

# Nice guys or potential cooperators when keeping promises: an experimental and Bayesian account for two explanations

*Said Jiménez*

*July 2019*

## 1 Introduction

Human beings are social animals and we get advantages from being so. Historically, belonging to a social group has provided protection and guaranteed access to resources such as food, currently there has been evidence that belonging to a social circle has a positive effect on longevity, physical and mental health (Holt-Lunstad, 2018; Tough, Siegrist, & Fekete, 2017). The interrelation between social closeness and cooperation is key to the formation of large groups of individuals without a genetic relationship, such groups form the basis of communities, societies, and nations, as well as probably constitute one of the most fundamental conditions for human survival (Fehr & Fischbacher, 2003; Fehr & Schurtenberger, 2018).

An important requirement for cooperation between individuals is communication, a meta-analysis of 45 studies reported a large positive effect ( $d = 1.01$ ) of communication in cooperation regardless of the communication medium (Balliet, 2010). In parallel, it is known that after communication sessions between completely unknown individuals there is an increase in subjective social closeness (Aron, Melinat, Aron, Vallone, & Bator, 1997). Thus, the findings suggest a positive relationship between communication, cooperation, and social closeness.

One of the forms of communication that have received the most attention in psychological game theory is the promise, in an interaction between sender and recipient, it is believed that promises influence beliefs of the recipient, generating trust and cooperation (Charness & Dufwenberg, 2006; Vanberg, 2008). However, there are situations where trust is betrayed, promises are broken or people deceive, for example, approximately 11% of people surveyed by the General Social Survey (GSS) in 2018 responded that they have had sex with someone different to their partner while they are married (Smith, Davern, Freese, & Hout, 2018). Another example, according to the Washington Post, President Trump has said 10,796 false or misleading statements in 869 days in office (Kessler, Rizzo, & Kelly, 2019).

In the laboratory, these transgressions to trust have been explicitly studied in experiments where subjects have incentives to lie or not keep their promises. However, the vast majority of these studies in economics, psychology and neurosciences have been conducted in people who do NOT know each other (to name a few, Gneezy, 2005; Baumgartner, Fischbacher, Feierabend, Lutz, & Fehr, 2009; Baumgartner, Gianotti, & Knoch, 2013; Charness & Dufwenberg, 2006;

Fischbacher & Föllmi-Heusi, 2013; Gneezy, Rockenbach, & Serra-Garcia, 2013; Mazar, Amir, & Ariely, 2008). In social psychology some studies address deception in close interpersonal relationships, however, they use self-report measurements (DePaulo, Ansfield, Kirkendol, & Boden, 2004; DePaulo & Kashy, 1998) or rather explore the development of deception detection skills in same-sex friends (Anderson, DePaulo, & Ansfield, 2002). Despite the widespread presence of interactions between human beings belonging to the same group, it is not known how social closeness between subjects can affect keeping or breaking promises.

In this manuscript, we explore the effects of three partners with different levels of social closeness on keeping or breaking promises, as well as on cooperation in a trust game. The above, we analyze it in the light of the two main motivations that have been pointed out in the literature regarding keeping promises (Baumgartner et al., 2009):

1. The *instrumental* suggests that promises are kept to make future cooperation easier.
2. The *intrinsic* mentions that the promises are kept to do what is morally right.

In our study, subjects perform a standard trust game in pairs with three phases: *first*, the trustee makes a promise to pay half of his earnings regardless of who his investor is; *second*, the investor receives the promise and decides if he invests his initial budget; *third*, in case the investor has given the budget, the trustee faces the decision to pay or not to pay half of his earnings. To evaluate the effect of social closeness, our subjects participate in the role of trustee in front of three partners with different levels of social closeness (zero, low and high) in the role of the investor: a *computer*, a *stranger*, and a *friend*.

The manipulations mentioned allow us to evaluate several hypotheses, the first one was proposed a priori and is derived from a larger research project (you can check the pre-registration of the hypothesis here: <https://osf.io/u97fd>), while the following are exploratory:

1. Social closeness will reduce the decision of breaking the promise. According to the *instrumental* motivation, we expect that subjects keep their promises with friends for the purpose of facilitating future cooperation, which can be extended even beyond the trials in the experiment. We also expect, although to a lesser extent, that the subjects keep promises to strangers, with the purpose of facilitating cooperation at least during the trials during the experiment. Finally, we anticipate that the participants break the promises to the computer because it is a partner without social closeness and they could not ensure cooperation in future trials. It should be noted that if the participants keep the promises to the computer, evidence would be given in favor of intrinsic motivation.
2. There will be an effect of social closeness on cooperation, regardless of promises. We expect that there is more probability of paying the *friend* than the *stranger* and more probability of paying the stranger than the *computer*. According to the *instrumental* motivation, the cooperation will be greater for partners with whom the subject anticipates greater cooperation in the future (friend > stranger > computer).
3. Finally, we will obtain two subsamples according to the cooperation rate of the subjects as was done in another study (Baumgartner et al., 2009). However, what in the aforementioned study was classified as a group of *honest* and *dishonest*, we will show

that in our sample it is not supported because if the groups differ in their payment rate it will be due to the difference in their level of commitment expressed by promises.

## 2 Methods

### 2.1 Subjects

We included 45 subjects (15 men), recruited from the National Autonomous University of Mexico, the age range from 19 to 33 years and their minimum educational level were bachelor’s degrees. Subjects went to the study with a friend considered *close* by themselves, who fulfilled the following characteristics: he was matched by sex, did not have a family bond and was not a person with a sentimental or sexual relationship. 30 subjects (15 men) performed the task within a Magnetic Resonance Imaging (MRI) scanner, however, their image data is not analyzed in the present work.

### 2.2 Task

All subjects performed an adaptation of the trust game with promises, using hypothetical monetary rewards (Baumgartner et al., 2009). The task was programmed in PsychoPy2 version 1.84.2 (Peirce et al., 2019; Peirce, 2008) and consisted of 24 trials between two players: trustee and investor. The trustee originally has 0 Mexican pesos and the investor has 2 pesos, the investor is presented with the opportunity to give his money to the trustee or keep it. If he invests his money it is multiplied by 5 pesos, so that the trustee has 10 pesos. Finally, the trustee decides to pay half to the investor or keep the 10 pesos.

The structure described is repeated in 24 trials, however, in 4 of the trials the trustee can send a promise to the investor, the promises were that *always*, *mostly*, *sometimes*, or *never* would pay back. Each promise was valid for three trials so that 12 of the trials have the effect of the promise and the other 12 do not. It is worth mentioning that the promises are not made to a particular partner, but the subject expresses their commitment level to pay back regardless of who their partners are in the next three trials in which the promise is valid.

