Diffusion of Responsibility and Social Damage*

Pablo Ignacio Soto-Mota† & Adrian Vargas-López ‡

October 8, 2020

Abstract

Diffusion of responsibility is a phenomenon widely studied in the psychological sciences. It happens when people diffuse their responsibility within a group in a given situation. The conditions in which it occurs encompass from crime to emergencies. In this study, we propose an experiment for examining the diffusion of responsibility and its relationship with social damage to the environment. We analyze the students' acceptance rate for increasing a private monetary pay-off instead of donating for reforestation efforts. Our main conclusion suggests that being in a smaller-size group decreases the environmental damage. In other words, we find that people diffuse their responsibility as the number of individuals in the group increases. We also detect mild evidence that individuals reduce their social damage whenever they are directly responsible (i.e., pivotal). These results add to the discussion of the diffusion of responsibility, open a stream of attention to environmental damages, and congregate both in the same conversation.

Keywords: Diffusion of responsibility, social damage, environmental damage, reforestation efforts.

*This version is a Working Paper. Let us know at pablo.soto.mota@nhh.no and a.vargaslopez@tec.mx before citing our results.

†Ph.D. Research Scholar at Norwegian School of Economics (NHH) - FAIR.

‡Ph.D. Candidate at Tecnológico de Monterrey (TEC) - EGobyTP.

1 Introduction

Evidence suggests that diffusing moral responsibility in a group is a strategy when people take actions that benefit them at the cost of generating social damage (Guerin, 2011). For example, individuals contribute to climate change by not reducing carbon emissions. Scientific research points out that leaving uncut the carbon emissions will produce irreversible changes in the environment and generate costs for all. However, some individuals still neglect these warnings. A reason might be that they feel they are not to blame individually, but that society is responsible as a whole, and they perceive it as unlikely that their actions will be decisive. In other words, they diffuse their responsibility in a bigger group and perceive themselves as indirectly responsible.

The studies regarding this social phenomenon began in the late '50s with Freeman et al. (1957) examining how customers gave lower tips as the diners group increased. A decade later, Darley and Latané (1968) analyzed bystanders' reactions to an emergency. In their pioneering work, they measured the time it took participants to react and to provide assistance in an emergency-like situation. Their experiment suggested the bystander's response as a function of other observers rather than the participants' indifference to the victim.

Many studies have followed-up these initial efforts in various settings and with minor tweaks to the original. The latest examples are the ones by Hortensius and de Gelder (2018) and by Falk and Szech (2017). The first study shows that an individual has a mix of distress (from the situation) and sympathy (for the victim), which plays an essential part in the way people react to the circumstances. The second study analyses the rates of people willing to kill mice by receiving a payment in return. These authors maintain that being pivotal reduces the number of moral transgressions in people (i.e., killing mice).

In this study, we investigate how the size of the group and being or not pivotal (i.e., responsible) in the choice of generating social damage affects the probability that a student will accept a deal with a small private benefit but with negative consequences for society. We hypothesize that if individuals are part of bigger groups and have the perception that their actions will not determine the outcome, they will be more willing

to take part in a socially harmful deal.

The game we propose asks participants to accept a socially undesirable deal or to reject it. The deal duplicates their potential monetary personal earnings if they win a lottery. Rejecting the deal implies that the subject remains with the initial sum and accepts a socially desirable outcome. Our experiment is different from the literature in two ways. The first is that we are not in the situation of an emergency. We create a socially desirable outcome by donating money for planting trees without the urgency posed by any crisis. The second is that we are the first to explore how the diffusion of responsibility operates for social damage made to the environment.

In the literature, we noticed that there is a discussion on public goods, collective action, and the size of a group. Given the nature of the social damage we use, it is arguable that our setting is similar to a public goods' scenario. There is no consensus on how the size of the group influences collective action on public goods (Poteete and Ostrom, 2004). The debate seems to be an ongoing thing where some find it highly problematic, and others do not. Given our results, we hope to add a bit to the discussion by incorporating the responsibility mechanism. An explanation that could shed light on how the size of a group increases social damage (or reduces the provision of a public environmental good).

The rest of the document is organized as follows. In the second section, we describe the literature regarding the diffusion of responsibility and explain the contribution this study makes. Next, we define our experimental design. In the fourth section, we present the participants in the experiment and some demographic variables. Later, we show our main results and additional analysis. In the sixth section, we present a theoretical model that informs our results. Finally, we conclude and give policy recommendations.

2 Literature Review

Diffusion of responsibility explains why people acting in a group feel less accountable than those operating individually (Guerin, 2011). In general, the diffusion of responsibility

explains why sending private e-mail requests one at a time increases the length and frequency of responses (Barron and Yechiam, 2002), and why in computer-mediated chat groups, as the group's size increases, it takes longer for an individual to receive help (Markey, 2000). It is also a way for athletes to account for their moral transgressions in competitions (Corrion et al., 2009). More concerning, though, is that judges give longer sentences when the criminal acts alone and shorter when committing the same crime but in a group (Feldman and Rosen, 1978).

Researchers have asked whether the diffusion of responsibility is exclusively bound to adults. It is not. Infants are prone to this phenomenon too (Plötner et al., 2015). The certainty of responsibility is crucial for children (Plötner et al., 2015). Since they are willing to help, as long as it is clear, it is their responsibility to do it. Additionally, we are also predisposed to increasing its effects when presented with a dehumanized version of victims and when our responsibility is not clear (Bandura et al., 1975).

The studies on this social phenomenon explore two essential elements. The first one features how people react and assist those in need during crises, known as the bystander effect. The second befalls when individuals delegate an unpopular decision as they play a version of the dictator's game—commonly known as the delegation of responsibility.

2.1 The bystander effect

Darley and Latané (1968) observe that the diffusion of responsibility is the main reason someone helps a victim differently when she is alone than in the presence of bystanders. They analyze the time it takes the participant to assist someone in need. People usually take longer when the victim and the participant are in the presence of others. Unfortunately, they take even longer if she perceives the bystanders are more fit to assist than herself (Hortensius and de Gelder, 2018).

The mechanisms that Hortensius and de Gelder (2018) highlight are the distress and the sympathy that the participant feels. Both are highly dependent on personality. As they conclude, personality factors dictate whether someone feels apathy for the victim and the situation. Bystander apathy, they argue, surges as a reflex from the competition

of the sympathy and distress the subject feels. Apathy wins whenever distress is higher than sympathy. However, the exact mechanism by which the bystander effect takes place remains underexplored.

The social cueing theory examines the bystander effect under the assumption that others can help (Barron and Yechiam, 2002). It alludes to the *cue value* of bystanders (Ross and Braband, 1973). The signal that a disabled participant (i.e., being blind) sends is different from the one sent by someone fully able. The blind person cannot help in an emergency condition, and that is the signal. The fully able participant interprets this and acts as if she was by herself. The cue value of bystanders determines whether someone feels they are facing an emergency alone or others can help. The evidence is mounting, and it flags that individuals find it harder to diffuse responsibility when facing a stressful situation alone (Bickman, 1972; Plötner et al., 2015).

