Promises, Expectations, and Social Cooperation*

Dorothee Mischkowski
University of Goettingen

Rebecca Stone
UCLA

Alexander Stremitzer†
UCLA

January 16, 2018

Abstract

Promising serves as an important commitment mechanism by operating on a potential cheater's internal value system. We present experimental evidence on what motivates people to keep their promises. First, they feel that they are duty-bound to keep their promises regardless of whether promisees expect them to (promising per se effect). Second, they care about not disappointing promisees' expectations, regardless of whether those expectations were induced by the promise (expectations per se effect). Third, they are even more motivated to avoid disappointing promisees' expectations when those expectations were induced by a promise (interaction effect). Clear evidence of some of these effects has eluded the prior literature due to limitations inherent to the experimental methods employed. We sidestep those difficulties by using a novel between-subject vignette design. Our results also shed light on how promising may contribute to the self-reinforcing creation of trust.

Keywords: cooperation, promises, expectations, trust, guilt aversion.

JEL-Classification: A13, D03, D64, C91, K12.

---

*The authors are grateful to Oren Bar-Gill, Gary Charness, Martin Dufwenberg, Russell Korobkin, Alexander Morell, Ernesto Reuben and seminar audiences at Harvard, UCLA, the Yale-UCL conference on the Philosophy of Contract, and the Bonn LawEcon workshop for helpful comments and suggestions. We thank Zhipeng Liao, Ben Nyblade and Henry Kim for their advice on statistical methods and Henry Kim, Rita Hsu and Danny Wong for excellent research assistance.

†University of Goettingen, Georg-Elias-Mueller Institute for Psychology, Gosslerstrasse 14, D-37073 Goettingen, Germany, dorothee.mischkowski@psych.uni-goettingen.de; UCLA Law School, 385 Charles E. Young Drive, 1242 Law Building, Los Angeles, CA 90095, rebecca.stone@law.ucla.edu, stremitzer@law.ucla.edu (corresponding author).

1 Introduction

Promises are ubiquitous in both private and in commercial settings. Casual observation, introspection, and a string of recent economic studies suggest that people are willing to keep promises even in the absence of third-party enforcement mechanisms and the second-party enforcement mechanisms that arise from repeated interactions (Ellingsen and Johannesson, 2004; Charness and Dufwenberg, 2006; Vanberg, 2008). Promises therefore seem to serve as important mechanisms of commitment, enabling people to solve fundamental problems of social cooperation.[1] In particular, promises facilitate the processes of exchange over time by diminishing the promisor's willingness to opportunistically exploit the promisee at least in the absence of an alternative avenue of enforcement.[2]

A clear understanding of the determinants of promise-keeping is important for institutional design, particularly given that a person's intrinsic reasons for keeping her promises are likely to interact with the extrinsic incentives that are provided by

[1] Promises serve as a first-party system of private governance. The distinguishing feature of a first-party system is that it operates on the potential cheater's internal value system, eliminating opportunistic behavior at the source, as opposed to eliminating it indirectly by the extrinsic incentives created by second-party and third-party enforcement systems. See Avinash Dixit's presidential address to the American Economic Association (Dixit, 2009). For third-party systems, see the vast economic literature on formal contracts beginning with Mirrlees (1976) and Holmström (1979). For second-party systems see the literature on relational contracting (Macaulay 1963, Klein and Leffler 1981, Bull 1987, Kreps 1996, MacLeod and Malcomson 1989, Levin 2003).

[2] An important class of situations where possibilities of such opportunism arise are those in which one party makes a non-contractible relationship-specific investment before a transaction takes place such that he may be "held up" by the other in later negotiations over the price, in which case he may not be able to guarantee himself a sufficient share of the return ex post to justify the investment. Because the investment is non-contractible, the legal system can't mitigate the hold-up problem, but an investor's preference for negative reciprocity may make his threat to punish his partner credible and so give rise to a second-party enforcement mechanism that mitigates the potential for hold up (Dufwenberg, Smith, and Van Essen, 2011). For classic discussions of the hold-up problem see Williamson (1979, 1985), Grout (1984), Grossman and Hart (1986), Hart and Moore (1988).

second- and third-party enforcement mechanisms (Benabou and Tirole, 2003). Yet the empirical literature to date has only started to shed light on exactly why people keep their promises.

Three types of reasons why people keep their promises, which are not necessarily mutually exclusive, have been suggested by the experimental economics literature. First, a promisor may feel bound to keep his promises insofar as he fears disappointing the promisee's expectations of performance, as in Charness and Dufwenberg's (2006) theory of "guilt aversion". We call this the expectations per se effect. There are a variety of underlying reasons why a promisor would care about the promisee's beliefs. He might, for example, believe that the promisee finds disappointment intrinsically painful, and feel an altrustic desire to prevent such pain. Or he might believe that the promisee is more likely to invest in reliance on a promise the more she believes that the promise will be kept such that he is more likely to harm her if he breaks the promise. If this is the only reason why people keep their promises absent self-interested reasons to do so, then keeping his promises would matter to the promisor only derivatively through the effect his promises had on the promisee's expectations—that is, only insofar as his promises made the promisee more likely to anticipate performance. If a promisee formed the same expectation of performance absent a promise, the promisor would feel equally bound to perform.[3]

Second, a promisor may feel duty-bound to keep his promises regardless of the

[3]A subject could, for example, form expectations of performance absent a promise after hearing the other party make a statement of intention. See Battigalli and Dufwenberg (2007, 2009) for a formal model of what they refer to as "simple guilt" according to which people have a preference to avoid guilt from letting others down. Formally speaking, their model builds on psychological game theory by incorporating others' beliefs into people's utility function. See Dufwenberg and Gneezy (2000) for experimental evidence of guilt aversion (which they called "let-down aversion" at the time).

promisee's expectations of performance (Vanberg, 2008). We call this the *promising per se effect*.[4] Such an effect is consistent with a Kantian conception of promising according to which a promise gives the promisee the moral right to demand performance, regardless of whether the promisee would be harmed were the promise not kept (see, e.g. Shiffrin, 2008).

Third, promises may interact with expectations. A promisor might be particularly concerned about not disappointing a promisee's expectations that were caused by his promise, such that the effect of promising is to make the promisor more sensitive to the promisee's expectations than he would have been had he not made a promise. We call this the *interaction effect*. Such an interaction effect is consistent with Scanlon's theory of promising, according to which promises are significant insofar as the promisor intentionally brings about a promisee's expectations of performance thus rendering him morally responsible for this expectation and so duty-bound not to disappoint it (Scanlon, 1998, pp. 295-327). Indeed, Scanlon's theory goes further by denying that there is a duty to keep a promise that hasn't caused the promisee to expect performance (p. 312). Thus, a purely Scanlonian promisor wouldn't exhibit the *promising per se effect*. The theory of "conditional guilt aversion" proposed by Ederer and Stremitzer (2017) assumes an extreme version of the interaction effect by supposing that promisees' expectations matter if and only if those expectations were generated by a promise contrary to the *expectations per se effect*.[5]

We employ a between-subject vignette study to explore these possible determi-

[4] Further experimental evidence consistent with a commitment-based explanation for promise keeping can be found in Braver (1995), Ostrom, Walker, and Gardner (1992), Ellingsen and Johannesson (2004), and Ismayilov and Potters (2012).

[5] Stone and Stremitzer (2016) propose an analogous conditional guilt aversion theory of the effect of reliance on promise keeping.

nants of promise keeping. We ask subjects to imagine that they are a prospective buyer of a product who has told a seller that he will buy the product from her upon her return from a trip out of town. Some subjects are told that they *promised* the seller that they would buy the good from her, while others are told that they simply told the seller that they merely *planned* to do so, explicitly stating that they were not making a promise. Subjects are then told that the seller has formed a belief about the likelihood that the buyer will actually buy the product from her. Some are told that she is certain he will buy the product from her, others are told that she believes that there is a fifty percent chance, while others are told that she is certain that he won't buy the product from her. Subjects are then asked how likely it is that they will buy the product from this seller despite having learned that a second seller is selling the same product at a lower price.

We use a vignette study in which subjects are asked to imagine that they interacted with a seller in this way as opposed to an experimental design in which such an interaction actually occurs and there are real monetary stakes because it is difficult to manipulate subjects' expectations about the behavior of others in a controlled fashion. Our novel design allows us to sidestep this problem by simply telling subjects what the seller believes about the likelihood that the buyer will perform. This enables us to reproduce and clarify key results of the prior literature, while also providing the first clear evidence of the *interaction effect* and the *promising per se effect*—all within the confines of a single experiment.[6]

We find evidence of all three hypothesized effects. Consistent with the *promising*

[6] See Section 3.1 for a more thorough discussion of the advantages and disadvantages of the vignette method relative to a design with real communication, monetary incentives and endogenous promising. See Section 5 for a discussion of our contributions to the literature.

per se effect, subjects are on average more inclined to perform when they promised to do so regardless of their counterparty's expectations. In particular, even if his counterparty was certain that he wouldn't perform, a subject is more likely to perform when he promised that he would do so. Consistent with the *expectations per se effect*, subjects are on average more likely to perform the greater are their counterparties' expectations of performance even if there was no promise. Thus, our data rejects the strong interaction effect implied by Ederer and Stremitzer's "conditional guilt aversion." However, we do find evidence of a weaker interaction effect. The average sensitivity of subjects' willingness to perform to their counterparties' expectations is higher when they made a promise.