The variation with respect to the original task is that in our experiment three partners were presented with three levels of social closeness in the role of the investor: *computer* (no closeness), *strange* (low closeness) and *friend* (high closeness), while the subject acts as the trustee. Each partner performs the role of investor in 8 of 24 total trials, however, both the promises and social closeness conditions were presented to the trustee in pseudorandom order.

Since our main interest was the trustee’s behavior, investors’ decisions were programmed *a priori* to give their amount in 6 trials and in 2 did not. The covert story for all our subjects was that they would be playing in real-time with their friend, the stranger (was told he would be another same-sex unknown person) and the computer. A diagram of the chronology of the experimental task with the duration of each phase in seconds is shown in Figure 1, each box from left to right represents a screen that was shown to the subjects sequentially. Section A<sub>1</sub> corresponds to an example of trials without promises and section A<sub>2</sub> to trials with promises.

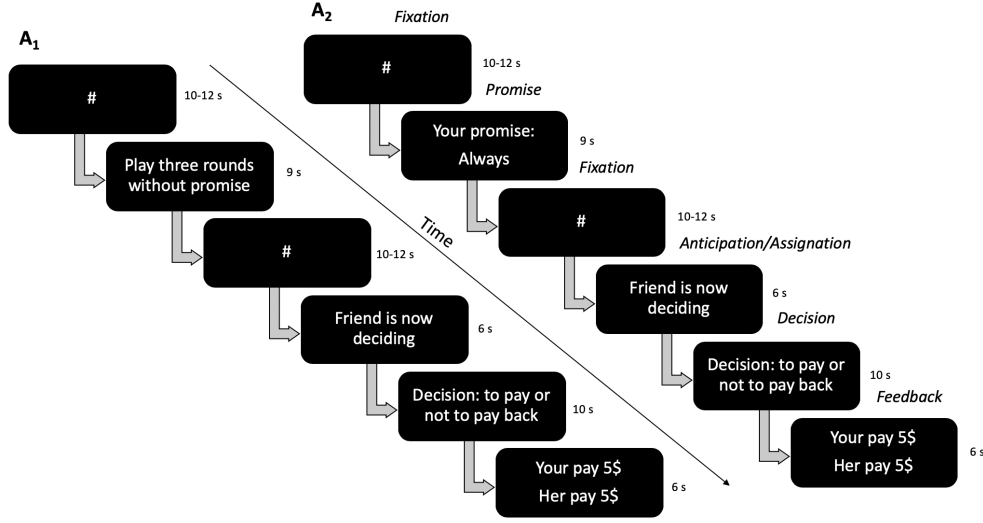


Figure 1: Trust game with promises

Fixation phase consisted only of a period in which the subject paid attention without performing any particular behavior, then, **promises phase** in part A<sub>1</sub> indicated to the subject that the following three trials they could decide without the effect of the promise, while in A<sub>2</sub> the subject was asked to decide between *always*, *mostly*, *sometimes* or *never* pay back. This was followed by another fixation period, subsequently, trials in both A<sub>1</sub> and A<sub>2</sub> continue in the same way.

In the **anticipation/assignment phase**, subjects were told who their partner was for that trial (computer, stranger or friend) and was given the message that their partner was making their decision. Later in the **decision phase**, the subject was informed if his partner invested his \$2 or not, he was reminded of his promise level (if they were trials with promises) and, in case of his partner had invested, he was asked to decide whether to pay back or not. Finally, in **feedback phase** payments for that trial were shown and the sequence was repeated.

## 2.3 Procedure

Subjects came with a same-sex friend considered close by him or her, it was emphasized that they must not have a romantic or family relationship with their friends, to try to exclude the effect on cooperation due to a consanguineous relationship or sexual attraction. Also, to ensure that the subjects and their companions had a similar degree of social closeness to each other, in the laboratory both responded to the "Inclusion of the Other in Self" IOS scale (Aron et al., 1997) without observing their partner responses. The scale consists of seven pairs of circles that vary in the degree of overlap between them, the respondent must select the pair of circles that best represents the subjective closeness to his partner.

Subjects were trained in the task by two of the researchers, the game was first explained verbally, and then the computerized interface was presented to them on a portable computer. They were told that they would make their decisions with four buttons on the computer keyboard, which corresponded to the levels of promises and with only two of these to make

their decision to pay back or not. Friends heard the explanation, they were told that they would make their decisions in a separate room on another computer. Subjects performed between 3 and 6 practice tests together with the researcher who exemplified the course of the game when the investor gave his budget. In the practice trials, all the phases indicated in Figure 1 were shown, however, the identity of the investors was replaced by question marks.

When the subjects did not report doubts we proceeded to take them without their companions to another room where they would make their decisions on a laptop, or, with the Lumina Cedrus response system in the case that it was a subject to which we also performed MRI. The subjects began the task believing that they would really play with the three partners of different social closeness. In the other room, another of the researchers performed the *debriefing* to the companion, they were told that the purpose of the research was to know if their mere presence had an effect on the subject’s decisions, but that to simplify the analysis their responses were programmed *a priori* so your participation ended there.

When the subject finished the task, *debriefing* was also carried out, he was told that the decisions of his partners had been programmed. Although no participant exercised their right, both subjects and their friends were told that they could withdraw their participation and informed consent if they considered disagreeing with any of the manipulations made by researchers.

## 3 Results

### 3.1 Descriptives

We analyzed 810 decisions about paying back or not (18 per subject), as well as 180 promises (4 per subject), 69% of the decisions, were pay back, while the proportion of the promises selected were 5.4%, 16.2%, 43.8% and 40.2%, respectively for promises *never*, *sometimes*, *mostly* and *always*. Figure 2 shows the payment proportion depending on the levels of promises and partners.

Breaking promises, that is, subject had promised *always* and subsequently decides not to pay back occurred only 12.4% of trials (19 times they did not pay back of 153 when promised *always*), of which, 8.50% (13/153) were trials with the computer, 3.92% (6/153) with the stranger and 0% (0/153) with the friend. It is essential to note that breaking the promise was not the most frequent decision, in 91.5% of trials when they promised *always* and had the computer as a partner they decided to pay back.