Mynatt and Sherman (1975) asked whether the success or failure of an outcome makes any difference in the responsibility attribution for team members. They find that teams whose advice leads to failure assume lower levels of responsibility. Combining an unfavorable outcome with responsibility, they claim, seems to be a critical factor. The big takeaway is that people may find it challenging to cope and assume their responsibility if the outcome is negative.

2.2 The delegation of responsibility

Individuals tend to delegate unpopular decisions to a third party (Bartling and Fischbacher, 2012; Coffman, 2011; Hamman et al., 2010; Oexl and Grossman, 2013). They do it because they want to bypass the potential punishment (Bartling and Fischbacher, 2012). When the delegee picks an unfavorable outcome for most, the punishment gets assigned to her in higher proportions. If the dictator chooses an unfair outcome, then a higher proportion of punishment is allocated to the dictator.

Dictators can delegate responsibility to chance. Whenever uncertainty is present, people tend to lower their punishment, both in total amount and as a proportion. It could be because they feel the dictator had nothing to do with the decision even though

she delegated her responsibility instead of choosing the fair allocation of resources in the first place.

Avoiding responsibility entails shirking a cognitive burden similar to guilt (Coffman, 2011). People feel they are to blame because they are aware of their actions. They are aware that others less fortunate suffer the most when the allocation of resources is unfair. However, using an intermediary is still the best strategy to reduce punishment (Coffman, 2011).

The problem with delegation and diffusion of responsibility is twofold. In the first place, people blame those that chose an unfavorable outcome, even if that person feels she is just following orders (Hamman et al., 2010), which means that intermediaries believe that their responsibility is limited. In the second place, dictators find it easier to cope with guilt when using an agent. They are cleared of blame as punishment allocation goes to the intermediary.

2.3 Why we are different from Falk and Szech (2017)

Recently, Falk and Szech (2017) suggested that individuals might not only diffuse their responsibility to avoid helping a person in need but that the same mechanism could explain why individuals in groups are more willing to generate social damage. In their study, they observed that people were more willing to accept killing mice for money when they were in a group and knew that their actions might not be relevant for the final choice. In other words, not being pivotal diffused their responsibility, and made their action more likely.

The study by Falk and Szech (2017) can be seen as another point of view of the *bystander effect*, where individuals decided not to perform a socially desirable action, as helping someone in need because they felt that others were more responsible for it. In the study by Falk and Szech (2017), individuals take a socially undesirable action, killing mice, because they felt that they were less responsible for it.

Our study builds on their paper but differentiates in two relevant dimensions. First, they do not investigate the group's size effect but only the fact of being pivotal or not.

We investigate both things and their potential interaction. Second, their understanding of what *not being pivotal* means is different from ours. In their experiment, *not pivotal* individuals had higher priming on the probability that the damage had already been generated, so that they could benefit without any responsibility. In contrast, as it will be evident in the next section, in this study, the action of *not pivotal* individuals matters. They add up towards the limit, but it is not decisive.

We argue that both ways of understanding pivotality are valid, though serve different purposes. The situation we propose is closer to global challenges as climate change or the overuse of resources where the limit is unknown, every action adds to it, and all will pay the social costs. Lastly, we consider that our design is more straightforward and avoids killing animals.

3 A theoretical explanation of our results, and one way to reinforce the hypotheses: The deal game

Suppose a situation in which t players are sequentially offered a deal with the following two options:

A Accept the deal and get utility $u > 0$. However, if at the end $s \leqslant t$ people accepted the deal, a social damage of size $d > 0$ will be generated. This means that all t individuals will have an inescapable loss of utility of size d whenever there is social damage.

B Reject the deal and have a utility of 0 if the social damage is not generated. Get $-d$ if the damage is generated.

Assume that individuals are pro-social, so they derive a loss of utility c from causing harm to others. This loss of utility comes from two sources:

1. A loss of utility from being part of a group that causes harm. In other words, a share of the responsibility for causing harm. We can denote this as $a(s, d)$ since the responsibility is divided across the s members.

2. A loss of utility from the marginal harm generated by the decision. This may be understood as being seen, or internally feel, as crucial to cause harm. Assume that being crucial means being the person who completes the share of s individuals needed to cause harm. In other words, being pivotal in the choice. Denote being pivotal as $p = 1$ and not being pivotal as $p = 0$. This loss of utility will be represented by $m(p, d)$.

Finally, assume that the t players believe that the probability that at least s individuals will accept the deal is π. Thus,

$$c(p, s, d) = m(p, d) + \pi a(s, d) \tag{1}$$

3.1 Conditions for acceptance

Given the setting of the deal game, a player will accept if:

$$u - \pi d - m(p, d) - \pi a(s, d) \geqslant 0$$

And, simplifying terms,

$$u - \pi[d + a(s, d)] - m(p, d) \geqslant 0 \tag{2}$$

Observe that for pivotal individuals (these are individuals that know that $s-1$ players already accepted the deal) $\pi = 1$ since they can be sure that the damage will be generated if they accept, and $m(p = 1, d) = d$, since the marginal harm they create with their decision is d. Thus, we can rewrite equation (2) as:

$$u - 2d - a(s, d) \geqslant 0 \tag{3}$$

Similarly, notice that there are types of not-pivotal individuals: *Type a:* the ones who know that less than $s-1$ players have accepted the deal, and *Type b:* the ones who know that more than s individuals already accepted the deal. For both types, $m(p = 0, d) = 0$, since in marginal terms they don't create any harm with their decision. Thus, for *Type*

a players, equation (2) can be rewritten as:

$$u - \pi[a(s,d) + d] \geqslant 0 \tag{4}$$

And for *Type b* players, the condition of acceptance is:

$$u - d - a(s,d) \geqslant 0 \tag{5}$$

4 Experimental design

Participants were part of a more extensive survey on beliefs and behaviors about food waste. For participating in the survey, they entered a raffle for one of ten prizes of MXN 400 (USD 18-19, depending on the exchange rate). Also, they had an initial 2% chance of winning MXN 300 (USD 13-14) if they agreed to participate. In the experiment, participants answered their willingness to accept a deal to duplicate their initial MXN 300 potential payment in exchange for increasing the chance of generating social damage. After the experiment, they answered a series of questions about their emotions, moral beliefs, and predictions of others' behaviors. One of these questions asked to guess the percentage of people who answered before. This question was incentivized with an extra potential payment of MXN 50 (USD 2-3) if their answer was closer to 5% of the correct number. Therefore, their maximum possible payment was MXN 650 (USD 29-30). The latter is equivalent to 5.2 days at the minimum wage in Mexico. The social damage they could generate was to eliminate a donation of MXN 400 to reforestation efforts. According to the donation platform, TeamTrees[1], the amount we set is sufficient to plant 20 trees.