Our findings have implications for understanding extralegal mechanisms of co-operation. They show that promising is a useful commitment mechanism and not simply because it creates expectations of performance in the promisee. Promising creates commitment independently of the promisee's expectations and enhances the commitment effect of those expectations.[7]

This raises the possibility of an interesting dynamic. Both our expectation per se effect and our interaction effect suggest that even in the absence of promising, more trust leads to higher cooperation and, conversely, the absence of trust undermines cooperation.[8] This suggests that promising may be associated with self-reinforcing spirals of trust or distrust. When parties initially trust one another, possibilities

[7] Of course, there is a sense in which it is not surprising that promising could have this multitude of effects. In a world in which promising had no force independent of the expectations promises create in promisees, promising wouldn't cause rational promisees to expect performance. Our results underline that promising is an important mechanism of commitment, a fact that is taken for granted in the design of much of American contract law. See, e.g., Restatement Second of Contracts §1, (1981).

[8] Reuben et al. (2009) show that distrust is self-fulfilling in a context where promises are absent.

for cooperation are enhanced, creating even more trust. Conversely, when parties initially distrust one another, cooperation is less likely, further undermining the development of trust between them. Our promising per se effect, however, gives parties a way of breaking out of a negative spiral of distrust. If promising is a way of creating commitment even in the absence of trust, promising can build trust where it is initially lacking leading to positive instead of negative self-reinforcing dynamics.

The remainder of the paper is organized as follows. Section 2 develops a more formal model of the possible reasons why people keep their promises. Section 3 describes the design and procedure of our experiment and derives the hypotheses we are going to test. Section 4 reports our results. We subsequently discuss these results and their contribution to the existing literature in Section 5. Section 6 concludes.

2 Theory

Suppose a buyer is thinking about buying a product from a seller. The buyer is unable to make the purchase immediately because he is out of town. But he informs the seller that he will purchase the product when he returns. He has a choice when communicating his intention to the seller: he can either promise the seller that he will purchase the product from the seller when he returns or simply tell the seller that this is his plan without making the seller any promises. After talking to the seller, however, he learns that a second seller is selling the same product at a lower price. And so he must decide whether to buy the product from the first seller at the higher price as he said he would or purchase the good from the second seller at the lower price. The first seller has formed an expectation about the likelihood that the buyer will purchase the product from her, and the buyer learns this expectation. To what extent does this expectation influence the buyer's willingness to buy the good from the first seller? And how is the buyer's willingness to buy the good from the first seller affected by whether or not he made the seller a promise?

Formally, the timeline is as follows. At the first stage, the buyer decides whether or not to make the first seller a promise $p \in \{0,1\}$. He can either tell the seller that he promises to buy the good when he returns, $p = 1$, or that while he plans to buy the good from the seller he doesn't promise that he will do so, $p = 0$. At the second stage, the seller forms an expectation about the likelihood that the buyer will buy the product from her, $e \in [0,1]$, and the buyer learns this expectation from a neutral third-party. (We assume that the he learns the expectation in this way so that he has no reason to question the truth of the reported expectation. The seller (or any other interested party) would have an incentive to deceive the buyer about this expectation if the seller's reported expectation is likely to influence the likelihood that the buyer ultimately buys the product from the seller.[9]) The buyer then learns that a second seller is selling the product at a lower price. Finally, the buyer chooses an action, $a \in \{0,1\}$, either buying the good from the first seller, $a = 1$, or buying it from the second seller, $a = 0$.

The buyer's behavior, and therefore his prediction about his own behavior, is determined by his preferences. In the experiment, we ask subjects to report their beliefs about the likelihood that they would buy from the first seller if they were in the position of the buyer in this scenario.[10] Thus, it is helpful to envisage a penultimate

[9]Ellingsen et al. (2010) ask the recipient in a dictator game to self-report her expectations. These expectations are then communicated to the dictator. This at least raises the possibility of strategic communication by the recipient, and/or the dictator discounting reported expectations.

[10]This mimics the question asked by Wilkinson-Ryan and Hoffmann (2015) for which our model provides an explicit justification. See Subsection 3.1 for a discussion of this design choice.

stage of the timeline during which the buyer forms a prediction about his own future behavior, $b(p, e) \in [0, 1]$, the likelihood that he will end up buying the product from the first seller. To make sense of the idea that the buyer might be uncertain about his own future behavior, suppose that there are random components of the buyer's utility function $U(a, p, e)$, ε_a, that are realized only at the final stage. We assume the ε_a are independently and identically distributed with $E[\varepsilon_a] = 0$. These random elements could result from uncertain subjective components of the buyer's utility or uncertain features of the environment that affect the utility of his options. They ensure that the buyer is uncertain about the content of his preferences, and hence his future decision, when he forms a prediction about this decision at the penultimate stage.

We suppose that the buyer's preferences are shaped by his moral attitudes towards promising and the seller's expectations—or, more precisely, his beliefs about the seller's expectations, but we elide this distinction by supposing that he learns the seller's expectations with certainty. Thus, the buyer may prefer an action that runs counter to his material self-interest, because his moral attitudes may cause him to sacrifice some of his own welfare for the sake of duty or an altruistic desire to promote the well-being of another.[11]

What attitudes towards promising and expectations are likely to be exhibited by people in the buyer's situation? We identify four stylized attitude types.

Self-interested: Classical economic theory predicts that the buyer cares only about his own material well-being. Such a buyer doesn't care about keeping his

[11] We use preferences to represent how the buyer chooses among the available actions and not necessarily to represent the buyer's welfare. Since the sole purpose of our model is to make predictions, we can remain agnostic about the relationship between his preferences and his welfare.

promises nor about satisfying the seller's expectations. He will buy from the first seller only when it is in his self-interest to do so.

Compassionate: A buyer might not care about keeping his promises but nonetheless display compassion towards the seller by promoting the seller's well-being. Such a buyer will care about satisfying the seller's expectations if the seller is likely to be harmed in some way when her expectations are thwarted, because, for example, she has relied on the buyer performing or she experiences psychological pain as a result of disappointed expectations.

A buyer may not exhibit compassion towards the seller but nonetheless care about doing his moral duty such that he is inclined to keep promises that he perceives to be morally binding. There are two possibilities here.

Kantian: The first is that the buyer believes that promises are morally binding simply by virtue of being promises, and therefore morally binding irrespective of the seller's expectations. The buyer might believe, for example, that the promise gives the seller a moral right to demand performance (see, e.g., Shiffrin, 2008).

Scanlonian: The second possibility is that the buyer believes himself duty-bound to honor expectations in others that he is morally responsible for and believes that promising is a way of taking responsibility for the expectations that others form in reliance on the promise (Scanlon, 1998, pp. 295-327). Unlike the compassionate buyer, he is indifferent towards the seller's expectations if he doesn't believe himself morally responsible for those expectations. The Scanlonian buyer may believe that he bears some moral responsibility for the seller's expectations even when he expressly told the seller that he was making no promise given that his mere statement of intent may have influenced the seller's expectations. What is important for our purposes is that such a buyer believes that he bears a greater degree of moral responsibility for the seller's expectation when he made a promise. Furthermore, unlike the Kantian, the Scanlonian believes that his moral duty depends on the seller actually forming some expectation that he will perform. Such a buyer will therefore be more likely to buy the good from the first seller the greater is the seller's expectation of performance, but only if the buyer made a promise to buy from the seller.

More formally, let $m(a)$ represent the material utility associated with action a, where $m(0) > m(1)$. Thus, a purely self-interested buyer, cares only about maximizing m, and so will always choose to buy the product from the second buyer.

If the buyer exhibits compassion then he cares about ensuring that his actions meet or exceed the seller's expectations regardless of whether he made a promise to do so. Formally, he cares about minimizing $j(e-a)$ where $j(x) > 0$, $j'(x) > 0$ if $x > 0$ and $j(x) = 0$, $j'(x) > 0$ otherwise.

If the buyer exhibits a Kantian disposition, then he cares about buying from the first seller, $a = 1$, whenever he promised to do so. Thus, formally, he cares about minimizing $pk(1-a)$ where $k(x) > 0$, $k'(x) > 0$ if $x > 0$ and $k(x) = 0$, $k'(x) > 0$ if $x = 0$.

If the buyer exhibits a Scanlonian disposition, then he cares about ensuring that his actions meet the seller's expectations only if he made a promise to do so. Formally, therefore, he cares about minimizing $pl(e-a)$ where $l(x) > 0$, $l'(x) > 0$ if $x > 0$ and $l(x) = 0$, $l'(x) > 0$ otherwise.

