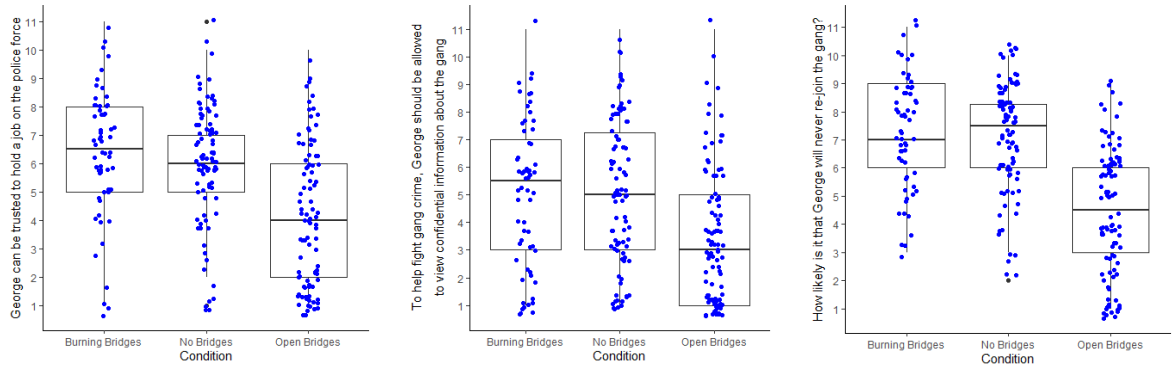
Medians for all measures are shown in Table 3. Condition had a significant impact on ratings of trust the George can work for police, $\chi^2(2)= 35.33$, $p<.001$, $\varepsilon^2=0.15$, that George can be trusted to view confidential police information, $\chi^2(2)= 18.27$, $p<.001$, $\varepsilon^2=0.08$, and likelihood that he will not rejoin the gang, $\chi^2(2)= 66.16$, $p<.001$, $\varepsilon^2=0.27$. Participants in the burning bridges condition rated George higher in all three measures compared to the open bridges condition. This suggests that burning bridges can influence perceptions of future behavior, by increasing belief George will accomplish his professed goals. Additionally, the no bridges condition was not significantly different from the burning bridges condition in all three measures (Figure 7). As in Step 1, this suggested that an external commitment is sufficient to signal future behavior, and adding the internal commitment of burning bridges does not gain the bridge burner any benefit over the person who had no bridges to begin with.

Table 3. Medians table for items in Step 3 that assess the degree to which participants judge George's likely future behavior. Items were ranked 0-10 (here coded as 1-11) with higher scores representing more perception of George leaving the gang. All items show ratings in

the burning bridges condition to be higher than in the open bridges condition with the no bridges condition not different from the burning bridges condition.

Item	Burning Bridges	No Bridges	Open Bridges
9. George can be trusted to hold a job with the police?	6.5	6	4
10. George can be trusted to view confidential police information about the gang?	5.5	5	3
11. How likely is it George will not re-join the gang?	7	7.5	4.5

Figure 7. Boxplots for three items in step 3. Blue dots are individual participant ratings with applied jitter. In all three measures, participants in the burning bridges condition gave significantly higher ratings than those in the open bridges condition, and no different from the no bridges condition.



Discussion

We show that burning one's bridges to a social group can convince others you are committed to a rival group. Across 11 measures we demonstrate that burning one's bridges 1) makes it harder to re-join the group to which bridges were burned, 2) makes the bridge burner appear more committed to the rival group to which bridges were not burned and 3) more likely to stay with the rival group and never rejoin the group to which bridges were burned.

Thus, burning bridges may be an effective way to signal loyalty to one group, when one's commitment is in question.

Burning bridges signals two types of commitment: internal and external commitment. When one has no bridges to begin with, this only signals external commitment. We demonstrate that external commitment alone is sufficient to signal commitment and intention to achieve the goal above when there is no commitment in place. However, by burning one's bridges he also signals internal commitment, e.g. strong motivation, and this further increases judgments of commitment and goal intention. However, when participants judge future behavior, e.g. following through on staying loyal to one group, external commitment is all that matters. There was no difference in perception when one intentionally burned his bridges, compared to when the bridges were not there to begin with.