The incentive to duplicate their private earnings in the experiment was small. If they accepted the deal, the payment they could win was of only MXN 12 (USD 0.54) in expected terms. It is equivalent to one-tenth of the daily minimum wage. However, we fixed the social damage at MXN 400, independent of how many people ended up earning

[1]TeamTrees is a donation platform. Each dollar donated plants one tree in a forest of high need around the world. For more information, go to https://teamtrees.org/.

money. In other words, the social damage was higher than the increase in expected private gains. This situation mimics the real choices of individuals concerning the environment. For example, wasting a resource such as water by taking to much time in the shower has a diminished positive impact on anyone's utility, but could lead to a substantial loss in social terms.

The experiment has a 2x2 design, in which we informed the participants of offering the deal to several individuals. This quantity was either 100 or 400. Also, we told them that the donation depended on the number of individuals that accepted the deal. We would not donate if the number of participants who accepted reached a threshold. The threshold was ten if the number of participants was 100, and 25 if the number of participants was 400. Finally, we primed them on the possibility of being pivotal or not in the decision to donate. The *Non-Pivotal* prime stated that if the person accepted, four others would be required to accept for canceling the donation. Non-pivotal participants had the same social norm on both sizes (20 out of 400 and 5 out of 100 are 5%). The *Pivotal* prime stated the possibility that the individual would be the one needed for reaching the threshold.

In the experiment, there are two varying factors: the threshold or size of the group needed to generate social damage (Big or Small), and being or not pivotal (Pivotal or Non-Pivotal). Thus, there are four experimental groups:

- Big group, pivotal condition (BP).
- Big group, non-pivotal condition (BNP).
- Small group, pivotal condition (SP).
- Small group, non-pivotal condition (SNP).

For example, a participant in the Big and Pivotal group would face the following choice:

Right now, in case of winning, we would give you MXN 300. However, you have the chance to double your prize by accepting the following deal.

Deal:
If you win, your reward will double. In other words, instead of receiving MXN 300, you would receive MXN 600. However, if 25 participants out of 400 that are receiving this offer accept this same deal, we will not donate the money to plant trees to the # TeamTrees project.

24 participants may have already accepted the deal. It means that if this is the case and you accept, we would not donate the $400 to the reforestation efforts.

In contrast, a person in the Small and Non-Pivotal group would decide after reading the following:

Right now, in case of winning, we would give you MXN 300. However, you have the chance to double your prize by accepting the following deal.

Deal:
If you win, your reward will double. In other words, instead of receiving MXN 300, you would receive MXN 600. However, if 10 participants out of 100 that are receiving this offer accept this same deal, we will not donate the money to plant trees to the # TeamTrees project.

5 participants may have already accepted the deal. It means that if this is the case and 5 other people accept, we would not donate the $400 to the reforestation efforts.

We argue that these two situations modify the opportunity to diffuse responsibility in others. Individuals may think they are part of a larger group and feel less responsible for the social damage. More people might contribute to the generation of social damage if they think their choice is not pivotal, even though they all contribute equally.

4.1 Hypotheses

We predict that both the size of the group and being pivotal or not have an effect on the propensity to generate social damage. Our main hypotheses are:

1. People in the Big group condition will generate more social damage than those in the Small group.

2. People in the Non-Pivotal condition will generate more social damage than those in the Pivotal condition.

Thus, the group that will generate more social damage is the group BNP, while the group that will generate the least social damage is the group SP.

5 Participants

We recruited participants by distributing a survey through the institutional e-mail accounts of two universities in Mexico City: the Center for Research and Teaching in Economics (CIDE) and the School of Social Science and Government at Tecnológico de Monterrey (TEC). Both institutions offer bachelor and graduate-level programs specialized in Economics, Public Policy, International Relations, and Law. The main difference between them is that CIDE is a public-funded institution, while TEC is a private university.

Authorities at both universities collaborated with us in distributing the link of the survey to their students. The distribution was done in three moments between February and April 2020. First, bachelor students at TEC received the link to the survey and the experiment. At a second moment, both bachelor and graduate students at CIDE received the survey. Finally, graduate students at TEC received the link. The survey was active for one week at each school. We sent the payments and donations one week after the link closed. It is crucial to notice that the second and third moments happened when regular courses got suspended in Mexico due to the coronavirus crisis. The suspension was for face-to-face courses. All the academic activities resumed later online.

We reached 911 students through their e-mails. However, only 307 (33.69%) answered at least one part of the survey or the experiment, and 280 completed it. We took into account valid answers from unique e-mails that were completed between 2 and 20 minutes[2]. After imposing the restrictions, we kept 206 valid answers (21.16% from TEC and 24.49% from CIDE).

Participants were randomly distributed into four groups corresponding to the experimental conditions. All groups were balanced across the socio-economic variables[3]. The only significant difference was between the SNP and BNP group with respect the variable Family Size (Difference: 0.752, p=0.05).

Table 1: Balance across experimental groups

	BP (N=55)	**BNP (N=46)**	**SP (N=65)**	**SNP (N=40)**
	Mean/SE	**Mean/SE**	**Mean/SE**	**Mean/SE**
Age	23.607 [0.806]	23.617 [0.798]	22.953 [0.661]	23.308 [0.807]
Gender	0.536 [0.067]	0.447 [0.073]	0.563 [0.063]	0.487 [0.081]
Family Size	2.304 [0.242]	1.915 [0.208]	2.344 [0.181]	2.667 [0.218]
Financial Aid	0.618 [0.066]	0.574 [0.073]	0.547 [0.063]	0.6 [0.078]
Ecofriendly Index	0.365 [0.015]	0.383 [0.015]	0.389 [0.015]	0.405 [0.021]

Students at both schools differ in socio-economic variables. Students at CIDE are on average older and live with fewer people. Also, a higher proportion of students at CIDE are male and receive financial aid. We underscore that all participants from CIDE answered the experiment after the coronavirus crisis started in Mexico, while only postgraduates from TEC participated in these circumstances. Thus, differences in the behavior across schools would not necessarily be the same in the absence of the crisis.

[2]According to Qualtrics, the estimated time for answering the experiment is between 5 and 8 minutes.

[3]*Gender* takes the value of zero if the participant is male. *Family size* is the reported number of people the student shares a house. *Financial aid* takes the value of one when the individual receives a scholarship or a loan to complete her studies. The *Ecofriendly index* was built based on six survey questions related to environmentally responsible habits. Higher numbers in this index indicate environmental awareness.