People might exhibit combinations of these attitudes. Thus, the utility function of buyer i can be expressed as:

$$U_i(a,p,e) = m(a) - \alpha_i j(e-a) - p[\beta_i k(1-a) + \gamma_i l(e-a)] + \varepsilon_a, \tag{1}$$

where $\alpha_i, \beta_i, \gamma_i \geq 0$ are parameters that describe the weight the buyer places on the aforementioned considerations. If $\alpha_i = \beta_i = \gamma_i = 0$, the buyer is self-interested. The strength of his compassionate attitudes is measured by α_i, the strength of his Kantian attitudes is measured by β_i, and the strength of his Scanlonian attitudes is measured by γ_i.

The buyer will choose to buy from the seller whenever

$$\begin{aligned} \Delta U_i &\equiv U_i(1,p,e) - U_i(0,p,e) \qquad (2)\\ &= \Delta m + \alpha_i j(e) + p\beta_i k(1) + p\gamma_i l(e) + \Delta\varepsilon \geq 0, \end{aligned}$$

where $\Delta m \equiv m(1) - m(0) < 0$ and $\Delta\varepsilon \equiv \varepsilon_1 - \varepsilon_0$, and therefore $E[\Delta\varepsilon] = 0$. Note that the larger is α_i, the more likely the buyer will buy from the first seller the greater are the seller's expectations even if the buyer didn't make a promise; the larger is β_i, the more the buyer thinks he is duty-bound to keep his promises regardless of expectations; and the larger is γ_i, the more likely the buyer will buy from the seller the greater are the seller's expectations if he made a promise.

Let $F(\cdot)$ be the c.d.f. of $\Delta\varepsilon$ and $f(\cdot)$ be the p.d.f. It follows from (2) that during the penultimate stage of the experiment the buyer's belief about his own future behavior will be given by:

$$\begin{aligned} b_i(p,e) &= \Pr[\Delta\varepsilon \geq -\Delta m - \alpha_i j(e) - p\beta_i k(1) - p\gamma_i l(e)] \\ &= 1 - F(-\Delta m - \alpha_i j(e) - p\beta_i k(1) - p\gamma_i l(e)). \end{aligned}$$

In order to ensure that $b_i(p,e) \in (0,1)$ we further make the technical assumption that $\Delta\varepsilon$ is uniformly distributed on $[-c,c]$ where

$$c > -\Delta m \text{ and } -c < -\Delta m - \alpha_i j(1) - p\beta_i k(1) - p\gamma_i l(1). \tag{3}$$

There are three important implications. First, regardless of whether a promise was made, then so long as the buyer exhibits some degree of compassion, $\alpha_i > 0$, b_i is increasing in the seller's expectation,

$$\frac{\partial b_i(1,e)}{\partial e} = f(\cdot)\left[\alpha_i j'(e) + \gamma_i l'(e)\right] \geq f(\cdot)\,\alpha_i j'(e) = \frac{\partial b_i(0,e)}{\partial e} > 0.$$

We call this the *expectations per se effect.* Second, so long as the buyer exhibits a Scanlonian predisposition, $\gamma_i > 0$, the rate of increase of b_i in the seller's expectation e is higher when a promise was made than when no promise was made. We call this the *interaction effect.* Third, so long as the buyer exhibits a Kantian predisposition, $\beta_i > 0$, b_i is always higher if a promise was made,

$$b_i(1,e) - b_i(0,e) = F(-\Delta m - \alpha_i j(e)) - F(-\Delta m - \alpha_i j(e) - \beta_i k(1) - \gamma_i l(e)) > 0$$

We call this the *promising per se effect.*

3 Design, Hypotheses & Procedure

3.1 Design

Subjects are asked to imagine that they are a prospective buyer of a good in a version of the scenario set out in the previous section. More specifically, they are asked to imagine that a seller, B, has offered to sell them a product that they are interested in buying for $100 once they get back from a trip out of town, but that just before returning from the trip they learn that another seller, C, is offering to sell an equivalent product at the lower price of $85.[12]

Subjects are then asked to indicate how likely they believe it is that they will buy the product from C (instead of B) under *one* of six randomly selected conditions. The conditions differ according to whether or not the buyer promised a seller that he will buy the good from her, and which expectations the seller forms about the likelihood that the buyer will in fact buy the good from her. Specifically, there are three "Promise conditions" and three "No Promise conditions," each one characterized by a particular expectation: "0% " "50%," or "100%". In the Promise conditions, subjects are asked to imagine that they promised the seller that they would buy the product when they returned ("*I promise I will buy it from you*"). In the No Promise conditions, subjects are asked to imagine that they simply informed the seller of their plans without making any promises ("*All I can say is that I plan to buy it from you, though I can't promise that I will do so*"). In the 0% conditions, subjects are told that the seller is sure that the buyer will not buy the product from her; in the 50% conditions, they are told that the seller thinks there is a 50% chance; and in the 100% conditions, they are told that the seller is sure that the buyer will buy the product from her.[13]

[12] We use a hypothetical decision involving small stakes so as to avoid raising the spectre of legal liability in subjects minds. Our assumption is that it won't occur to subjects that they might be sued for breaking a promise to buy a product priced at just $100 without any express indication that this could be a possibility in the description of the scenario.

[13] See Appendix A for the text of the Vignette and Appendix E for screenshots of the instructions. There is arguably something incongruous about the 0%/Promise condition and the 100%/No Promise condition. It might seem odd to suppose that the seller would feel certain that the buyer wouldn't perform his promise given the limited information about the buyer that the seller apparently has. Likewise, it might seem odd to suppose that the seller would feel sure that the buyer would buy the good from her when the buyer expressly told the seller that he wasn't making any promises. But people vary in their degrees of optimism and pessimism, and so these conditions sim-

We employ a between-subject design whereby each subject is exposed to only *one* of the six conditions in order to minimize possible demand effects.[14] In each condition, subjects are asked to rate the likelihood that they will choose the product from the second seller.[15] Specifically, subjects are asked to select one of seven options: "No way," "Very unlikely," "Unlikely," "50:50, "Likely," "Very likely," and "With certainty." The advantage of framing the question as a likelihood, as opposed to simply asking subjects whether they will perform or not, is that it gives us a more continuous measure of subjects' willingness to perform. Arguably, this framing also makes it psychologically easier for a subject to reveal his preference for not performing than would be the case if a subject had to choose between definitely performing or definitely not performing. This may be particularly important in a study in which subjects face no monetary consequences from their decisions so that reporting a willingness to perform is a very cheap way for a subject to make himself feel good about himself.

A limitation of our design is that the scenario subjects are presented with is purely hypothetical and subjects are paid a fixed fee for completing the survey. Thus, they have no pecuniary incentive to answer honestly. But it is difficult to study the effects of promising and expectations in a controlled manner in an incentivized experiment in which there is real interaction between subjects and subjects choose whether or

ply capture beliefs formed by people on the extreme ends of the optimism or pessimism spectrum. And presumably when people say they are "sure" something will occur, what they mean to convey is that they believe it to be extremely likely–not that they believe that the alternative is literally impossible.

[14] Had we instead asked each subject to respond to all six conditions, it would have become apparent to subjects that we were studying the effects of promising and expectations, and subjects may distort their answers to conform to their beliefs about our hypotheses, or, more minimally, to create a false impression of consistency.

[15] See Wilkinson-Ryan and Hoffmann (2015) for a similar design.

not to make promises to one another, because it is difficult to manipulate subjects' expectations in a systematic fashion under these circumstances.[16] One possibility would be to configure the experiment so that different subjects ought rationally to form a different expectation about some event by manipulating the exogenous uncertainty that they face (Ederer and Stremitzer, 2017). But the event that is ultimately of interest here is not a move of nature. It is an action by a subject who may or may not have made a promise. And expectations about such actions will be informed by subjects' priors about the likelihood that people stick to their stated plans and/or keep their promises.

A second possibility would be to have subjects report their endogenously formed beliefs (Ellingsen et al., 2015). But we are not ultimately interested in a promisee's beliefs about the actions of the promisor (the seller's beliefs in our scenario). Rather, we are interested in the promisor's second-order beliefs about the expectations of the promisee (the buyer's beliefs about the expectations of the seller). And if the subjects' decisions had an effect on the payments they receive, we couldn't safely rely on the former to honestly report his expectations to the latter. This is because the promisee would have an incentive to overstate his expectations, if, as we hypothesize, the promisor is more likely to perform when he believes that the seller expects him to perform. One advantage of our vignette design is that we can circumvent these problems by simply telling promisors what the promisee expects them to do in each scenario.

[16]Ellingsen et al. (2011) use self-reporting of recipients' beliefs which admits the possibility of strategic communication. Reuben et al. (2009) use a multistage game, where the experimenter reports recipients' expectations from a previous stage game assuming expectations stay constant over time. Bellemare et al. (2017) compare different approaches to test for guilt aversion that have been used in the experimental literature to date.