These results support our hypothesis that burning bridges is a credible signal of commitment and can be used to influence the judgments of others in an intergroup context. One limitation of this study is that participants are likely to sympathize with one group, e.g. the police, and be against the other, e.g. the gang. In this way, participants may be judging from the perspective of an ingroup member, however, our hypothesis proposes that burning bridges signals equally to ingroup and outgroup members. In a proposed follow-up study, we alter the vignette so participants take the role of outgroup members judging a bridge burner and measure the say items.

Proposed Study 2b- Police signaling loyalty to gang

Study 2a demonstrated burning bridges as a credible signal of commitment to one group over another. One limitation of this finding is that participants are likely to take the side of the group to which the individual is signaling loyalty, and not the group he is betraying. Will participants see someone as equally committed and trustworthy if instead they

are leaving a beloved group to join a hated outgroup? Assuming people are more likely to identify with police over a criminal gang, Study 2a shows that burning bridges is a credible commitment signal to the ingroup, but is it just as good a signal to the outgroup? We propose to test this in a follow-up study.

Method

We will recruit 300 participants from Prolific. Participants will read a vignette similar to study 2a, except it will describe the inverse situation: a former police officer aiming to join a criminal gang. Conditions will be analogous to study 2a, and measures, including exploratory items will be identical.

Summary

It is just as important to understand how burning bridges is understood by the outgroup as it is by the ingroup. As we will explore in studies 3 and 4, often the negative impacts on society caused by burning bridges come from outgroup perceptions. We will go on to test how burning bridges functions as a signal of group loyalty in political partisanship.

Proposed Study 3: Burning Bridges to signal partisan loyalty

Study 3 will test burning bridges in a political context. Studies 2a and 2b tested judgments of burning bridges in two sides of an intergroup conflict where the judge was either for (Study 2a) or against (Study 2b) the group being joined. Now we test both at once: We will see how someone on the political right who burns bridges to the left is judged by both the right and left, and vice versa. We predict that burning bridges is a credible signal of commitment regardless of which side the observer is on. We expect ratings of perceived commitment to be higher for bridge burners, regardless of whether bridges are being burnt on the political left or right and whether the partisan judge affiliates with the left or right. In contrast, we expect character judgments like whether the bridge burner is trustworthy, moral,

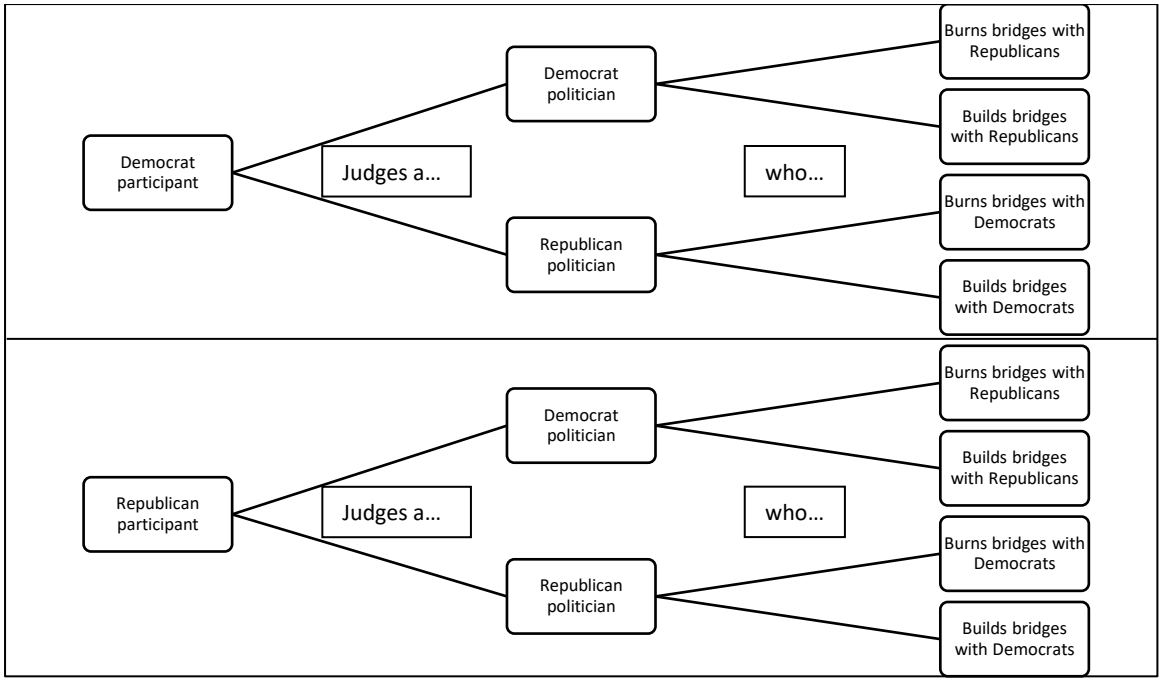
or competent to differ between observers and who they are judging. Our model predicts that burning bridges is a deliberate sacrifice of reputation with one group to improve it in a rival group. Thus, we predict that partisans who affiliate with the group to which bridges are burned will judge the bridge burner negatively on these items compared to partisans who affiliate with the group the bridge burner is trying to join.