### 3.2 Hierarchical Bayesian Modeling

All models presented below were programmed in R via the **brms** package (Bürkner, 2017; R Core Team, 2019), which performs the inference using sampling by Markov chain Monte Carlo through **Stan** (Stan Development Team, 2018). For each model, the posteriors distributions of all parameters were approximated with four chains of 2000 iterations each, the first 1000 iterations of each chain were discarded (burning period), for a total of 4000 post-burning samples. Models convergence was evaluated through visual inspection of the chains

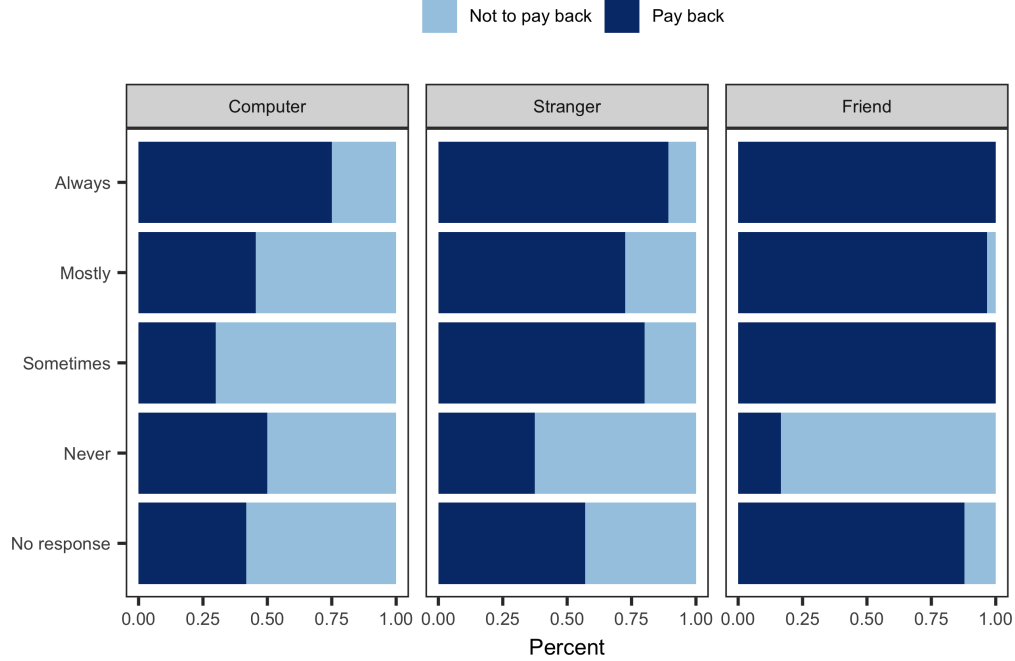


Figure 2: Pay back proportion by partners and promises levels

and calculation of the  $\hat{R}$  statistic, which for all parameters was 1, which is interpreted as convergence. Vaguely informative prior distributions were used for the parameters of interest in the models, which allows the data to dominate the inference, also assumes as little as possible regarding the nature of the phenomenon, which could be adequate for the current state of evidence in the problem we are studying (McElreath, 2018).

### 3.2.1 Social closeness between partners

To assess social closeness, 30 (15 men) of our subjects and their friends responded to the IOS scale (Aron et al., 1997; Fareri & Delgado, 2014; Sip, Smith, Porcelli, Kar, & Delgado, 2015), which consist of seven pairs of circles that vary in the degree of overlap and represent the social closeness that an individual perceives with respect to the other. A Hierarchical and Cumulative Bayesian Model (See section: Commitment expressed in the promises), supports the hypothesis that there are no differences in subjective social closeness between subjects and their friends, with a posterior evidence ratio of 2.74 in its favor.

### 3.2.2 Effect of social closeness on breaking promises

To evaluate the hypothesis that social closeness would reduce the behavior of breaking the promise, we filtered all trials in which subjects promised that they would *always* pay back and then decided not to pay. Subsequently, Bayesian inference was used to assess the effect of partners on the decision to break the promise at the individual and population level. For this purpose, a Hierarchical Bayesian Model was carried out that assumes that the uncertainty in partners effect on the decision to break the promise varies depending on each individual, however, it also assumes that these variations belong to common population distributions

(Gelman & Hill, 2006).

$$y_i \sim \text{Bernoulli}(\theta_i)$$

$$\text{logit}(\theta_i) = \mathbf{X}\beta + \mathbf{Z}u$$

In the previous model, the decision to break the promise  $y_i$  comes from a Bernoulli distribution with probability  $\theta_i$ , the goal of the Hierarchical Model is to predict each decision through the linear combination of the effects of each partner, transformed by its inverse link function logit (Bürkner, 2017). In this model,  $\beta$  and  $u$  are coefficients at the population level and individual level respectively, while  $\mathbf{X}$ ,  $\mathbf{Z}$  are their corresponding design matrices. In this case, the population coefficients correspond to the presence of partners with no social closeness  $\beta^{computer}$ , low  $\beta^{stranger}$  and high  $\beta^{friend}$ .

Figure 3 shows the posterior probability of breaking the promise depending on the partner, circles correspond to the medians of the distributions of the posterior estimates effects, the thick bar, and the thin bar correspond to the interval of 50% and 95% posterior probability, respectively. It is clear that the probability of breaking the promise decreases as social closeness increases, however, we calculate the reasons for evidence for the following hypotheses:

- There is a greater probability to break the promise to the computer than the stranger.
- There is more probability to break the promise to the stranger compared to a friend.

The evidence ratio, which is the ratio between the posterior probability of the mentioned hypotheses and their corresponding alternative hypotheses was 7.47 and 8.64 respectively in favor of the previous hypotheses. As can be seen in Figure 5 in the estimate for  $\beta^{computer}$ , when there is no social closeness we have a 95% posterior probability that the parameter for promise-breaking is between 4-38%, although it is a small proportion, it clearly excludes the probability that promises are not broken.

### 3.2.3 Effects of promises and social closeness on cooperation

To evaluate the effect of the experimental conditions on the decision to pay back, we used all trials where partners invest. Again, a Hierarchical Bayesian Model was fit, which assumes that the promises and partners have an effect that varies for each individual, however, it also assumes that these variations belong to common population distributions. The model estimates that the probability  $\theta_i$  of the decision to pay back is based on the effect of the presence of promises  $\beta^{promise}$ , as well as partners with low  $\beta^{stranger}$  and high  $\beta^{friend}$  social closeness.

Figure 4 shows the posterior distributions of the  $\beta$  coefficients in the logit scale, the center line represents the median of the distribution and the shaded area corresponds to the interval of 50% posterior probability. As can be seen, more than 95% of the posterior density of the population coefficients is greater than 0, which indicates strong evidence of the effect of experimental conditions on the decision to pay back, so that the presence of promises and the stranger increase clearly the probability of paying, but the presence of the friend is the