Table 2: Balance across schools

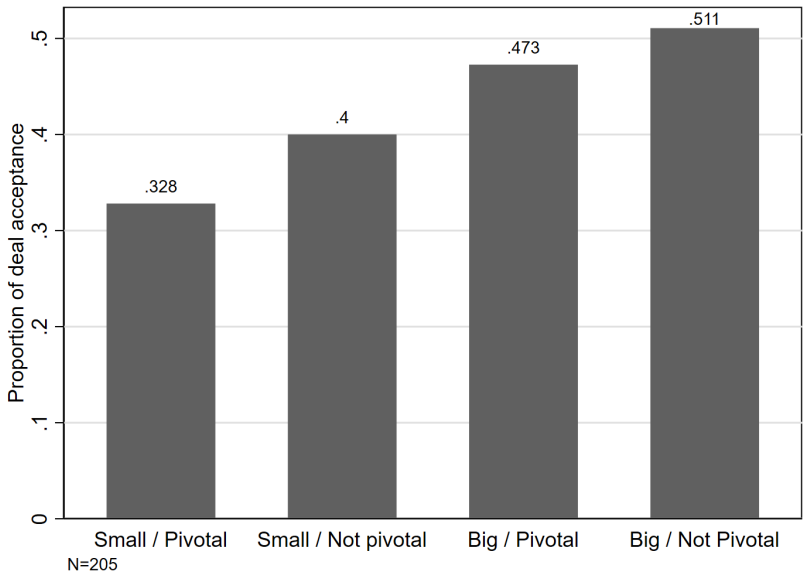
	CIDE (N=97)	**TEC (N=109)**	**t-test difference**
	Mean/SE	**Mean/SE**	
Age	24.309 [0.608]	22.505 [0.461]	1.805**
Gender	0.402 [0.050]	0.606 [0.047]	-0.203***
Family Size	2.031 [0.160]	2.514 [0.143]	-0.483**
Financial Aid	0.825 [0.039]	0.367 [0.046]	0.458***
Ecofriendly Index	0.392 [0.0]	0.372 [0.143]	0.020

The value displayed for t-tests are the differences in the means across groups. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

6 Results

As predicted, the small and pivotal conditions are related to a lesser acceptance of the deal (Graph 1). However, only the difference between the SP and the BNP group is statistically significant. This result suggests that individuals in bigger groups, and when they do not feel crucial for the decision, have a higher acceptance for generating social damage. We argue that the lack of statistical significance could be due to the small incentive in the experiment.

Graph 1. Proportion of deal acceptance by group



In Table 3 we present the results of regressing the dichotomic variable of acceptance of the deal (*Accept*) on the conditions (*Small & Pivotal*). In columns (1) to (3), and (4) to (6) we present the regression with and without the interaction of *Small & Pivotal*, this is the term *Small * Pivotal*. In columns (2) and (5) we add a set of controls relative to personal beliefs and the moment when participants completed the experiment (*Controls1*). In columns (3) and (6) we add socioeconomic controls (*Controls2*). Notice that the intercept, α_0, is the proportion of acceptance for the *Big and Not Pivotal* (BNP) group.

$$Accept = \alpha_0 + \beta_1 Small + \beta_2 Pivotal + \gamma_1 Controls1 + \gamma_2 Controls2 + \varepsilon$$

We find a more precise result of what we observed in Graph 1 by comparing the acceptance rates across groups (Table 3). We see that the *Small* condition has a significant effect ($p < 0.1$) on the acceptance rate. Moreover, the *Pivotal* condition affects in the expected direction, though it is not statistically significant. Finally, adding controls does not change the effects. In the next subsection, we will explain that the controls are not related to the proportion of acceptance; furthermore, we balanced them across groups.

In contrast, the university of the participant is related to her acceptance. A significantly higher proportion of CIDE students accepted the deal. We have explained that students at both schools have different characteristics. However, we cannot discern if

the difference is because of those characteristics or if the coronavirus crisis had a more significant role. Further, when we add socio-economic variables as controls, we observe that the school's effect is no longer statistically significant. In this case, the effect of *Gender* is significant, suggesting that females at TEC rejected the deal more than other subgroups.

Finally, adding an interaction variable ($Interaction = Small * Pivotal$) eliminates the statistical significance of the *Small* condition. We prefer to lean towards the interpretation of the model without the interaction. The reason is that its coefficient is not significant, suggesting that the effect of one condition does not vary at different levels of the other.

6.1 Additional Mechanisms

We added five questions to explore other possible mechanisms for the acceptance of the deal. Three of these additional mechanisms are related to the effect of being pivotal. The remaining two explore if predicted guilt or pro-environmental preferences could change the likelihood of an individual's acceptance. As noted in Table 4, none of these mechanisms have a statistically significant correlation with accepting the deal. Nevertheless, from this analysis, we conclude useful lessons for future research.

From the additional mechanisms that explored the effect of pivotality, we observe that the priming effect was not enough to change their subjective probability of being pivotal. Also, we do not find evidence that the proposed mechanisms explain the effect. However, we observe hypothetical situations in which participants think that being or not pivotal changes their moral responsibility. We presume that in an experiment with more robust priming of pivotality, we could find a more visible effect.

In the first additional mechanism, we asked participants to predict the percentage of people that answered the experiment before them. The rationale is that a higher expectance could set a firmer belief that we reached the necessary threshold. This prior makes the experimental variation of the primes (*Pivotal, Non-Pivotal*) less believable and reduces the moral cost of accepting. Also, since participants could answer the experiment over one week, it would be reasonable for people who answered on the last days to have

Table 3: Regression results

	Without interaction			**With interaction**		
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
	Accept	**Accept**	**Accept**	**Accept**	**Accept**	**Accept**
Small	-0.130*	-0.117*	-0.14*	-0.106	-0.086	-0.143
	[0.070]	[0.069]	[0.074]	[0.110]	[0.109]	[0.116]
Pivotal	-0.068	-0.07	-0.002	-0.049	-0.051	0.000
	[0.071]	[0.070]	[0.075]	[0.101]	[0.099]	[0.103]
People before		0.003	0.002		0.003	0.002
		[0.002]	[0.002]		[0.002]	[0.002]
Pr. No Donation		-0.001	-0.002		-0.001	-0.002
		[0.001]	[0.001]		[0.001]	[0.001]
Days		-0.014	-0.028		-0.014	-0.028
		[0.021]	[0.023]		[0.021]	[0.001]
Tec Student		-0.217***	-0.101		-0.216***	-0.101
		[0.070]	[0.081]		[0.070]	[0.081]
Interaction				-0.039	-0.052	0.004
				[0.143]	[0.141]	[0.149]
Age			-0.051			0.051
			[0.075]			[0.075]
Age Squared			-0.001			-0.001
			[0.001]			[0.001]
Gender			-0.259***			-0.259***
			[0.078]			[0.078]
Family Size			0.007			0.007
			[0.026]			[0.026]
Financial aid			0.259			0.26***
			[0.091]			[0.091]
Bachelor			0.193			0.191
			[0.173]			[0.173]
Ecofriendly index			-0.349			-0.349
			[0.361]			[0.362]
_cons	0.522***	0.662***	-0.23***	0.512	0.651***	-0.229
	[0.064]	[0.115]	[0.260]	[0.075]	[0.119]	[1.173]
N	197	197	167	197	197	168

a higher prior.