A further advantage of our vignette design is that we can cleanly study the effect of promises by comparing a scenario in which the potential buyer makes a promise with a scenario in which the potential buyer makes a statement of intention but explicitly disclaims a promise. Previous experiments study scenarios in which subjects have the opportunity to communicate with one another and exchange promises with scenarios in which either subjects cannot communicate with one another at all (Charness and Dufwenberg, 2006) or potential promisors aren't able to communicate with those on the receiving end of their performance decisions (Vanberg, 2008). These experiments cannot cleanly disentangle the effects of promising from other effects of communication, and it is not clear how experiments that make promising endogenous would be able to do so. It might be possible to discern more subtle distinctions in subjects' communications in an experiment in which free-form communication is possible, though we are doubtful that it is easy to distinguish between promises and mere statements of intent, given that some statements of intent will imply (or be taken to imply) promises in context. Another possibility would be to give subjects the option of choosing between two pre-coded messages, e.g, "I promise to buy" and "I plan on buying but do not promise." But this would introduce selection effects, because there are likely to be systematic differences between subjects who make a promise and those who don't.

3.2 Hypotheses

We are now in a position to formulate the hypotheses that flow from our theory. First, the *promising per se effect* entails that subjects will report that they will be more likely to buy from the initial Seller B in the Promise condition than in the No

Promise condition in all three expectation conditions.

Hypothesis 1 *Subjects' reported likelihood of buying from Seller B in the Promise conditions will exceed their reported likelihood in the No Promise conditions for all levels of the Seller B's expectations (H1).*

Explanation. The *promising per se effect* means that promising makes the buyer more willing to buy from Seller B even when B has zero expectations of performance. We should observe reported likelihoods consistent with (H1) so long as some subjects have a Kantian disposition, since Kantians as we have defined them are inclined to do as they promised while exhibiting no particular disposition to be nice to others if they haven't made a promise.

Second, the *expectations per se effect* means that subjects will report that they are more likely to buy from Seller B the higher is her expectation that they will do so, regardless of whether a promise was made.

Hypothesis 2 *Subjects' reported likelihood of buying from Seller B will be greater, the higher are Seller B's expectations (H2).*

Explanation. The *expectations per se effect* means that the buyer becomes more willing to buy from Seller B the greater are B's expectations even if he made no promise. We should observe reported likelihoods consistent with (H2) so long as some subjects have a compassionate disposition, since compassionate agents don't like disappointing another's expectation even if there was no promise.

Finally, *the interaction effect* entails that subjects will report a greater increase in their willingness to buy from B as the seller's expectation increases when a promise was made.

Hypothesis 3 *An increase in Seller B's expectations cause a greater increase in subjects' reported likelihood of buying from Seller B in the Promise conditions than in the No Promise conditions (H3).*

Explanation. The *interaction effect* means that the buyer becomes more sensitive to Seller B's expectations when he made a promise. We should observe reported likelihoods consistent with (H3) so long as some subjects have a Scanlonian disposition, since Scanlonian agents care about not disappointing the expectations of others more when they feel responsible for those expectations as a result of having made a promise.

3.3 Procedure

We programmed the vignettes using Qualtrics and recruited 169 subjects from Amazon Mechanical Turk's pool of Master Workers and 614 from the general pool of MTurk workers who had a HIT approval rate of 95% or greater.[17] We determined our sample size using G Power analysis and a simulation based on pilot data.[18] Roughly

[17] We also conducted three pilot studies on 50 subjects who were recruited on Amazon Mechanical Turk. In the pilots we recruited only Mechanical Turk Master Workers. According to Amazon, these are "elite groups of Workers who have demonstrated accuracy on specific types of HITs [Human Intelligence Tasks] of a certain type with a high degree of accuracy across a variety of Requesters." Initially, we planned on restricting our final study to Master Workers who hadn't participated in the pilots. However, when we ran the study with this restriction, after an initial flurry of around 150 responses, the response rate slowed down to about one response per hour, suggesting that we had used up most of the pool of Master Workers. Thus, after getting 169 responses from Master Workers, we decided to eliminate the restriction and recruit 614 subjects from the general pool of MTurk Workers who had a HIT approval rate of 95% or greater.

[18] We conducted an a priori power analysis using G*Power (Faul, Erdfelder, Buchner, and Lang, 2009) that revealed a required sample size of $N = 713$ to detect a conservatively estimated small sized effect ($f2 = .02$) in a linear multiple regression ($n = 3$predictors) with a high power of $1 - \beta = .90$ and an alpha level of .05. We recruited a slightly larger sample of $N = 773$.

34% of subjects were MTurk Master workers.[19] All recruited subjects completed the survey.

After subjects had responded to the vignette, they were asked several control questions to ascertain whether they had understood the scenario. We also asked subjects other questions that were designed to assess how carefully and honestly they answered the questions. We used these questions to create additional robustness checks for our results. Finally, we asked subjects questions to ascertain their demographic characteristics.[20]

Before subjects proceeded to the main part of the experiment we announced that the task would take around 5-7 minutes and that we would pay subjects $1 for participating in the study. The announced hourly wage was therefore $9-12 per hour, which is well above the current national minimum wage ($7.25 per hour) and much higher than the wages paid in typical MTurk Studies.[21] On average, our subjects actually took 4 minutes and 16 seconds so that the effective average hourly wage was $14 per hour.

[19]These are the 169 subjects recruited when we restricted recruitment to Master Workers. In addition, 97 of those 614 workers from the general pool self-identified as mTurk Master Workers by answering affirmatively the following question in the post experiment questionnaire: "Are you an mTurk Master Worker (your response to this questions will have no effect on your payout)? (Yes/No/ I don't know what an mTurk Master Worker is.)" We classified all subjects who answered "Yes" as mTurk Master Workers.

[20]The questions of the post experiment survey can be found in Appendix B along with subjects' responses. Appendix E contains the screenshots.

[21]Horton and Chilton (2010) find a median hourly wage of $1.38 and Mason and Watts (2009) report a typical payment of $0.01-$0.10 per Human Intelligence Task (HIT).

4 Results

4.1 Descriptive Presentation of Data

Table 1 and Figure 1 summarize the means and medians of the choice variables by treatment condition for all of our subjects excluding those who incorrectly answered at least one of the control questions.[22] As a descriptive matter, our data is in line with our hypotheses. First, consistent with hypothesis (H1) and the *promising per se effect*, the mean and median likelihood of performance are higher in the Promise condition than in the No Promise condition for all levels of seller B's expectations. Thus, in Figure 1, the lines representing the mean and median likelihood of performance in the Promise condition are higher than the corresponding lines for the No Promise condition. Crucially, this also is the case when expectations are zero. Second, consistent with hypothesis (H2) and the *expectations per se effect*, the mean and median likelihood of performance are higher the greater are seller B's expectations in the Promise conditions and, crucially, even in the No Promise conditions (though only weakly so for the median in the No Promise condition).[23] Thus, in Figure 1 the lines representing the mean and median likelihood of performance in both conditions are upward sloping (except the line for the median in the No Promise condition, which is flat between expectations 50% and 100%). Finally, consistent with hypothesis (H3) and the *interaction effect*, the mean likelihood of performance increases more if there

[22] All our reported statistical results hold irrespective of whether we include or exclude participants who incorrectly answered the control questions, whether we include or exclude our pilot data, whether we include or exclude MTurk Master Workers, and whether we include or exclude those participants who answered our post-experiment survey questions in a way that makes us doubt they took the experiment seriously (see regression tables).

[23] However, the fact that the median in the No Promise condition is exactly the same for expectations 50% and 100% is likely an artifact of the discontinuous scaling of our choice variable.

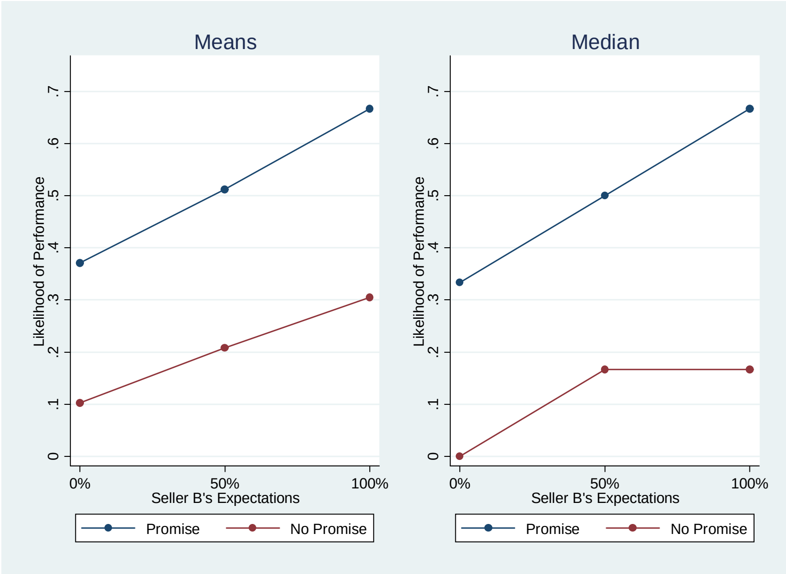


Figure 1: Mean and Median Likelihood of Performance Across Conditions.

was a promise than if there was none. Thus, in Figure 1, the lines representing the mean and median likelihood of performance in the Promise conditions are steeper than the corresponding lines for the No Promise conditions. The graphs suggest that this effect is even more pronounced for the median likelihood of performance.