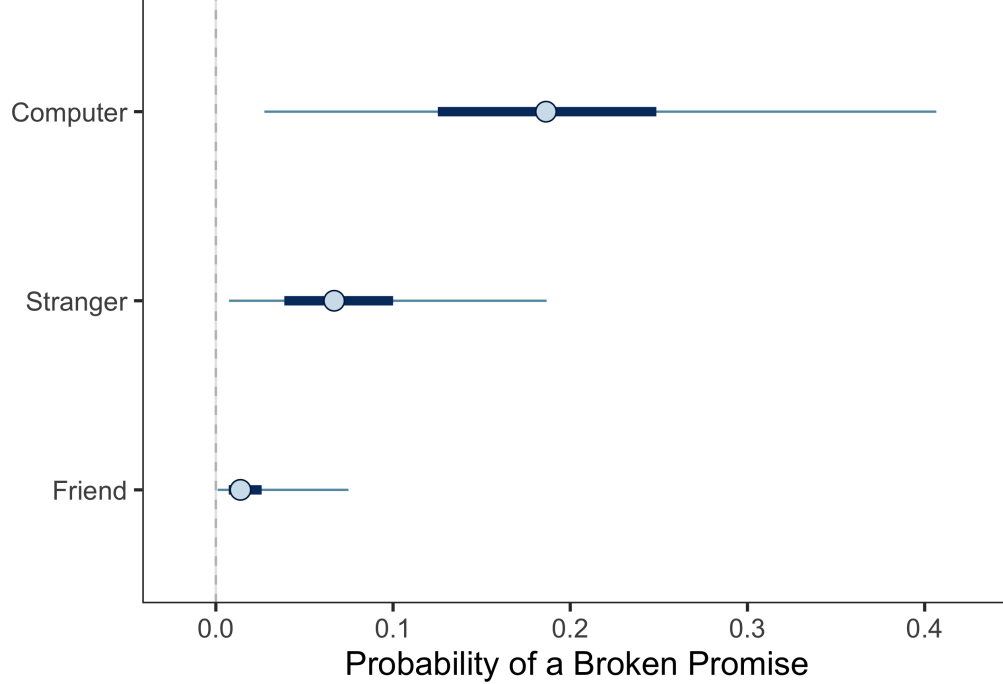


Figure 3: Posterior probability of Promise Breaking

condition that has the greatest effect on this behavior. Posterior distributions clearly show that the probability of cooperation increases as social closeness does.

Also, Figure 5 presents the posterior predictive distribution in contrast to the proportion of decisions to each partner during the promises phase, each panel corresponds to one of the first twelve subjects. The responses on the dotted line would indicate that the subject paid back random to that partner. Posterior predictive distribution simulates observations of the model and compares them with the actual data, it helps us to identify if the model is sufficiently close to the process that generated the data (Lee & Wagenmakers, 2014; Schad, Betancourt, & Vasishth, 2019). It can be seen that there is a great correspondence between the subject's responses and predictions of the model, even in cases where the model is "wrong" (for example, subject 1), the observed response is in the range of a predicted standard deviation, which gives credibility to estimates.

### 3.2.4 Effects of social closeness on cooperation varying by promises

In this model, it is assumed that social closeness has an effect that varies for each level of promise, which implies that there are levels of promise that are more sensitive than others to the effect of the partners on the decision to pay back. Again, a Hierarchical Bayesian Model was made to estimate the effect of partners at the population level and the variations depending on the promise levels. Table 1 summarizes the subsequent distributions of the model coefficients in the logit scale, including point estimates, standard errors and Bayesian Credibility Intervals of 95%. An estimate similar to the previous models can be observed, with strong evidence of the effect of social closeness on the decision to pay back. Although