Table 4: Correlation between accepting and the additional variables

	Accepted
People Before	0.066
Pr. No Donation	-0.006
H. Sit. 1 (Pivotal)	0.082
H. Sit. 2 (Size)	-0.053
Ecofriendly index	-0.358

We incentivized this first mechanism. We told participants that if the difference between their prediction and the real value were less than 5%, we would add MXN 50 to their potential payment. Graph 2 in Appendix 2 shows that participants' prediction and the actual value of people who answered before are positively correlated. The results show that 41 participants (19.9%) predicted within 5% from the correct answer. However, the acceptance rate and their prediction do not correlate (Table 4). We conclude that the participants did not update their belief about reaching the threshold by their predictions.

In the second mechanism, we asked them to estimate the probability that the threshold would be reached and, therefore, the probability that we would not donate (Graph 3, Appendix 2). We expected to have an estimate of one, or at least near it. However, the results show that people's estimate is below this amount. It is worth noticing that the moral cost for a person is lower when she thinks: 1) her action will not cause a damaging effect; and 2) the damage will happen nonetheless. We observe that there is not a statistically significant difference between the estimation of those who accepted the deal and those who rejected it (Diff. 0.517, SE.3.808). Furthermore, the distribution of the estimated probabilities of no donation for those who accepted the deal has a higher median.

Similar to Graph 3, Appendix 2, if subjects completely believed the experimental prime in the Pivotal condition, we would expect that every individual in that condition who accepted the deal to estimate the probability of no donation as one. However, we do not observe this (Graph 4, Appendix 2). 91% of participants in that situation estimated

lower probabilities than one. Their mean probability is the same as those in the non-pivotal condition. We conclude that the prime failed the credibility test and suppose that the observed null-effect could change with a more precise and concrete pivotal situation.

The third mechanism asked participants about their feelings of guiltiness concerning the deal. We asked participants who accepted the deal to report how guilty they felt. In contrast, we asked those who did not accept to predict the feelings of guiltiness from those who accepted. We do not find differences in their answers, implying that their predictions are correct (Graph 5, Appendix 2). We suggest that those who did not accept the deal are more sensitive to the feeling of guiltiness. However, since this question was immediately after the experiment, the answers from those who accepted the deal could be biased in a self-serving way.

The fourth mechanism asked participants two questions that presented two hypothetical scenarios, similar scenarios to the one in the experiment. Participants had to decide if two fictional characters were equally or unequally responsible for generating social damage. In Hypothetical Situation 1, one fictional character was pivotal, and the other was not. In Hypothetical Situation 2, one character was part of a big group, while the other was part of a small one. We wanted to elicit their moral views and to answer if they are related to the acceptance rate.

Graph 6 in Appendix 2 presents the proportion of individuals who believe that the fictional characters are not equally responsible. In other words, the proportion of participants who believe that not being pivotal or being part of a big group, dilute their responsibility. We do not find statistically significant differences between the participants who accepted the deal and those who did not.

We observe a significant difference between the opinions in both situations. A higher proportion of participants agree that being pivotal changes moral responsibility more than the size of the group, a finding that contradicts our results. In the experiment, the size of the group was more important. But we also found that the pivotal condition was not credible. We expect that in further studies, we will find an effect with a more robust pivotal condition.

Finally, in the fifth mechanism, we constructed an (*Ecofriendly Index*) with six survey questions related to environmentally responsible habits[4]. It is reasonable to expect that, since the social damage in the experiment is related to the environment, the people with pro-environmental preferences would reject the deal more often. However, we do not find evidence of this.

7 Discussion and policy recommendations

We found evidence that the group's size needed to generate social damage is relevant for the acceptance of the deal. Our results also suggest that being pivotal may affect the deal's acceptability. We did not find other mechanisms related to our outcome. The primary explanatory variable for the acceptance of the generation of social damage was the school of the participant. We argue that these results respond to the small incentives in the experiment and the lack of statistical power due to the relatively low participation of the students reached by e-mail. However, our findings make us expect that we could see more definite results in a controlled setting with higher incentives and sufficient participants.

The main policy recommendation we propose is, despite being non-pivotal, making people believe that everything is not lost and make them think they are pivotal could decrease social damage. Additionally, and in line with the pivotality belief, telling individuals that their help is needed remains a piece of sound advice. Remember that researchers have already answered that even infants need to feel responsible if we want them to help. The work that Greta Thunberg is doing in this regard, from our results and the literature agreement, seems vital.

Having the belief that we are pivotal in a particular setting is not enough. Our role must be clear too. We recommend making us believe we are relevant and responsible within a clear role. It will help us set ourselves in the problem the society wants to solve, see our share of responsibility, and diffuse no more. In this regard, the individualization

[4]We built the index through a Likert-scale of agreeableness over six statements. The statements were: I am a consumer who avoids wasting resources; I avoid wasting resources even if abundant; I avoid worrying about material goods; I usually lead an environmentally responsible lifestyle; I have environmentally friendly eating habits; My daily routines are friendly to the environment.

of responsibility seems essential. One way to think about this is to make individuals see the local damage and to act locally for global purposes.

We recommend that any intervention is made clear enough (responsibility-wise) and easy enough for individuals to pursue. The latter is to avoid an excess burden of responsibility that people might feel. Additionally, we urge authorities to create interventions that provide feedback on local actions that assist those involved in perceiving themselves relevant.

Diffusion of responsibility is a social phenomenon that can explain why we destroy the environment. It is a harmful phenomenon for the environment and our future as a species. We need tools to offset its consequences. As we see, one of our best answers is the illusion of pivotality and how individuals position themselves in a problem with global magnitudes. Perhaps we should all feel as responsible and as pivotal as Greta in order to do something. More research on this regard is needed, however.

Appendix 1: Discussion on deception

We define deception in experiments as an act or statement intended to make subjects believe something that is not true (Rousu et al., 2015), for example, the use of confederates, changing the rules of the experiment without informing participants or promising inexistent payments. In disciplines such as Psychology, deception is a common practice to increase the experimental control of researchers (Hertwig and Ortmann, 2008). However, there is an implicit ban on the use of deception in Experimental Economics (Rousu et al., 2015).

The central argument used to defend the ban on the use of deception is that it could lead to problems of identification in future experiments (Colson et al., 2016). In Economics, we assume that participants react to credible incentives. If, because of previous experiences, participants think that some statements in the experiment instructions might be false, their actions could be changed unintendedly. Thus, the use of deception generates a public-good problem: while it might be useful in a particular experiment, it puts the design of all the others at risk.

We do not challenge the notion that a majority of Economic researchers hold that deception is not acceptable in experiments. However, we observe that our experimental design raises the question of whether it implies the use of deception. As an answer, we defend that the design does not imply use deception in any reasonable sense. Here, we offer two arguments for this. We observe two potential points of concern:

1. We told some participants that we were offering the deal to 100 individuals, and to others, we told them that we were offering it to 400.