Figure 2 provides a more complete picture of our results by showing the full distributions of the choice variable by treatment condition. Consistent with the above description of the effects on means and medians, there is more probability mass at higher ends of the distribution in the Promise condition than in the No Promise condition. And as expectations increase, probability mass is shifted to the right in both the Promise and No Promise conditions. Finally, Figure 2 illustrates the nature of the interaction effect. In the No Promise condition, as expectations increase, all the distributions remain skewed to the right, and the increase in the

Table 1: Mean and Median Likelihoods of Cooperation

	Expectations		
	0%	50%	100%
No Promise			
Mean	.10	.21	.30
Median	.00	.17	.17
	(N=120)	(N=105)	(N=116)
Promise			
Mean	.37	.50	.67
Median	.33	.50	.67
	(N=106)	(N=123)	(N=135)

mean willingness to perform arises because more probability weight is shifted to the right-hand tail as expectations increase. By contrast, in the Promise condition, the skewness of the distribution shifts from right to left. This illustrates why the medians show a more pronounced interaction effect than the means. Comparing the distributions for expectations of 100% also suggests that outliers are likely to dampen the statistical significance of the interaction effect in statistical tests based on sample means as the distribution in the No Promise condition is skewed to the right, while the distribution in the Promise condition is skewed to the left.

4.2 Baseline Specification

We use a number of different statistical tests to test our hypotheses. First, as a baseline, we used a standard OLS regression model to test all three of our hypotheses. Second, even though violations of normality assumptions about the distribution variables in a standard OLS model are not generally seen as a problem when, as here, the sample size is large, we ran the non-parametric Wilcoxon ranksum test as a robustness check, along with the t-test, to test hypotheses (H1) (the *promising per*

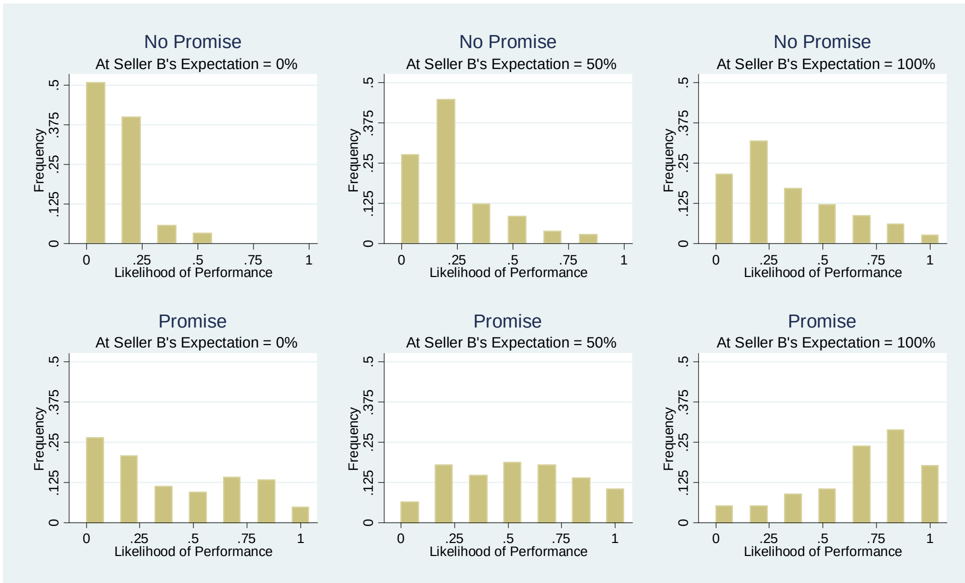


Figure 2: Distribution of Responses Across Conditions.

se effect) and (H2) (the *expectations per se effect*).[24] Third, we performed a bootstrapping procedure which enabled us to employ a non-parametric test of hypothesis (H3) (the *interaction effect*). Fourth, we ran a robust regression in order to deal with concerns about the influence of outliers on our results. Fifth, because we only have data for three discrete levels of expectations, we ran a categorical regression.[25] The discrete nature of our expectations variable is unlikely to be a problem for the standard OLS model given that as a conceptual matter our data is interval data. But we nonetheless ran the categorical regression as a robustness check. Finally, to analyze

[24]These two tests are unavailable for the interaction effect (H3) given our between-subject design. To test the interaction effect we would have to compare the difference in cooperation rates for different expectation levels in the Promise and the No Promise condition. The Wilcoxon ranksum test would allow us to test for the difference between unmatched data but, as we have to test for the difference in differences, we would need within-subject data for different levels of expectation to conduct this test.

[25]This test is equivalent to the ANOVA F-test.

more subtle distributional effects we ran a quantile regression on the medians.

Table 2 in Appendix C reports the results from the standard OLS model:

$$b = \alpha * promise + \beta * expectation + \gamma * promise \times expectation + \varepsilon.$$

We find that promising raises the average likelihood of performance by 26 percentage points even if expectations are zero. This effect is statistically significant at the 1% level. Moreover, promising increases the average likelihood of performance at each level of expectations and most notably for zero expectations (Wilcoxon ranksum test, $p < 0.01$; t-test, $p < 0.01$).[26] This is strong evidence in support of our hypothesized *promising per se effect* as embodied in hypothesis (H1).

Subjects are also more likely to perform if the seller has higher expectations even in the absence of a promise. Even in the No Promise condition, a shift in expectations from 0 to 100% increases the average likelihood of performance by 20 percentage points. This effect is statistically significant at the 1% level. Moreover, for both the Promise and, more crucially, the No Promise condition, the average likelihood of performance increases with greater expectations for each pairwise comparison (Wilcoxon ranksum test, $p < 0.01$; t-test, $p < 0.01$).[27] This strongly supports our hypothesized *expectations per se effect* as embodied in hypothesis (H2).

Finally, in line with our hypothesized *interaction effect* as embodied in hypothesis (H3), a shift of expectations from 0% to 100% increases the average likelihood of performance by 9 percentage points more in the Promise condition compared to

[26] Z values for the Wilcoxon ranksum test are 6.4, for 7.8, and 8.7 for expectation levels 0, 50%, 100% , respectively.

[27] In the Promise condition, Z values for the Wilcoxon ranksum test are 3.4, 4.3, 6.6 for the (0% 50%), (50% 100%), and (0% 100%) comparison, respectively. In the No Promise condition these values are 4.3, 2.7, 6.5. Tables 4 and 5 in Appendix C provide results from categorical regressions in the standard and the robust version.

the No Promise condition. This effect is statistically significant at the 5% level.[28] The non-parametric test of (H3) using a two-step bootstrapping procedure yields significance at the 10% level ($p = 0.07$).[29]

4.3 Alternative Specifications

We find that a robust regression yields even stronger results, suggesting that outliers are not driving our findings (see Table 3 in Appendix C). As we conjectured above when looking at the distribution of outcomes across treatments, the interaction effect becomes stronger ($\gamma = 0.19$) and significant at the 1% level.

One concern about applying standard OLS to our data is that it treats expectations as interval data. While it makes sense to treat expectations in this way as a conceptual matter, we only elicit subjects' responses for three different levels of expectations. Therefore, as a robustness check, we dummy coded the expectations variable and ran a categorical regression (see Table 4 in Appendix C). We find that the comparison between 0% and 100% is statistically significant ($\gamma = 0.19$, $p < 0.01$). The (0% 50%) and (50% 100%) comparisons are not significant ($p = 0.46$, $p = 0.22$). However, they go in the right direction. Moreover, our data are very well behaved as the estimated effects sizes are roughly half the (0% 100%) comparison, as we would expect. Hence, the lack of significance is likely simply due to a lack of power: more data is needed to find a statistically significant effect when the effect is smaller.[30]

[28] The lower significance level is to be expected, given that we need higher statistical power to detect an interaction effect. See note 32 for a discussion of the possibility that a boundary effect may have caused us to underestimate the size of the interaction effect.

[29] The bootstrapping procedure requires us to simulate synthetic samples and is discribed in detail in Appendix D.

[30] Similarly, a joint test of the equality of each of the treatments, which is statistically equivalent to the ANOVA F-test, is not significant in our leading data pool (p = 0.14, see Column 1 of Table

Table 6 in Appendix C reports results from a quantile regression run at the median.[31] Interestingly, we find that for the median subject the entire effect of expectations on subjects' likelihood of performing works through the *interaction effect*, effectively eliminating the *expectations per se effect.* This reveals a subtle feature of our data. In the No Promise condition, a minority of subjects respond to higher expectations, but not a majority—hence, the nonmoving median. In the Promise condition, a majority of subjects respond to higher expectations, but there is a sizable minority that doesn't—hence, the bigger movement of the median than the mean.[32]

Finally, the two-step bootstrapping exercise described in Appendix D allows for a non-parametric test of hypothesis (H3) (the *interaction effect*) and yields statistical significance at the 10% level ($p = 0.07$).

---

4 in Appendix C). However, the joint test may be too conservative as it treats the fact that we see no statistically significant difference in the (0% 50%) comparison as equally as troubling as if we saw no statistically significant difference in the (0% 100%) comparison. Still, running a robust version of the categorical regression restores significance on all fronts (see Table 5 in Appendix C, reported p-values are equivalent to the robust version of the ANOVA F-test).