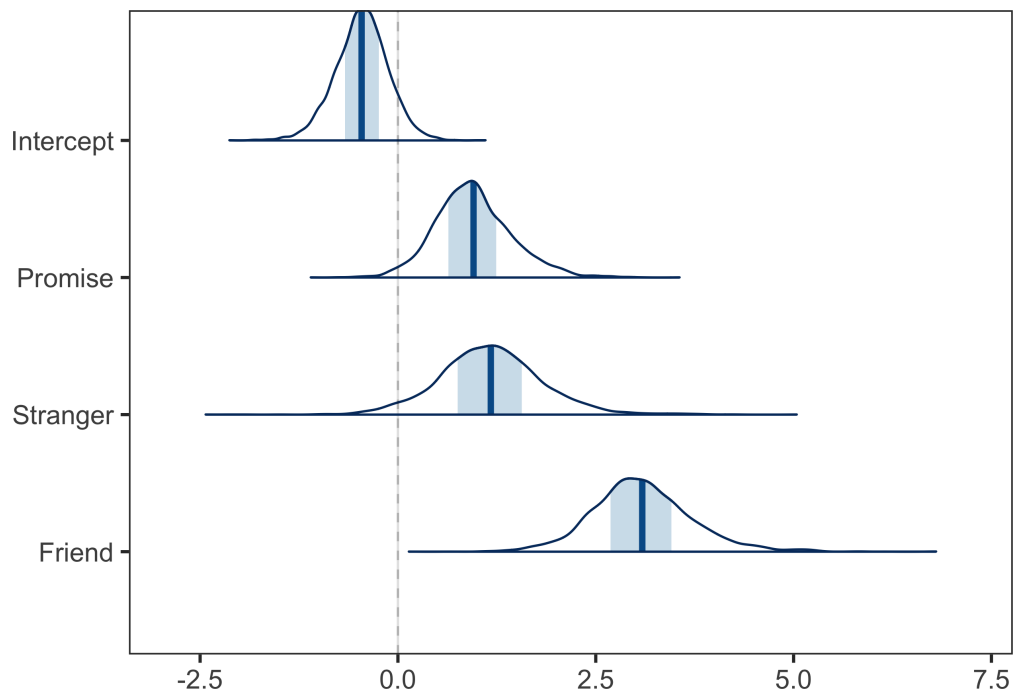


Figure 4: Posterior estimates in logit scale

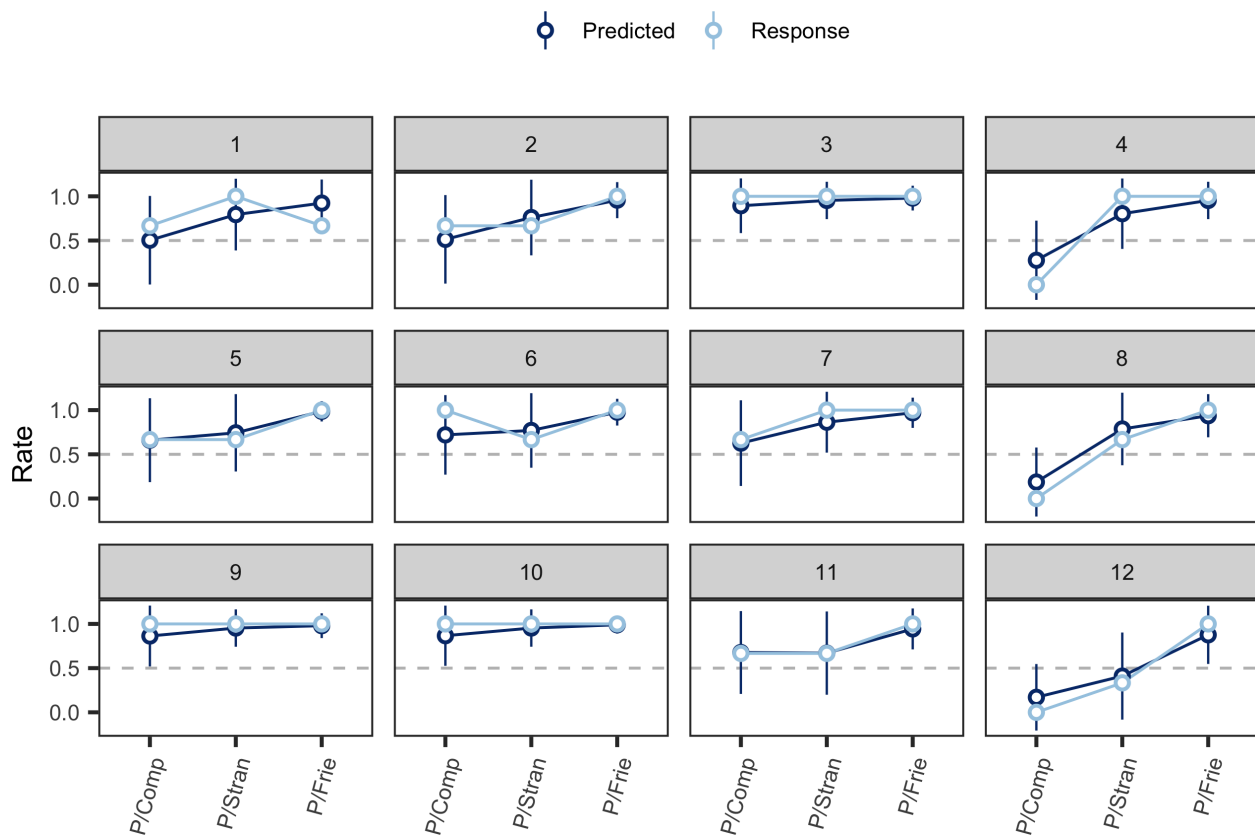


Figure 5: Posterior predictive over pay back rate

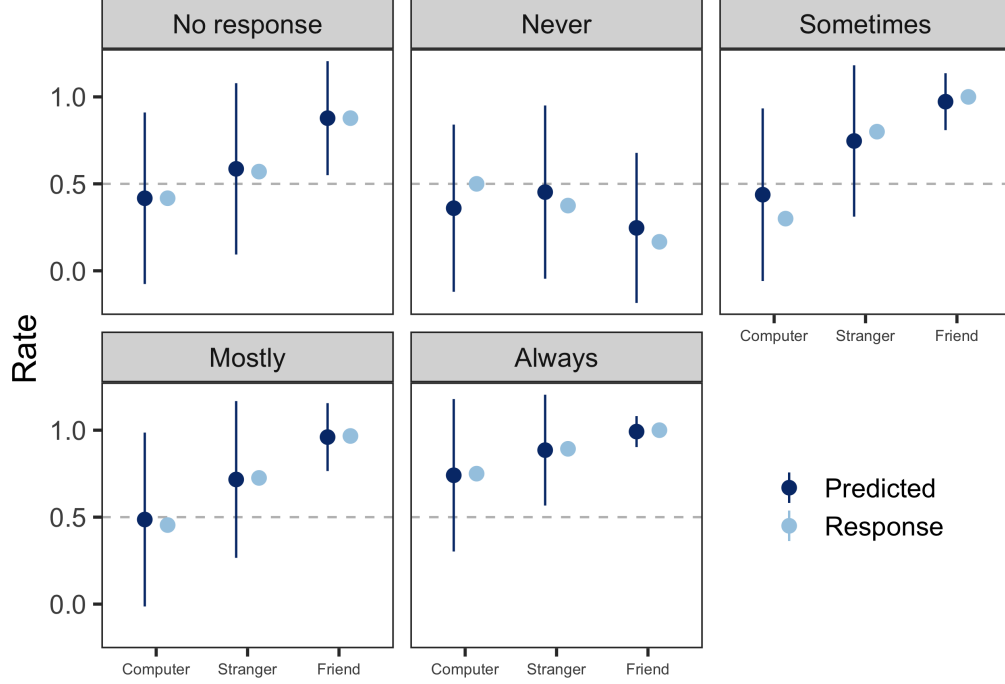


Figure 6: Posterior predictive over pay back rate, Social Closeness Varying effects by Promise levels

the credibility range for the friend’s effect includes 0, the evidence ratio that the effect is greater than zero is 15.81 with 94% of posterior probability.

On the other hand, Figure 6 shows the posterior predictive distribution, compared to the payment rates of all individuals to the different partners and their variation by the level of promise. With the exception of the promise *never*, a monotonic positive effect of partners is observed at all levels of promise, however, it can be observed how the effect of social closeness varies depending on the strength of the promise, mainly for the decisions towards the computer.

Table 1: Posterior coefficients estimates

Term	Estimate	Est.Error	95 % CI	
			Lower	Upper
Intercept	-0.153	0.536	-1.005	0.617
Stranger	0.829	0.410	0.128	1.470
Friend	2.084	1.248	-0.151	3.975

### 3.3 Cumulative Bayesian Modeling

#### 3.3.1 Commitment expressed in promises

In the original study of promises, authors chose to divide their sample into two according to the hierarchical clustering technique with Ward’s method (Baumgartner et al., 2009). In this way they obtained two sets of participants that were different in their payment rates, despite the fact that they both made very high promises, for this reason, the authors named the group that paid little *dishonest* and the group that paid a lot *honest*. In a similar exercise, in the present study we performed the hierarchical grouping technique to obtain a solution of 2 groups and we found two similar sets in  $n$  that we call the group “Low” and “High” (Low = 25, High = 20), with quite different payment ratios (Low = 58%, High = 83%), and a 95% confidence interval of 20% to 32% in the difference favor of the High group.

Although it seems a similar result to that reported in that paper, we explore the pattern of promises of both groups to determine if it was possible to classify our subjects as *honest* and *dishonest*. If we hypothesize that the commitment to pay would be reflected in the level of promises selected, a group of *dishonest* people could generate in their partners the belief that they will pay by choosing a high level of promises (mostly or always) and, later, betraying that trust when deciding not to pay back.

In order to explore whether the groups represent populations that do not differ in the level of commitment expressed in the promises, we use a Cumulative Bayesian Model, which assumes that the levels of promises are an observed ordinal variable  $Y$  that originates from the categorization of a continuous latent variable  $\tilde{Y}$ , for this case, *the expressed commitment to pay back* (Bürkner & Vuorre, 2019). The degree to which the subjects of the High group differ from the Low group, in normal standard deviations ( $z$ -values), on the latent scale of  $\tilde{Y}$ , has a point estimate of 0.95, which implies that the High group has 0.95  $z$ -values greater commitment to pay back than the Low group.