Method

We will recruit 400 American participants from Prolific. Using in-built filters from Prolific, half of our sample will be affiliates of the Republican party, and half will be affiliates of the democratic party. In the survey we will also measure participants' levels of partisan animosity (Stancato & Keltner, 2021). We anticipate our predicted effects are strongest among partisans with the highest levels of animosity, with a moderating relationship.

Participants will be shown mock social media posts by would-be politicians running for office. Each politician is said to either be 1) Republican or 2) Democrat and the content of the post either 1) Burns bridges with the opponent group by refusing to compromise or work with them or 2) Builds bridges with the opponent group by promising compromise and collaboration. Participants will be given the same measures as in Studies 2a and 2b, including the exploratory items adapted for the social media posts

Figure 8. Proposed design of conditions for Study 3. Participants will be recruited who are either self-reported Republicans or Democrats. Participants will each be assigned to rate four mock social media posts by four different politicians, each of which falls into one of the four categories on the far right of the figure.



Summary

Previous studies tested the burning bridges model using abstract situations, here we apply it to a real-world intergroup conflict: political polarization. By testing the model in this way, we effectively measure both ingroup and outgroup perceptions of a bridge burner and can see both status loss in the outgroup as well as status gain for the ingroup.

Study 4: Professing extreme beliefs as burning bridges

Previous studies applied the burning bridges hypothesis to intergroup conflict and political partisanship, but there is another context in which burning bridges may be a useful signal: extreme beliefs. In his discussion of the burning bridges phenomenon, Mercier (2020) provides examples of extreme and baffling beliefs that people publicly profess: that the Earth is flat, that the Holocaust did not happen, and North Korean politicians proclaiming that Kim Jong Un can teleport. Even academic scholars make claims that seem ridiculous to most other academics: a philosopher argues that genetics is just as unscientific as creationism (cited in Coyne, 2018) and a postmodern scholar proclaims that pandemics were caused by “global

industrial resource extraction" instead of more ordinary causes like poor hygiene (cited in Coyne, 2017). A key feature of these ridiculous claims is that they are rejected by *most* of the academic community, but *not* by the particular group the speaker belongs to, in this case philosophers and post-modernists. In this way, Mercier (2020) argues, these extreme claims act as a unique identity marker and a signal of group commitment by burning bridges to other academic disciplines, by making the speaker look ridiculous to them. So too, conspiracy theories around the covid vaccine, 5g towers or politicians involvement in human trafficking may be publicly endorsed because the speaker wishes to signal commitment to one group by burning their bridges to the rest of society.