2. We told participants a statement based on information we do not possess at the moment. The statement we tell participants is: "*X participants may have already accepted.*"

The first argument for defending we did not deceive participants is that both statements are true in their meaning. We can argue that we sent the deal to more than 900 participants for the first point of concern. The amount of 900 is enough to form two

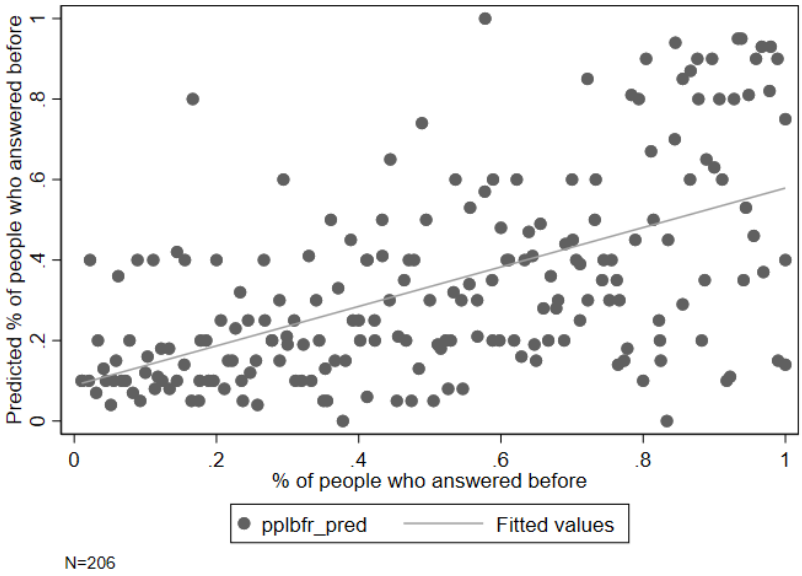
groups of 400 and one of 100. Also, we were offering the deal to more than 400 participants; thus, it follows that we were offering it to 100, 35, or any other number smaller or equal to 400. Therefore, it is not a false statement. In contrast, it would have been a false statement to say, "*We are offering this deal only to 100 participants*". We did not say that.

For the second point of concern, we emphasize that it is a statement about a possibility and not a fact. Notice: the true value of a statement does not depend on the information of the person declaring it but in its content. As a consequence, to qualify as true or false the statement, it does not matter if we had information or not about how many people had already accepted the deal. It only matters if it was indeed true that it was a possibility that some people had already accepted. Since every participant could answer at any moment, it was indeed a possibility. Therefore, it is not a false statement. Hence, we did not deceive participants.

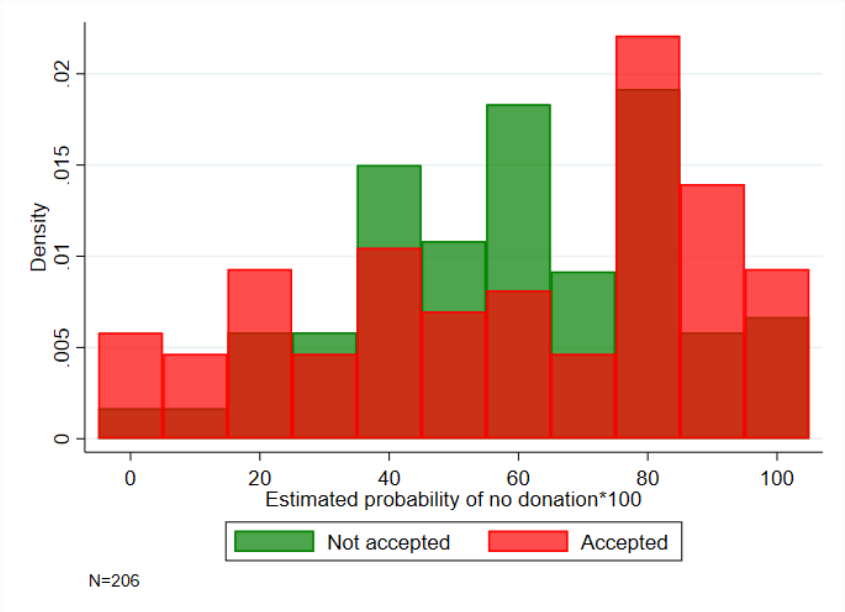
The second argument is that the statements were true in its consequences. In other words, participants would have no basis to feel deceived since the consequences of what we told them was true. Therefore, their experience in the experiment does not affect their behavior in future studies. Even if conceding that the statements were misleading because they tried to make people believe something that could be false, the consequences of those statements were true. In other words, when choosing to donate for all, we used the highest possible limit (25). So, even when we told some people that the limit was 15, we were willing to donate after more acceptances. We did it to achieve that no participant could feel deceived. Thus, since the statements were true in its consequences, the design does not imply deception.

Appendix 2: Graphs

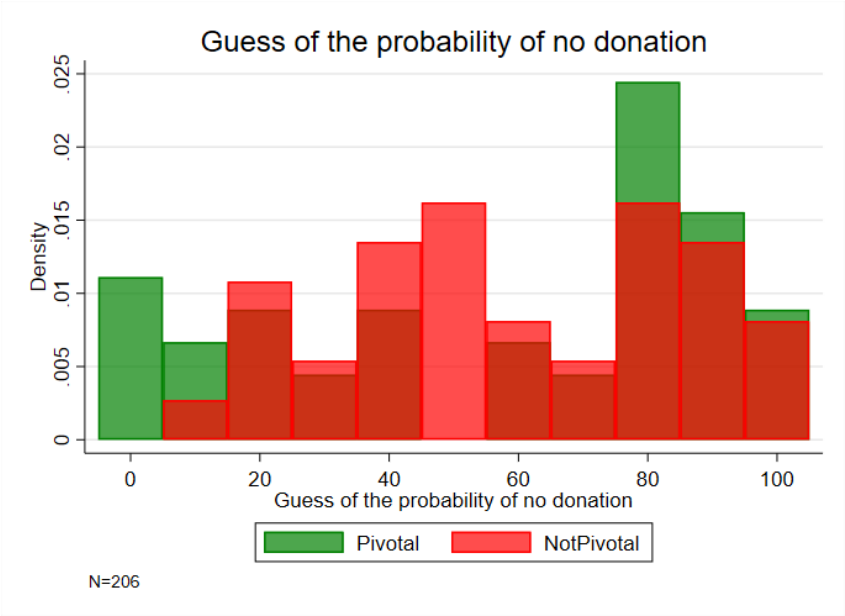
Graph 2. Predicted and actual % of people before