The 95% Bayesian Credibility Interval indicates that the High group is between 0.40 to 1.51  $z$ -values of difference from Low. So we can conclude with at least a 95% probability that people belonging to the High group expressed in their promises a greater commitment to pay back than the subjects of the Low group. If we look at Figure 7, the probability of choosing the different levels of promises varies depending on the group, the one who had the highest percentage of decisions to pay back is more likely to choose *always* (High group), while the group that had the lowest percentage of decisions to pay back are more likely to choose *mostly* (Low group).

According to our data, we could not justify the classification of our groups according to their honesty or dishonesty, at least not with the technique used in Baumgartner et al. (2009). Since people were quite consistent with keeping what they promised to pay.

## 4 Discussion

The objective of this study was to evaluate the effect of social closeness on keeping and breaking promises, as well as on cooperation in a trust game. Our results give evidence

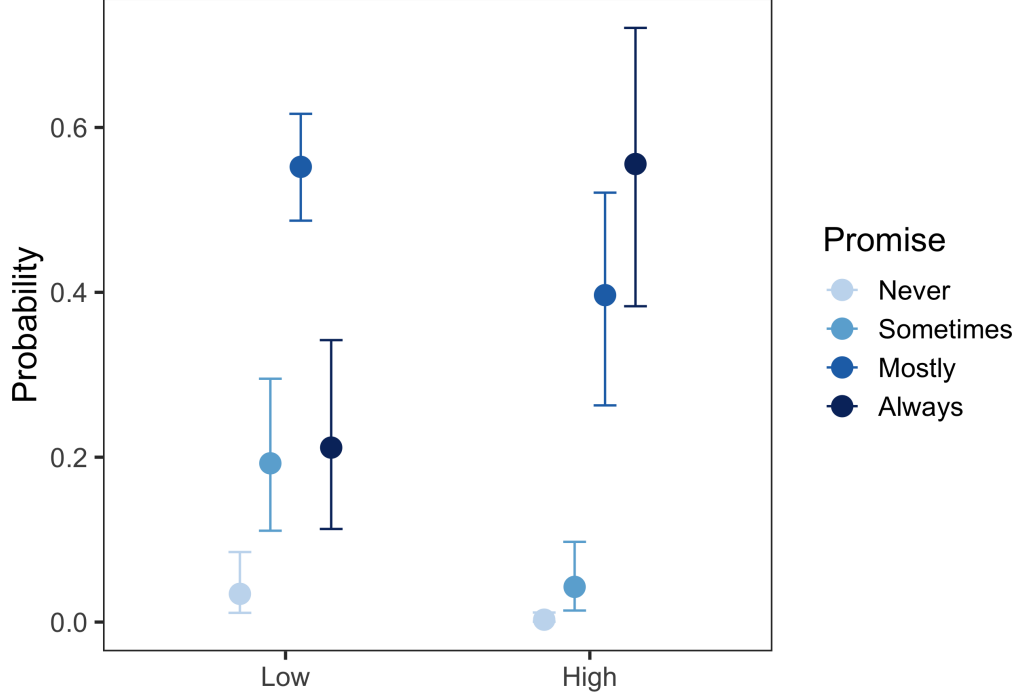


Figure 7: Promise choices by group

that zero social closeness increases the probability of breaking the promise and it decreases monotonically as social closeness increases. Likewise, as social closeness increases, so does cooperation; *high* social closeness has an effect on decisions that surpasses even those that all other experimental conditions. The most interesting finding was that when subjects express high commitment to cooperate (through choosing that they would *always* pay back) probability that they comply is also quite high, however, there is a proportion of transgressions that according to our inferences are between 4 - 38% with 95% certainty.

To our knowledge, this is the first experimental study that incorporates social closeness as a predictor of the decision to break promises with socially relevant partners. In previous studies, the participants remain anonymous during the course of the tasks (Baumgartner et al., 2009), their measurements to evaluate transgressions are self-reported (DePaulo et al., 2004; DePaulo & Kashy, 1998), they do not directly quantify the breaking of promises (Charness & Dufwenberg, 2006; Vanberg, 2008), or as we will discuss later, they use heterogeneous measures of social closeness (Glaeser, Laibson, Szeinkman, & Soutter, 2000).

Our experiment is probably the first to include a partner without social closeness (*computer*), our data give evidence that the mere fact of considering that subjects play with a human diminishes the probability of breaking the promise. It is approximately 7.5 times more likely to break the promise to the computer than to the stranger, even though that partner was not known. In our study, zero social closeness is important because it allows us to explore the *intrinsic* motivation to keep the promises, if humans keep their word because it is morally correct, we would anticipate that probability of breaking the promise was very low in the three partners and, particularly, the inference regarding the probability of breaking the promise to

the computer would include zero.

The probability of breaking the promise without social closeness, although low, is clearly greater than zero, which provides evidence that contradicts, at least in degree, the hypothesis of *intrinsic* motivation. We cannot exclude the possibility that the subjects keep promises mainly morally motivated, because even in the case of the computer, the subjects keep a large proportion, but we can exclude that this is their only motivation.

On the other hand, high social closeness allows us to explore the *instrumental* motivation to keep promises. In our study, it is almost 9 times more likely to break the promise to the stranger than to the friend; if human beings keep their word to facilitate cooperation in the future, we would anticipate that with a partner of high social closeness (with whom they are very likely to cooperate, even after the experiment) there would be less chance of breaking the promise. In fact, in our study, the promise to friends was not broken on any occasion, so the inference, in this case, is obvious.

Our findings could be explained by the social norm called *conditional cooperation*, which indicates that the belief that other people cooperate at high levels also induces high levels of cooperation (Fehr & Schurtenberger, 2018). Thus, the lack of certainty regarding the decisions that the computer would make could explain the probability of breaking the promise that exists towards this partner. Similarly, the information that the subjects have regarding friend's behavior -even before the experiment- can explain the high levels of cooperation and keeping of promises towards this partner.

Regarding cooperation, it is not the first investigation in which the effect of social closeness is evaluated, however, the nature of experimental manipulation has been diverse. For example, in one study the contribution in a public goods game was evaluated during several trials in a group made up of the same individuals (*partners*), compared to another group of new different subjects for each trial (*strangers*) (Croson, 1996). It was assumed that in the condition of *partners* the cooperation would be greater because it is a stable group compared to the group of *strangers*. The mentioned study does not give details regarding the recruitment of participants, so we could assume that even in the condition of *partners* these are individuals who do not consider themselves socially close.

In another experiment, an increase in cooperation in a trust game in pairs was reported when individuals are socially close (Glaeser et al., 2000), in this study, subjects know each other and the researchers carefully measured several variables regarding their social connection, however, some individuals who arrived together were allowed to perform the task between themselves and others were paired with unclear criteria. The above allows social closeness between partners to be heterogeneous, allows for romantic relationships and assumes that two individuals who come together to class are considered close to each other. Also, since it is a study of *one-shot*, it excludes the possibility of evaluating how the same subject varies his behavior based on different levels of social closeness.