[31] As mentioned above, it is important to be careful when interpreting quantile regressions on an output variable that is not continuous. However, we checked and found that the coefficients are relatively stable around the median.

[32] That a sizeable minority does not respond to expectations in the Promise condition might be an artifact of our design. Performance is bounded above. Subjects cannot perform at a higher level than performing "for sure." So, for those subjects who would already perform with a high likelihood for low expectations because they feel duty-bound to honor their promise no matter what, there is not much scope for increasing their likelihood of performance as promisees' expectations increase. Our data is consistent with this explanation. A quantile regression run at the first quartile has the same coefficient for the interaction effect as a quantile regression run at the median.($\gamma = .33$, $p < .01$). However, the coefficient of the interaction effect for a quantile regression run at the third quartile becomes negative ($\gamma = -.17$, $p < .01$). The coefficients become even more negative for regressions on higher quantiles (See Table 7 in Appendix C). This possible boundary effect also suggests that we underestimate the size of the mean interaction effect.

5 Discussion

Our results provide a more nuanced picture of the determinants of promise keeping than the experimental literature has done to date. We are, to our knowledge, the first to provide clear and direct evidence for a promising per se effect. The previous literature, notably Charness and Dufwenberg (2006) and Vanberg (2008), provides evidence that a dictator's promise to a recipient in a dictator game increases the likelihood that she will perform. However, both papers document this effect for the cases where the recipient's performance expectations are positive. It would therefore be wrong to infer from the previous literature that promises matter irrespective of promisees' expectations. Indeed, if dictators only care about a recipient's expectations if they have made a promise, the increased performance levels that arise when there was a promise could be explained entirely by the interaction effect. In other words, it is not clear from the previous literature that promises have any effect in the absence of positive expectations. A Scanlonian promisor wouldn't feel duty-bound to keep his promises if the promisee is sure he won't perform, as Scanlon's "profligate pal" example shows. To empirically isolate the promising per se effect it is therefore crucial to show that promises matter even if expectations are zero, which we are able to demonstrate.[33]

Our finding that expectations matter even in the absence of promising is in line with the theory proposed by Charness and Dufwenberg (2006) and results by Reuben

[33]It is inherently difficult to create zero expectations in the recipient of a promise in a controlled fashion if this promise arises naturally out of the interaction of two experimental subjects. This is presumably the reason why this result has remained elusive up until now. Our vignette study enables us to demonstrate the existence of a pure *promising per se effect* because we simply tell our promisors that the promisee's expectations are zero.

et al. (2009), Bellemare et al. (2011) and Regner and Harth (2014),[34] while it runs counter to Vanberg (2008) and Ellingsen et al. (2010), who find evidence that expectations are not independently significant.[35]

Our findings also contradict the strong version of the interaction effect that is embodied in Ederer and Stremitzer's (2017) theory of conditional guilt aversion. But the interaction effect that we identify supports a weak version of this theory, since promising makes a party more sensitive to her counterparty's expectations.[36] The interaction effect therefore accounts for the fact that the previous literature has found clear evidence for the relevance of expectations if there was a promise (Ederer and Stremitzer, 2017), but, as mentioned above, at best mixed evidence for their relevance if there was no promise. This paper is the first to provide clear empirical evidence for the presence of an interaction effect.[37]

[34] However, Reuben et al. (2009) employ a lost wallet game, where arguably there is a preexisting duty to return the wallet, which could have a similar effect to a promise.

[35] However, Vanberg (2008) isolates expectations from promising, by destroying the promissory link between the subjects but, in doing so, destroys any link between the two parties. Therefore, he really compares a relationship of promising with no relationship at all. Likewise, in Ellingsen et al. (2010) there is no relationship between the parties. Parties play a dictator game in which the experiments communicate the recipients' elicited expectations to their dictators (unbeknownst to the recipients).

[36] Morell (2014) finds that dictators care more about recipients' expectations if they are perceived as "ingroup" rather then "outgroup." It may be that the main effect of promising is to create a sense of responsibility in the dictator similar to the sense of responsibility a dictator might feel for an ingroup recipient.

[37] It is possible that subjects make implicit assumptions about the extent to which the seller relies on the buyer's statement and that those assumptions vary systematically across our six conditions such that our results are driven by subjects' different beliefs about the seller's reliance rather than by the variables we are interested in. While we cannot rule out this possibility, we are doubtful that differential assumptions about reliance can explain the observed interaction effect or the promise per se effect (though it might be that the expectations per se effect is really an "expectations and reliance per se effect"). This is because a rational seller's reliance ought to be a function of her beliefs about the buyer's behavior, and while it is true that promising often makes a promisee more confident that the promisor will perform, our vignettes tell subjects what the seller's beliefs are. Holding constant those beliefs, the buyer has no reason to suppose that the seller will rely more when he made a promise, at least assuming that the buyer assumes that the seller is rational.

A notable feature of our design is that we ask subjects to imagine that parties communicate with one another even when the buyer makes no promise. This suggests that, holding expectations constant, it is the promise, and not merely the fact that the buyer and seller communicated with one another, that increases the buyer's willingness to perform in the Promise condition. But this of course means that we can't determine whether expectations matter independently of communication.[38] In order to determine whether expectations matter independently of communication, we would need to construct a vignette in which a party learns his counterparty's expectations of performance in the absence of any prior communication between the parties.[39]

6 Conclusion

Our paper provides a unified framework for studying the effect of promises and expectations on performance. We are able to document a *promising per se effect*, according to which promises matter regardless of the promisee's expectations, an *expectations per se effect*, according to which expectations matter even in the absence of a promise, and an *interaction effect*, according to which promising makes a promisor more sensitive to a promisee's expectations.

However, our between-subject data don't allow us to determine whether our findings result from the presence of different pure types in the subject population (Compassionate, Kantian, Scanlonian), or by subjects who exhibit all three dispositions in

[38] Bicchieri and Sontuoso (2015) provide some evidence that suggests that the fact of communication could be important here.

[39] The evidence presented by Reuben et al. (2009) suggests that recipients' expectations in a dictator game might be important even absent communication.

a weighted combination. In order to understand better the composition of the subject population we would need within-subject data–that is, data obtained by asking every subject to consider what they would do in all six conditions. A problem with eliciting such data, and one of the motivations behind our between-subject design, is that it introduces the problem of experimenter demand effects.[40] Nonetheless, such data may give us important information about the composition of the subject population. Pilot data obtained from such a within-subject design suggests that there is indeed considerable heterogeneity. 62% percent of our subjects conform to our model.[41] Of these subjects, 58% exhibit a Compassionate disposition, 75% exhibit a Kantian disposition, and 39% exhibit a Scanlonian disposition over at least part of the range. Only 22% of these subjects exhibit all three dispositions. We leave a more systematic inquiry into this heterogeneity of subject types and the possible relationship of those types to personality traits discussed in the psychology literature to future research.

References

Battigalli, P., and M. Dufwenberg (2007): "Guilt in games," *The American Economic Review*, 97(2), 170–176.

[40]Presenting subjects with all six conditions makes salient to subjects' the variables that we are studying, raising the concern that subjects' responses will be distorted. Subjects might for example modify their answers in order to conform with their beliefs about the researcher's hypotheses or to create a false impression of consistency. See Charness, Gneezy, and Kuhn (2012).

[41]This means that they exhibit a likelihood of performance that is: (i) weakly increasing in expectations; (ii) weakly greater when there is a promise; and (iii) increasing at a weakly greater rate in expectations when there is a promise. Most of the remaining subjects behave irregularly on at least one of the three dimensions, while the remainder are contrarian on at least one of the three dimensions by exhibiting a likelihood of performance that is either (i) strictly decreasing in expectations over at least part of the range; (iii) strictly lower when there is a promise over at least part of the range; or (iii) increasing at a strictly lower rate in expectations when there is a promise.

——— (2009): "Dynamic psychological games," *Journal of Economic Theory*, 144(1), 1–35.

Benabou, R., and J. Tirole (2003): "Intrinsic and extrinsic motivation," *The Review of Economic Studies*, 70(3), 489–520.

Bicchieri, C., and A. Sontuoso (2015): "I cannot cheat on you after we talk," *The Prisoner's Dilemma*, pp. 101–114.

Braver, S. L. (1995): "Social Contracts and the Provision of Public Goods," in *Social Dilemmas: Perspectives on Individuals and Groups*, ed. by D. Schroeder. Praeger, New York.

Bull, C. (1987): "The Existence of Self-Enforcing Implicit Contracts," *Quarterly Journal of Economics*, 102(1), 147–160.

Charness, G., and M. Dufwenberg (2006): "Promises and Partnership," *Econometrica*, 74, 1579–1601.

Charness, G., U. Gneezy, and M. A. Kuhn (2012): "Experimental methods: Between-subject and within-subject design," *Journal of Economic Behavior & Organization*, 81(1), 1–8.

Dixit, A. (2009): "Governance institutions and economic activity," *The American Economic Review*, 99(1), 3–24.

Dufwenberg, M., and U. Gneezy (2000): "Measuring Beliefs in and Experimental Lost Wallet Game," *Games and Economics Behavior*, 30(2), 163–182.