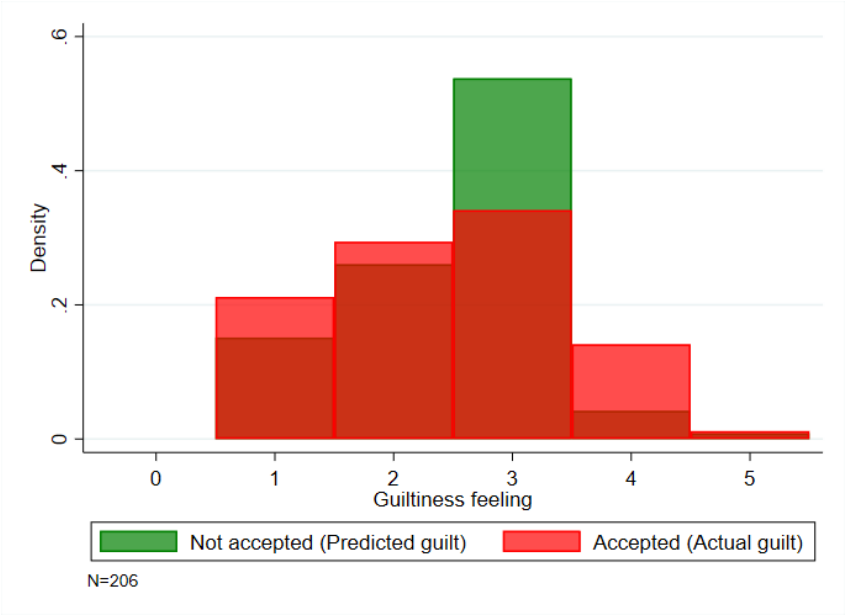
Graph 3. Estimated probability of no donation



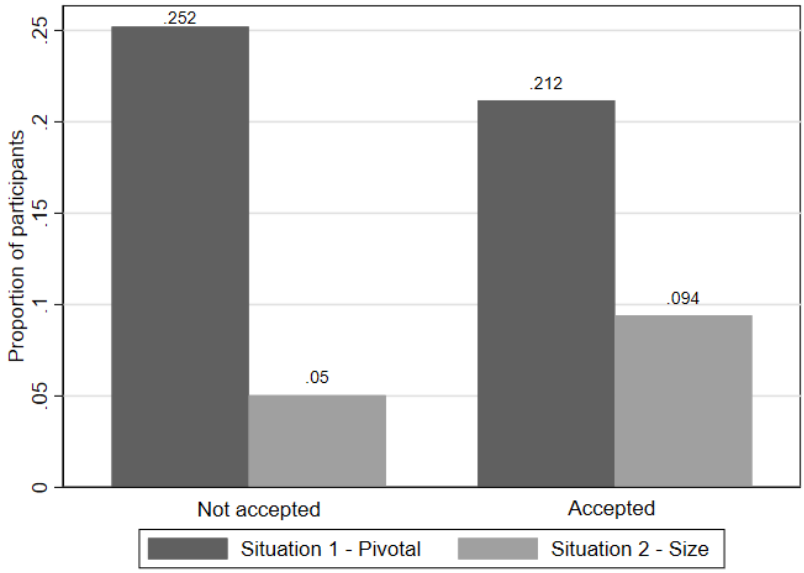
Graph 4. Estimated probability of no donation pivotal/not pivotal



Graph 5. Predicted and actual feelings of guilt



Graph 6. Hypothetical situations



Appendix 3: Theoretical model

7.1 Full information scenario

In full information, sequential games, players know how many have accepted the deal and how many are left to answer. It is possible to assume that payments are made after the decisions of all players. Therefore, they know if they are pivotal or not. Thus, the condition for acceptance for the pivotal individuals are (3), (4), and (5). It is easy to see that since we have assumed $d > 0$, and we know that $\pi \in (0, 1)$:

$$2d + a(s, d) > d + a(s, d) > \pi[a(s, d) + d]$$

The restriction for the pivotal players is stricter than for the non-pivotal *Type b*. At the same time, this is more restrictive than for the non-pivotal *Type a*. With this in mind, we can state the two central hypotheses of our study:

***Diffusion of responsibility* hypothesis.** $a(s, d)$ is decreasing for s. It means that the individual receives a lower share of the moral cost of being a part of the guilty group whenever the number of people in that group increases. Therefore, the necessary utility she needs in return for accepting the deal is lower.

***Diffusion of being pivotal* hypothesis.** Pivotal subjects will be more reluctant to accept the deal or will need a higher reward of u, than non-pivotal individuals. Similarly, Non-pivotal individuals type *Type b* will accept less the deal than the ones with *Type a*.

As an illustration, observe that in the simplest case, where we think that all s members of the guilty group share the responsibility equally, we could write $a(s, d) = \frac{d}{s}$ which is in line with the hypothesis. There, from (3), (4), and (5) we get, respectively:

$$u \geqslant d[\frac{2s + 1}{s}] \tag{6}$$

$$u \geqslant \pi d[\frac{s+1}{s}] \tag{7}$$

$$u \geqslant d[\frac{s+1}{s}] \tag{8}$$

Since $s \geqslant 1$, if $u \geqslant 3d$, all players will accept the deal and the social damage will be generated. If $2\pi d \leqslant u < 3d$, only $s-1$ will accept and the social damage will not be generated. And if $u < 2\pi d$ no one will accept the deal. If $s = 1$, then:

$$u \geqslant 3d \geqslant 2d \geqslant 2\pi d$$

It implies that $3d$ is whenever the participant accepts, the damage will be done, $2d$, the damage is already done, whereas $2\pi d$, the damage is not done yet and the participant despite being non-pivotal, could think she is. It is remarkable how π plays a role in helping the participant believe she could be pivotal.

7.2 Partial information scenario

Imagine a situation with incomplete information in which the deal is offered sequentially to players. They do not know how many players have answered before, nor how many accepted. The players have uncertainty about whether they are pivotal or not. Then, suppose two things: 1) the player receives the payments after everyone decides, and 2) the information of accepting/rejecting is available to everyone, but not the order of decisions. Finally, suppose that players have an intrinsic aversion to being pivotal. Thus, if P is the estimated probability that she is pivotal, the player will accept the deal if:

$$P[u - 2d - a(s,d)] + (1-P)[u - \pi[a(s,d) + d]] \geqslant 0 \tag{9}$$

Notice that (9) is the weighted sum of two mutually exclusive events. The first is when the individual is pivotal and gets a utility of $u - 2d - a(s,d)$. The second is not pivotal and gets an utility of $u - \pi[a(s,d) + d]$. It is relevant to mention that, in this case, it is

not possible to distinguish between *Type a* and *Type b* non-pivotality. The reason is that the player cannot know if the damage was already done or not, so her utility in the not pivotal case depends on her belief of the probability that s players will accept (π).

An alternative way to understand this is to imagine that the game is not sequential but simultaneous[5]. All players decide at once, but know that they are going to count their decisions sequentially. As a voting scenario, for example. In this case, players know that someone will be pivotal, but they do not know who will be.

From (9) we make the necessary algebraical manipulations to get:

$$P \leqslant \frac{u - \pi[a(s,d) + d]}{2d + a(s,d) - \pi[a(s,d) + d]} \tag{10}$$

Recall that P is a probability, so $P \in [0,1]$. However, if $P = 0$ we are back into the non-pivotal situation (*Type a*), from the previous section. Thus, the condition for acceptance is that $u \geqslant \pi[a(s,d) + d]$. Besides, if $P = 1$, then (10) is the pivotal scenario, from the previous section.