To try to homogenize social closeness between our subjects and their partners, participants performed the task with a same-sex match friend considered close by themselves, in none case was it a relative and, although we understand that there may be same-sex couples and that the subjects may have hidden their relationship, they were emphasized that to be part

of the study their companion could not be a romantic or sexual partner. Besides, Bayesian analysis of the IOS scale showed evidence in favor that the degree of social closeness between 30 of our 45 subjects and their friends was the same. Although the results of the two studies mentioned are practically in the same direction as the findings of our experiment, establishing the similarity in effects magnitudes is not possible due to the design differences, the performed tasks, and the statistical procedures used.

A methodological contribution of our study is that, in the literature of promises, it is the first to use Bayesian inference tools: *Hierarchical Models* and *Cumulative Models*. The former (aka multilevel models), allowed us to model the effect of the experimental conditions on the individual response, these models assume that lower level observations (e.g. decisions of each individual) are nested in higher level units (e.g. individual subjects). Within-subjects designs have traditionally been analyzed with repeated measures AN(C)OVA, however, *Hierarchical Models* grant the advantages of naturally dealing with unbalanced data, including categorical and/or numerical predictors, explicitly incorporate individual variability, among others (Bürkner, 2017; Gelman & Hill, 2006; McElreath, 2018; Vuorre & Bolger, 2018).

In our case, as shown in Figure 5, the *Hierarchical Model* allowed us to capture very precisely how the experimental conditions affected the decisions of each subject and the differences between them. For example, there are obviously motivated *instrumental* subjects such as number 8 or 12 who are very sensitive to the identity of their investors and modify their cooperation under social closeness. At the same time, there are notably *intrinsic* subjects such as 9 and 10 who cooperate all the time regardless of who their partners are. The *Hierarchical Model* naturally includes this information for the estimation of population effects, which, if not considered, would lead to inaccurate inferences.

On the other hand, the *Cumulative Model* allowed us to capture the strength of the commitment expressed through promises. A very frequent problem that has been pointed out recently is that analyzing ordinal data with methods that assume that observations are *metric* can lead to serious inference errors (Liddell & Kruschke, 2018). The *Cumulative Models* assume that the observed responses come from the categorization of a continuous latent variable (Bürkner & Vuorre, 2019), so in our study, we could identify that the groups with different payment proportions, which in other articles have been called *dishonest* and *honest*, differ in the commitment they express through their promises. Considering that the tools to perform the *Cumulative Models* are relatively recent, it would be valuable to replicate published studies (e.g., Baumgartner et al., 2009, 2013) considering the new alternatives of inference.

Although the *Hierarchical* and *Cumulative* models are not tools exclusively for Bayesian inference, their application from this approach represents several advantages compared to frequentist statistics. Classic problems of the previous one such as multiple comparisons, the decision of when to stop collecting subjects (stopping rule), or use of planned comparisons versus *post hoc*, are not factors that affect the support provided by the data to the hypothesis from the Bayesian approach (Dienes, 2011). In Bayesian inference, hypothesis testing makes formal use of probability to express the plausibility of theories, in our case we were able to obtain evidence ratios regarding the extent our data support hypotheses, regardless these were proposed *a priori* or *post hoc*.

A limitation of this paper is the use of hypothetical rewards compared to real rewards. A relatively recent study on decision-making reported that there is more loss aversion when subjects have real monetary rewards compared to hypothetical in a risk task (Xu et al., 2016). However, other works do not report differences between the use of both types of rewards in self-control, temporal and social discount tasks (Johnson & Bickel, 2006; Locey, Jones, & Rachlin, 2011; Rachlin & Jones, 2006). Likewise, it can be argued that our results have theoretical congruence (Fehr & Schurtenberger, 2018) and are in the same direction as other works that use real money (Croson, 1996; Glaeser et al., 2000), so there are not many reasons to expect that other types of rewards would modify our results.

Another possible limitation is the use of multiple trials with each partner. Although the decision to use repeated measures during our design serves more the purpose of reducing contaminant sources and increasing internal validity (Maxwell, Delaney, & Ken, 2018), to have more clarity regarding the difference in the *intrinsic* and *instrumental* motivations, the studies could benefit from only include one trial for each partner. Play a *one-shot* with each partner avoids the possibility that multiple trials could generate the belief that the computer can also vary its behavior according to the decisions of the *trustee*.

## 5 Conclusions

Subjects keep their promises by a combination between the *intrinsic* and *instrumental* motivations. They are predominantly *nice guys* because they keep their word even with investors with no social closeness. However, its trustworthiness is far from perfect, since there is a small, but reliably greater than zero, proportion of betrayals committed towards this partner. It seems that if subjects cannot assure that their investor will be a *potential cooperator*, dishonesty may emerge in the form of broken promises. The theoretical predictions of the social norm *conditional cooperation* frame the finding that social closeness decreases the probability of breaking the promise and increases cooperation. In the same way that suggests the correspondence between commitment expressed through promises and subsequent behavior.

## Bibliography

- Anderson, D. E., DePaulo, B. M., & Ansfield, M. E. (2002). The development of deception detection skill: A longitudinal study of same-sex friends. *Personality and Social Psychology Bulletin*, 28(4), 536–545. <https://doi.org/10.1177/0146167202287010>
- Aron, A., Melinat, E., Aron, E. N., Vallone, R. D., & Bator, R. J. (1997). The experimental generation of interpersonal closeness: A procedure and some preliminary findings. *Personality and Social Psychology Bulletin*, 23(4), 363–377. <https://doi.org/10.1177/0146167297234003>
- Balliet, D. (2010). Communication and cooperation in social dilemmas: A meta-analytic review. *Journal of Conflict Resolution*, 54(1), 39–57. <https://doi.org/10.1177/0022002709352443>