Dufwenberg, M., A. Smith, and M. Van Essen (2013): "Hold-Up: With a Vengeance," *Economic Inquiry*, 51(1), 896–908.

Ederer, F., and A. Stremitzer (2017): "Promises and Expectations," *Games and Economic Behavior*, 106, 161–178.

Efron, B., and R. J. Tibshirani (1993): "An introduction to the bootstrap Chapman & Hall," *New York*, 436.

Ellingsen, T., and M. Johannesson (2004): "Promises, Threats and Fairness," *Economic Journal*, 114(495), 397–420.

Ellingsen, T., M. Johannesson, S. Tjøtta, and G. Torsvik (2010): "Testing Guilt Aversion," *Games and Economic Behavior*, 68(1), 95–107.

FAUL, F., E. ERDFELDER, A. BUCHNER, AND A.-G. LANG (2009): "Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses," *Behavior research methods*, 41(4), 1149–1160.

GROSSMAN, S. J., AND O. D. HART (1986): "The costs and benefits of ownership - A theory of vertical and lateral integration," *Journal of Political Economy*, 94(4), 691–719.

GROUT, P. (1984): "Investment and Wages in the Absence of Binding Contracts: A Nash Bargaining Approach," *Econometrica*, 52 (2), 449–460.

HART, O., AND J. MOORE (1988): "Incomplete Contracts and Renegotiation," *Econometrica*, 56(4), 755–785.

HOLMSTRÖM, B. (1979): "Moral Hazard and Observability," *The Bell Journal of Economics*, 10, No. 1, 74–91.

HORTON, J. J., AND L. B. CHILTON (2010): "The Labor Economics of Paid Crowdsourcing," in *Proceedings of the 11th ACM Conference on Electronic Commerce*, EC '10, pp. 209–218, New York, NY, USA. ACM.

ISMAYILOV, H., AND J. J. M. POTTERS (2012): "Promises as Commitments," *Tilburg University, Center for Economic Research, Discussion Paper*, 2012-064.

KLEIN, B., AND K. B. LEFFLER (1981): "The role of market forces in assuring contractual performance," *The Journal of Political Economy*, pp. 615–641.

KREPS, D. M. (1996): "Corporate culture and economic theory," *Firms, Organizations and Contracts, Oxford University Press, Oxford*, pp. 221–275.

LEVIN, J. (2003): "Relational Incentive Contracts," *American Economic Review*, 93(3), 835–857.

MACAULAY, S. (1963): "Non-Contractual Relations in Business: A Preliminary Study," *American Sociological Review*, 28(1), 55–70.

MACLEOD, B., AND J. MALCOMSON (1989): "Implicit Contracts, Incentive Compatibility, and Involuntary Unemployment," *Econometrica*, 57, 447–80.

MASON, W., AND D. J. WATTS (2010): "Financial incentives and the performance of crowds," *ACM SigKDD Explorations Newsletter*, 11(2), 100–108.

MIRRLEES, J. (1976): "The Optimal Structure of Authority and Incentives Within an Organization," *Bell Journal of Economics*, 7, 105–31.

Morell, A. (2015): "The Short Arm of Guilt: Guilt Aversion Plays Out More Across a Short Social Distance," *Preprints of the Max Planck Institute for Research on Collective Goods*.

Ostrom, E., J. Walker;, and R. Gardner (1992): "Covenants With and Without a Sword: Self-Governance Is Possible," *American Political Science Review*, 86(2), 404–417.

Reuben, E., P. Sapienza, and L. Zingales (2009a): "Is Mistrust Self-Fulfilling," *Economics Letters*, 104, 89–91.

——— (2009b): "Is mistrust self-fulfilling?," *Economics Letters*, 104(2), 89–91.

Scanlon, T. (1998): *What We Owe to Each Other*. Cambridge MA, Harvard University Press.

Shiffrin, S. V. (2008): "Promising, Intimate Relationships, and Conventionalism," *Philosophical Review*, 117 No. 4, 481–524.

Stone, R., and A. Stremitzer (2016): "Promises, Reliance, and Psychological Lock-in," *UCLA School of Law, Law-Econ Research Paper No. 15-17*.

Vanberg, C. (2008): "Why Do People Keep Their Promises? An Experimental Test of Two Explanations.," *Econometrica*, 76, 467–1480.

Wilkinson-Ryan, T., and D. A. Hoffman (2015): "The Common Sense of Contract Formation," *Stanford Law Review*, 67, 1269–1301.

Williamson, O. E. (1979): "Transaction-Cost Economics: The Governance of Contractual Relations," *Journal of Law and Economics*, 22, 233–61.

——— (1985): *The Economic Institutions of Capitalism*. Free Press, New York.

APPENDIX A: VIGNETTES

Seller B is offering a product for sale for $100 that you are interested in buying. You are currently out of town for three days and therefore unable to go to B's shop and buy the product immediately. But B may have the opportunity to sell the product to somebody else in the meantime, so you promise B that you will buy the product upon your return. The conversation proceeds as follows:

B says: "*I would be willing to sell the product to you, but someone else might offer to buy it in the meantime. Why should I wait to sell the product to you?*"

You say: "*Well, I promise I will buy it from you upon my return.*"

B immediately concludes that there is a [0%/50%/100%] chance that you are going to keep your promise to buy the product from him. Imagine you know this.

On the day you want to buy the product from B, you accidentally learn that another seller (C) is offering to sell an equivalent product at the price of $85, which is $15 less than the price that B is charging.

So the situation is this: C is offering to sell you the product at a lower price. **You have made a promise to B** to buy the product from him and the product is still available. **You also know that [B is sure that you will not/B thinks there is a 50% chance that you will/B is sure that you will**] keep your promise to buy the product from him.

How likely is it that you would choose to buy the product from **the second seller C** in this scenario?

Here are the scenarios in which the buyer makes no promise:

Seller B is offering a product for sale for $100 that you are interested in buying. You are currently out of town for three days and therefore unable to go to B's shop and buy the product immediately. But B may have the opportunity to sell the product to somebody else in the meantime. The conversation proceeds as follows:

B says: "*I would be willing to sell the product to you, but someone else might offer to buy it in the meantime. Why should I wait to sell the product to you?*"

You respond: "*All I can say is that I plan to buy it from you, though I can't promise that I will do so.*"

B immediately concludes that there is a [0%/50%/100%] chance that you are going to buy the product from him. Imagine you know this.

On the day you want to buy the product from B, you accidentally learn that another seller (C) is offering to sell an equivalent product at the price of $85, which is $15 less than the price that B is charging.

So the situation is this: C is offering to sell you the product at a lower price. **You have made no promise to B** to buy the product from him but the product is still available. **You also know that [B is sure that you will not/B thinks that there is a 50% chance that you will/B is sure that you will] buy the product from him.**

How likely is it that you would choose to buy the product from **the second seller C** in this scenario?

APPENDIX B: POST EXPERIMENT QUESTIONNAIRE

The control questions were:

1) In the given scenario, did you make a promise to B to buy his product? (Yes/No),

2) Please indicate the expectations B had whether you are going to buy his product. (0%/ 50%/ 100%);

3) Who made the better offer? (the first seller B, the second seller C).

78 out of 783 answered at least one of these control question incorrectly. We use subjects' responses to these questions to construct a robustness check in our statistical tests below.

The questions to assess carefulness and honesty of subjects were:

1) I didn't take the scenario seriously. I just wanted to earn the $1.00 fee as quickly as possible. (Yes: 12 out of 783);

2) I carefully read the instructions. (No: 4 out of 783);

3) I chose my answers to make myself seem like a good person. (Yes: 69 out of 783);

4) This is the first time I have completed this survey. (No: 16 out of 783). We have no good explanation for these 16 subjects who self-reported having taken the survey before. We provided links to participants which were only good for a single log in. We implemented filters preventing subjects (as identified by their MTurk IDs) from participating who had participated in pilots of our experiment or similar experiments we had run in the past. So the only reason for the 16 self-reported repeat takers could be that subjects have multiple MTurk IDs or mistakenly checked the wrong box.

The demographic questions were:

1) What is your age? (age was between 18 and 74 with and average age of 35)

2) What is you gender? (49% female)

3) What is your highest level of schooling? (Master's degree or more: 11 %; Bachelor's degree: 41%; Associate's degree: 16%; Vocational or technical certificate/ diploma after high school (such as cosmetics): 7%; Highschool diploma: 24%; I did not complete Highschool: 1%)

4) Is English your first language? (Yes: 98%).

APPENDIX C: TABLES

Table 2: Standard OLS Regressions

	(1)	(2)	(3)	(4)
Promise	.26***	.24***	.25***	.26***
	(.03)	(.03)	(.04)	(.03)
Expectations	.20***	.18***	.21***	.21***
	(.03)	(.03)	(.04)	(.04)
Promise × Expectations	.09**	.11**	.12**	.08*
	(.05)	(.04)	(.06)	(.08)
Cons	.10***	.13***	.10***	.10***
	(.02)	(.02)	(.04)	(.02)
R^2	.36	.33	.36	.33
N	705	783	486	650

Note: Standard Errors in Parentheses. $*, **, ***$ indicate significance at the 10%, 5%, 1% level, respectively. Column (1): Data excluding those who failed control questions. Column (2): Data including those who failed control questions. Column (3): Data excluding those who failed control questions and master workers. Column (4): Data excluding those who failed control questions and those who did not pass the filter constructed on the basis of the post experiment survey.