We are interested, however, in the case where $0 < P < 1$. If this is true, then we can directly obtain from (10):

$$u \geqslant P[2d + a(s,d)] + (1 - P)[\pi[a(s,d) + d]] \tag{11}$$

The interpretation of (11) is simple. It states that the payment for accepting the deal should be bigger than the expected costs of accepting. Notice that from (11) we can make two predictions based on our hypothesis:

***Diffusion of responsibility* prediction:** in a condition with a larger necessary group, more people will accept the deal.

***Diffusion of being pivotal* prediction:** participants primed with a pivotal intervention will be more reluctant to accept the deal.

[5]This is not adding assumption. In this case, there is not full information about the actions of other players; the sequential game is equivalent to a simultaneous one.

7.3 Comparison with Falk, A. & Szech, N. (2017)

Observe that equation (10) has an equivalent interpretation to the *Belief-dependent dichotomy* in the article by Falk, A. & Szech, N. (2017). As in that article, it is also true in our model that for a small enough probability of being pivotal (expressed by P here and by $p(8)$ in that article) opting for the morally problematic option (this is, accept the deal) becomes more attractive. Thus, our model also generates the central hypothesis in that study: increasing the belief of being pivotal will reduce the proportion of people opting for the morally problematic option.

However, our model has three differential characteristics:

1. Our model is explicitly about a *public bad*. In other words, social damage that, if generated, subtracts d from the utility of every player, no matter their individual choices. In the article by Falk, A. & Szech, N. (2017), this is not the main focus. Moreover, they do not consider if participants in their experiments lose utility whenever they kill mice, even when they rejected the deal.

2. Our model integrates the effects of being pivotal and the minimum size of the guilty group (s). Thus, our model is useful in making predictions about it. In the article by Falk, A. & Szech, N. (2017), they only consider the case where $s = 1$.

3. Our model explicitly models the belief about the actions of other participants: π.

7.4 Out of scope elements in our model

This model has the objective to explain the main intuition of our hypothesis, explain our experimental design, and interpret our findings. However, it has at least two elements that are out of scope:

1. We are assuming perfectly homogeneous individuals in three dimensions.

   (a) We are assuming that all individuals receive the same reduction from the *public bad*. In other words, that they value equally the *public good*, that was destroyed. This is not necessarily the case.

(b) We are assuming that all individuals have the same belief about the actions of others (π).

(c) We are assuming that all individuals have the same belief about the probability of being pivotal (P).

The analysis using heterogeneity in these dimensions would imply different conditions for acceptance and more possible solutions in both scenarios (full and partial information).

2. In our *Diffusion of responsibility hypothesis*, we are assuming that individuals care about the absolute number of people in the guilty group. It is possible to assume instead that they care about the proportion of the guilty group concerning the total number of participants, $\frac{s}{t}$. It is an empirical question in itself to determine one or the other. It would be more natural to think that ten individuals who accepted the deal feel less guilty in a population of 20 than in a population of 100.

   By taking this into account, our main predictions would not change, but we would be able to add predictions about different levels of t. We would predict that for a more substantial proportion of $\frac{s}{t}$, more people would accept the deal, and we would also be able to make predictions about the effects of changes in t.

References

Bandura, A., Underwood, B., and Fromson, M. E. (1975). Disinhibition of aggression through diffusion of responsibility and dehumanization of victims. *Journal of Research in Personality*, 9(4):253–269.

Barron, G. and Yechiam, E. (2002). Private e-mail requests and the diffusion of responsibility. *Computers in Human Behavior*, 18(5):507–520.

Bartling, B. and Fischbacher, U. (2012). Shifting the blame: On delegation and responsibility. *Review of Economic Studies*, 79(1):67–87.

Bickman, L. (1972). Social influence and diffusion of responsibility in an emergency. *Journal of Experimental Social Psychology*, 8(5):438–445.

Coffman, L. C. (2011). Intermediation Reduces Punishment (and Reward). *SSRN Electronic Journal*.

Colson, G., Corrigan, J. R., Grebitus, C., Loureiro, M. L., and Rousu, M. C. (2016). Which Deceptive Practices, If Any, Should Be Allowed in Experimental Economics Research? Results from Surveys of Applied Experimental Economists and Students. *American Journal of Agricultural Economics*, 98(2):610–621.

Corrion, K., Long, T., Smith, A. L., and D'Arripe-Longueville, F. (2009). "It's not my fault; it's not serious": Athlete accounts of moral disengagement in competitive sport. *Sport Psychologist*, 23(3):388–404.

Darley, J. M. and Latané, B. (1968). Bystander Intervention in Emergencies: Diffusion of Responsibility. *Journal of Personality and Social Psychology*, 8(4):377–383.

Falk, A. and Szech, N. (2017). Diffusion of being pivotal and immoral outcomes. *KIT Working Paper Series in Economics*, (No. 111).

Feldman, R. S. and Rosen, F. P. (1978). Diffusion of responsibility in crime, punishment, and other adversity. *Law and Human Behavior*, 2(4):313–322.

Freeman, S., Walker, M. R., Borden, R., and Latané, B. (1957). Diffusion of Responsibility and Restaurant Tipping: Cheaper by the Bunch. pages 584–587.

Guerin, B. (2011). Diffusion of Responsibility. *The Encyclopedia of Peace Psychology*, First Edit.

Hamman, J. R., Loewenstein, G., and Weber, R. A. (2010). Self-interest through delegation: An additional rationale for the principal-agent relationship. *American Economic Review*, 100(4):1826–1846.

Hertwig, R. and Ortmann, A. (2008). Deception in experiments: Revisiting the arguments in its defense. *Ethics and Behavior*, 18(1):59–92.

Hortensius, R. and de Gelder, B. (2018). From Empathy to Apathy: The Bystander Effect Revisited. *Current Directions in Psychological Science*, 27(4):249–256.

Markey, P. M. (2000). Bystander intervention in computer-mediated communication. *Computers in Human Behavior*, 16(2):183–188.

Mynatt, C. and Sherman, S. J. (1975). Responsibility attribution in groups and individuals: A direct test of the diffusion of responsibility hypothesis. *Journal of Personality and Social Psychology*, 32(6):1111–1118.

Oexl, R. and Grossman, Z. J. (2013). Shifting the blame to a powerless intermediary. *Experimental Economics*, 16(3):306–312.

Plötner, M., Over, H., Carpenter, M., and Tomasello, M. (2015). Young Children Show the Bystander Effect in Helping Situations. *Psychological Science*, 26(4):1–8.

Poteete, A. R. and Ostrom, E. (2004). Heterogeneity, Group Size and Collective Action: The Role of Institutions in Forest Management. *Development and Change*, 35(3):435–461.

Ross, A. S. and Braband, J. (1973). Effect of increased responsibility on bystander intervention, II: The cue value of a blind person. *Journal of Personality and Social Psychology*, 25(2):254–258.

Rousu, M. C., Colson, G., Corrigan, J. R., Grebitus, C., and Loureiro, M. L. (2015). Deception in experiments: Towards guidelines on use in applied economics research. *Applied Economic Perspectives and Policy*, 37(3):524–536.