Method

We will recruit 400 participants from Prolific, 100 per condition. Participants will read a short story in which they take the role of a member of a small group that either 1) is greatly disliked by much of the general populace for their beliefs or 2) is treated as ordinary by the general populace. Participants will be asked to judge a prospective member of the group who either 1) publicly proclaims support for the most extreme version of the group's beliefs or 2) is not public about his support for the group. The full 2x2 design is shown in Figure 9. We will collect measures similar to Study 2a, but adapted to the vignette.

Figure 9. Conditions in Study 4.

	Fringe Group	**Ordinary Group**
Extreme public support	*Condition A*	*Condition B*
Moderate public support	*Condition C*	*Condition D*

We predict that participants in the fringe group which is disliked by the general populace will perceive higher commitment from the prospective member who proclaims

extreme support for the group's ideals (condition A) compared to the prospective member who proclaims only moderate support (condition C). We also predict this difference will be greater than the difference in the analogous conditions of condition B and D, when the group is seen as ordinary. We predict that for the Fringe Group, proclaiming extreme public support will burn bridges to the general populace, because the group is so disliked, and thus be a more credible signal. For the ordinary group which is well-liked by the populace, bridges are not burned and thus it is not a credible signal of commitment.

Summary

Burning bridges is an effective signal of commitment, but this may be especially true for fringe or extremist groups. Because merely proclaiming support for this group makes one disliked by outgroup members, supporting the group itself becomes an act of burning bridges, and so signals commitment. We aim to show that publicly supporting a fringe group is perceived as burning bridges to the outgroup to a higher degree than supporting an ordinary, well-liked group; furthermore, publicly supporting a fringe group should better convince members of that group that you are committed, compared to an ordinary group.

References

Abramowitz, A. I., & Webster, S. (2016). The rise of negative partisanship and the nationalization of US elections in the 21st century. *Electoral Studies, 41*, 12-22.

Bryan, G., Karlan, D., & Nelson, S. (2010). Commitment devices. *Annual review of Economics*, *2*(1), 671-698.

Coyne, J. (2018) "The University of Edinburgh and the John Templeton Foundation royally screw up evolution and science (and tell arrant lies) in an online course," Why Evolution is True, https://whyevolutionistrue.wordpress.com/2018/03/25/the-university-of-edinburgh-and-the-john-templeton-foundation-royally-screw-up-

evolution-and-science-and-tell-arrant-lies-in-an-online-course/ (accessed January 6, 2023).

Coyne, J (2017). “A postmodern holiday: Recent nonsense from the humanities,” Why Evolution is True, https://whyevolutionistrue.wordpress.com/2017/01/10/a-postmodern-holiday-recent-nonsense-from-the-humanities/(accessed January 6, 2023).

Crimston, C. R., Selvanathan, H. P., & Jetten, J. (2022). Moral polarization predicts support for authoritarian and progressive strong leaders via the perceived breakdown of society. *Political Psychology*, *43*(4), 671-691.

De Hooge, I. E., Breugelmans, S. M., & Zeelenberg, M. (2008). Not so ugly after all: when shame acts as a commitment device. *Journal of personality and social psychology*, *95*(4), 933.

Gidron, N., Adams, J., & Horne, W. (2019). Toward a comparative research agenda on affective polarization in mass publics. *APSA Comparative Politics Newsletter*, *29*, 30-36.

Giné, X., Karlan, D., & Zinman, J. (2010). Put your money where your butt is: a commitment contract for smoking cessation. *American Economic Journal: Applied Economics*, *2*(4), 213-35.

Hahm, H., Hilpert, D., & König, T. (2022). Divided by Europe: affective polarisation in the context of European elections. *West European Politics*, 1-27.

Hall, D. L., Cohen, A. B., Meyer, K. K., Varley, A. H., & Brewer, G. A. (2015). Costly signaling increases trust, even across religious affiliations. *Psychological science*, *26*(9), 1368-1376.

Hopp, T., & Vargo, C. J. (2019). Social capital as an inhibitor of online political incivility: An analysis of behavioral patterns among politically active Facebook users. *International Journal of Communication*, *13*, 21.