- Baumgartner, T., Fischbacher, U., Feierabend, A., Lutz, K., & Fehr, E. (2009). The Neural Circuitry of a Broken Promise. *Neuron*, 64(5), 756–770. <https://doi.org/10.1016/j.neuron.2009.11.017>
- Baumgartner, T., Gianotti, L. R. R., & Knoch, D. (2013). Who is honest and why: Baseline activation in anterior insula predicts inter-individual differences in deceptive behavior. *Biological Psychology*, 94(1), 192–197. <https://doi.org/10.1016/j.biopsycho.2013.05.018>
- Bürkner, P.-C. (2017). brms : An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, 80(1). <https://doi.org/10.18637/jss.v080.i01>
- Bürkner, P.-C., & Vuorre, M. (2019). Ordinal Regression Models in Psychology: A Tutorial. *Advances in Methods and Practices in Psychological Science*, 2(1), 77–101. <https://doi.org/10.1177/2515245918823199>
- Charness, G., & Dufwenberg, M. (2006). Promises and partnership. *Econometrica*, 74(6), 1579–1601. <https://doi.org/10.1111/j.1468-0262.2006.00719.x>
- Croson, R. T. (1996). Partners and strangers revisited. *Economics Letters*, 53(1), 25–32. [https://doi.org/10.1016/S0165-1765\(97\)82136-2](https://doi.org/10.1016/S0165-1765(97)82136-2)
- DePaulo, B. M., Ansfield, M. E., Kirkendol, S. E., & Boden, J. M. (2004). Serious Lies. *Basic & Applied Social Psychology*, 26(2/3), 147–167. [https://doi.org/10.1207/s15324834basp2602&3\\_4](https://doi.org/10.1207/s15324834basp2602&3_4)
- DePaulo, B. M., & Kashy, D. A. (1998). Everyday Lies in Close and Casual Relationships. *Journal of Personality and Social Psychology*, 74(1), 63–79. <https://doi.org/10.1037/0022-3514.74.1.63>
- Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on? *Perspectives on Psychological Science*, 6(3), 274–290. <https://doi.org/10.1177/1745691611406920>
- Fareri, D. S., & Delgado, M. R. (2014). Differential reward responses during competition against in- and out-of-network others. *Social Cognitive and Affective Neuroscience*, 9(4), 412–420. <https://doi.org/10.1093/scan/nst006>
- Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature*, 425(6960), 785–791. <https://doi.org/10.1038/nature02043>
- Fehr, E., & Schurtenberger, I. (2018). Normative foundations of human cooperation review-article. *Nature Human Behaviour*, 2(7), 458–468. <https://doi.org/10.1038/s41562-018-0385-5>
- Fischbacher, U., & Föllmi-Heusi, F. (2013). Lies in disguise—an experimental study on cheating. *Journal of the European Economic Association*, 11(3), 525–547. <https://doi.org/10.1111/jeea.12014>
- Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. (p. 651). <https://doi.org/10.2277/0521867061>
- Glaeser, E., Laibson, D., Scainkman, J., & Soutter, C. (2000). Measuring Trust. *The Quarterly Journal of Economics*, 115(3), 811–846.



- Gneezy, U. (2005). Deception: The role of consequences. *American Economic Review*, 95(1), 384–394. <https://doi.org/10.1257/0002828053828662>
- Gneezy, U., Rockenbach, B., & Serra-Garcia, M. (2013). Measuring lying aversion. *Journal of Economic Behavior and Organization*, 93, 293–300. <https://doi.org/10.1016/j.jebo.2013.03.025>
- Holt-Lunstad, J. (2018). Why Social Relationships are Important for Physical Health: A Systems Approach to Understanding and Modifying Risk and Protection. *Ssrn*, (October 2017), 1–22. <https://doi.org/10.1146/annurev-psych-122216-011902>
- Johnson, M. W., & Bickel, W. K. (2006). Within-subject comparison of real and hypothetical money rewards in delay discounting. *Journal of the Experimental Analysis of Behavior*, 77(2), 129–146. <https://doi.org/10.1901/jeab.2002.77-129>
- Kessler, G., Rizzo, S., & Kelly, M. (2019). President Trump has made more than 10,000 false or misleading claims. Retrieved from [https://www.washingtonpost.com/politics/2019/06/10/president-trump-has-made-false-or-misleading-claims-over-days/?utm{\\\_}term=.8565a16aa5e8](https://www.washingtonpost.com/politics/2019/06/10/president-trump-has-made-false-or-misleading-claims-over-days/?utm{\_}term=.8565a16aa5e8)
- Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian Cognitive Modeling*. <https://doi.org/10.1017/cbo9781139087759>
- Liddell, T. M., & Kruschke, J. K. (2018). Analyzing ordinal data with metric models: What could possibly go wrong? *Journal of Experimental Social Psychology*, 79(November 2017), 328–348. <https://doi.org/10.1016/j.jesp.2018.08.009>
- Locey, M. L., Jones, B. A., & Rachlin, H. (2011). Real and hypothetical rewards in self-control and social discounting. *Judgment and Decision Making*, 6(6), 552–564.
- Maxwell, S. E., Delaney, H. D., & Ken, K. (2018). *Designing experiments and analyzing data: a model comparison perspective* (3rd edition, pp. 611–612). New York: Routledge.
- Mazar, N., Amir, O., & Ariely, D. (2008). The Dishonesty of Honest People: A Theory of Self-Concept Maintenance. *Journal of Marketing Research*, 45(6), 633–644. <https://doi.org/10.1509/jmkr.45.6.633>
- McElreath, R. (2018). *Statistical rethinking: A bayesian course with examples in R and stan*. <https://doi.org/10.1201/9781315372495>
- Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., ... Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1), 195–203. <https://doi.org/10.3758/s13428-018-01193-y>
- Peirce, J. W. (2008). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics*, 2. <https://doi.org/10.3389/neuro.11.010.2008>
- Rachlin, H., & Jones, B. (2006). Social Discounting. *Psychological Science*, 17(4), 283–286.
- R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>

- Schad, D., Betancourt, M., & Vasishth, S. (2019). Toward a principled Bayesian workflow in cognitive science. Retrieved from [osf.io/b2vx9](https://osf.io/b2vx9)
- Sip, K. E., Smith, D. V., Porcelli, A. J., Kar, K., & Delgado, M. R. (2015). Social closeness and feedback modulate susceptibility to the framing effect. *Social Neuroscience*, *10*(1), 35–45. <https://doi.org/10.1080/17470919.2014.944316>
- Smith, T., Davern, M., Freese, J., & Hout, M. (2018). GSS Data Explorer | NORC at the University of Chicago. Retrieved from <https://gssdataexplorer.norc.umd.edu/variables/5067/vshow>
- Stan Development Team. (2018). RStan: The R interface to Stan. Retrieved from <http://mc-stan.org/>
- Tough, H., Siegrist, J., & Fekete, C. (2017). Social relationships, mental health and wellbeing in physical disability: A systematic review. *BMC Public Health*, *17*(1), 1–18. <https://doi.org/10.1186/s12889-017-4308-6>
- Vanberg, C. (2008). Why Do People Keep Their Promises? An Experimental Test of Two Explanations1. *Econometrica*, *76*(6), 1467–1480. <https://doi.org/10.3982/ECTA7673>
- Vuorre, M., & Bolger, N. (2018). Within-subject mediation analysis for experimental data in cognitive psychology and neuroscience. *Behavior Research Methods*, *50*(5), 2125–2143. <https://doi.org/10.3758/s13428-017-0980-9>
- Xu, S., Pan, Y., Wang, Y., Spaeth, A. M., Qu, Z., & Rao, H. (2016). Real and hypothetical monetary rewards modulate risk taking in the brain. *Scientific Reports*, *6*. <https://doi.org/10.1038/srep29520>