Table 3: Robust Regressions

	(1)	(2)	(3)	(4)
Promise	.24***	.22***	.21***	.23***
	(.03)	(.03)	(.04)	(.03)*
Expectations	.18***	.16***	.18***	.18***
	(.03)	(.03)	(.04)	(.04)
Promise × Expectations	.19***	.19***	.24***	.18***
	(.05)	(.05)	(.06)	(.05)
Cons	.10***	.12***	.10***	.10***
	(.02)	(.02)	(.03)	(.02)
N	705	783	486	650

Note: Standard Errors in Parentheses. $*, **, ***$ indicate significance at the 10%, 5%, 1% level, respectively. Column (1): Data excluding those who failed control questions. Column (2): Data including those who failed control questions. Column (3): Data excluding those who failed control questions and master workers. Column (4): Data excluding those who failed control questions and those who did not pass the filter constructed on the basis of the post experiment survey.

Table 4: Categorical Regressions

	(1)	(2)	(3)	(4)
Expectations per se				
(0 100)	$.17^{***}$	$.18^{***}$	$.21^{***}$	$.21^{***}$
(0 50)	$.09^{***}$	$.09^{***}$	$.10^{**}$	$.12^{***}$
(50 100)	$.06^{***}$	$.09^{***}$	$.11^{**}$	$.09^{**}$
Joint Test	$p < .01^{***}$	$p < .01^{***}$	$p < .01^{***}$	$p < .01^{***}$
Interaction Effect				
(0 100)	$.09^{**}$	$.11^{**}$	$.12^{**}$	$.08^{*}$
(0 50)	.04	.04	.07	.02
(50 100)	.06	.07	.12	.07
Joint Test	$p = .14$	$p < .05^{**}$	$p < .13$	$p = .22$
N	705	783	486	650

Note: Standard Errors in Parentheses. $*$, $**$, $***$ indicate significance at the 10%, 5%, 1% level, respectively. Column (1): Data excluding those who failed control questions. Column (2): Data including those who failed control questions. Column (3): Data excluding those who failed control questions and master workers. Column (4): Data excluding those who failed control questions and those who did not pass the filter constructed on the basis of the post experiment survey.

Table 5: Robust Categorical Regressions

	(1)	(2)	(3)	(4)
Expectations per se				
(0 100)	$.17^{***}$	$.16^{***}$	$.19^{***}$	$.18^{***}$
(0 50)	$.09^{***}$	$.08^{**}$	$.09^{**}$	$.10^{***}$
(50 100)	$.08^{**}$	$.08^{**}$	$.10^{**}$	$.08^{**}$
Joint Test	$p < .01^{***}$	$p < .01^{***}$	$p < .01^{***}$	$p < .01^{***}$
Interaction Effect				
(0 100)	$.19^{***}$	$.18^{***}$	$.23^{***}$	$.17^{***}$
(0 50)	$.08^{*}$	.07	$.12^{**}$	.06
(50 100)	$.11^{**}$	$.11^{**}$	$.11^{*}$	$.11^{**}$
Joint Test	$p < .01^{***}$	$p < .01^{***}$	$p < .01^{***}$	$p < .01^{***}$
N	705	783	486	650

Note: Standard Errors in Parentheses. $*, **, ***$ indicate significance at the 10%, 5%, 1% level, respectively. Column (1): Data excluding those who failed control questions. Column (2): Data including those who failed control questions. Column (3): Data excluding those who failed control questions and master workers. Column (4): Data excluding those who failed control questions and those who did not pass the filter constructed on the basis of the post experiment survey.

Table 6: Quantile Regressions at Median

	(1)	(2)	(3)	(4)
Promise	.17***	.17***	.17***	.17***
	(.06)	(.05)	(.06)	(.06)*
Expectations	.00	.00	.00	.00
	(.06)	(.05)	(.07)	(.06)
Promise × Expectations	.33***	.33***	.33***	.33***
	(.09)	(.07)	(.09)	(.09)
Cons	.17***	.17***	.17***	.17***
	(.04)	(.04)	(.04)	(.04)
N	705	783	486	650

Note: Standard Errors in Parentheses. *, **, * * * indicate significance at the 10%, 5%, 1% level, respectively. Column (1): Data excluding those who failed control questions. Column (2): Data including those who failed control questions. Column (3): Data excluding those who failed control questions and master workers. Column (4): Data excluding those who failed control questions and those who did not pass the filter constructed on the basis of the post experiment survey.

Table 7: Quantile Regressions at Different Quantiles

	Quantiles				
	.25	.5	.75	.85	.95
Promise	.00	.17**	.50***	.67***	.67***
	(.04)	(.06)	(.04)	(.05)	(.07)
Expectations	.17***	.00	.33***	.50***	.50***
	(.05)	(.06)	(.04)	(.06)	(.07)
Promise × Expectations	.33***	.33***	−.17**	−.33***	−.50***
	(.07)	(.09)	(.06)	(.08)	(.10)
Cons	.00	.17***	.17***	.17***	.33***
	(.03)	(.04)	(.03)	(.04)	(.05)
N	705	705	705	705	705
Pseudo R^2	0.16	0.27	0.30	0.26	0.22

Note: Standard Errors in Parentheses. $*$, $**$. $***$ indicate significance at the 10%, 5%, 1% level respectively. Data excluding those who failed control questions.

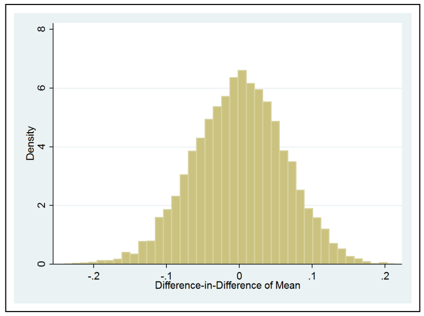
APPENDIX D: BOOTSTRAP

We performed a two-step bootstrapping procedure in order to employ a non-parametric test of H3 (the *interaction effect*). Let $\hat{\theta}_e$ and θ_e be the estimators of the mean likelihood of performance in the Promise and No Promise conditions respectively when the promisee's expectations are e. We observe that $\left(\hat{\theta}_1 - \theta_1\right) - \left(\hat{\theta}_0 - \theta_0\right) = 0.1$. That is, the difference in means between the Promise and the No Promise samples is higher if promisees' expectations are 100% than if they are 0%. We want to know the probability with which we would observe this positive difference-in-difference of means by chance. In other words, we want to test the null hypothesis $\left(\hat{\theta}_1' - \theta_1'\right) - \left(\hat{\theta}_0' - \theta_0'\right) = 0$, where $\hat{\theta}_1', \theta_1', \hat{\theta}_0', \theta_0'$ are the means of the underlying distributions from which our samples are drawn.

We can do so in two steps (see, e.g., Efron and Tibshrani, 1993, pp. 220-223). First, we recenter the original samples to conform with the null hypothesis. Specifically, we subtract from each observation in each of the four samples the respective sample means and then add the mean effect of promising to each observation in the two Promise samples. In other words, if the mean for the combined No Promise samples is $\bar{x}$ and the mean for the combined Promise samples is $\bar{y}$, we add $(\bar{y} - \bar{x})$ to each observation in the two Promise samples.[42]

We then create four synthetic samples – of sample sizes equal to our real samples – by randomly drawing with replacement from each of the four samples. We can then calculate the difference-in-difference, $\left(\hat{\theta}_1'' - \theta_1''\right) - \left(\hat{\theta}_0'' - \theta_0''\right)$ where $\hat{\theta}_1'', \theta_1'', \hat{\theta}_0'', \theta_0''$ are the means of these synthetic samples. After 10,000 iterations, we obtain a simulated distribution of the differences-in-differences of the means that would arise if the null hypothesis were true (that is, if the difference of means between the Promise and the No Promise conditions was equal across different levels of expectations.

[42] By subtracting the sample means, we make our data conform to the hypothesis $\hat{\theta}_1' = \theta_1' = \hat{\theta}_0' = \theta_0' = 0$. In doing so, we eliminate all three of our hypothesized effects from our data. By adding back $(\bar{y} - \bar{x})$ to the observations in the promise samples, we effectively add back the *promise per se effect,* so that our data ends up conforming to our less restrictive null hypothesis $\left(\hat{\theta}_1' - \theta_1'\right) - \left(\hat{\theta}_0' - \theta_0'\right) = 0$. We don't add back in the *expectations per se effect*, since doing so leaves this hypothesis unchanged.



Simulated distribution.

The area under the curve to the right of the observed estimator 0.1 corresponds to the probability that a greater or equal difference-in-difference would have been observed if the null hypothesis were true. This value, 0.07, is small enough to permit rejection of the null hypothesis at the 10% level.