Imhoff, R., & Lamberty, P. (2020). A bioweapon or a hoax? The link between distinct conspiracy beliefs about the coronavirus disease (COVID-19) outbreak and pandemic behavior. *Social Psychological and Personality Science*, *11*(8), 1110-1118.

Kurzban, R., & Christner, J. (2011). "Are supernatural beliefs commitment devices for intergroup conflict?" In J. P. Forgas, A. Kruglanski, & K. D. Willimas (Eds.), *The psychology of social conflict and aggression* (pp. 285–300). Sydney Symposium of Social Psychology, vol. 13). New York: Taylor and Francis.

Lee, F. E. (2015). How party polarization affects governance. *Annual review of political science*, *18*, 261-282.

McAndrew, F. T. (2021). Costly signaling theory. *Encyclopedia of evolutionary psychological science*, 1525-1532.

McCoy, J., Rahman, T., & Somer, M. (2018). Polarization and the global crisis of democracy: Common patterns, dynamics, and pernicious consequences for democratic polities. *American Behavioral Scientist*, *62*(1), 16-42.

Pew Research Center. (2022). The polarization in today's Congress has roots that go back decades. Retrieved December 1, 2022, from https://www.pewresearch.org/fact-tank/2022/03/10/the-polarization-in-todays-congress-has-roots-that-go-back-decades/

Stancato, D. M. & Keltner, D. Awe, ideological conviction, and perceptions of ideological opponents. *Emotion* **21**, 61–72 (2021).

Van Prooijen, J. W., Krouwel, A. P., & Pollet, T. V. (2015). Political extremism predicts belief in conspiracy theories. *Social Psychological and Personality Science*, *6*(5), 570-578.

Wagner, M. (2021). Affective polarization in multiparty systems. *Electoral Studies*, *69*, 102199.

Appendix 1. Study 2 survey items

Step 1: Do participants think that George's bridges are burned with his gang?

1. At this point, if George wanted to re-join the Six Group, how difficult would that be for him? (0-10, Extremely Difficult- Extremely easy) (reverse coded)
2. At this point, how do you think the Six Group views George? (0-10 Very Negatively-Very Positively) (Reverse coded)
3. At this point, how do you think the police view George? (0-10 Very Negatively-Very Positively)

Step 2: Do participants think George is credibly signalling commitment to leave the gang?

Signal of Intention

4. How **committed** is George to leaving the Six Group for good? (0-10 not at all committed – Extremely committed)
5. How much does George **intend** to leave the Six Group for good? (0-10 does not at all intendent to leave – Completely intends to leave)

Signal of Attitude

Please rate how much the following characteristics describe George.
George is…

6. Loyal to the Six Group (0-10, strongly disagree – strongly agree) (Reverse coded)
7. Loyal to the police and the NGCA (0-10, strongly disagree – strongly agree)
8. How **tempted** do you think George will be to return to the Six Group? (0-10 not at all tempted-Very tempted) (reverse coded)

Step 3: Do participants think George will leave the gang?

9. How much do you agree with this statement: George can be **trusted** to hold a job on the police force (0-10 Strongly disagree – strongly agree)
10. How much do you agree with this statement: To help fight gang crime, George should be allowed to view confidential information that the police have about the Six Group (0-10 Strongly disagree – strongly agree)
11. How likely is it that George will continue helping the police, and never re-join the Six Group? (0-10, Very unlikely to very likely)
12. Please explain your reasoning (free response)

Exploratory Items

Please rate how much the following characteristics describe George.
George is…

13. Friendly and warm (0-10, strongly disagree – strongly agree)
14. Moral (0-10, strongly disagree – strongly agree)
15. Competent (0-10, strongly disagree – strongly agree)
16. Trustworthy (0-10, strongly disagree – strongly agree)
17. Foolish and unwise (0-10, strongly disagree – strongly agree)
18. Strong-willed (0-10, strongly disagree – strongly agree)
19. Honest (0-10, strongly disagree – strongly agree)
20. Brave (0-10, strongly disagree – strongly agree)