Worth Keeping but Not Exceeding: Asymmetric Consequences of Breaking Versus Exceeding Promises

Ayelet Gneezy and Nicholas Epley

Social Psychological and Personality Science 2014 5: 796 originally published online 8 May 2014
DOI: 10.1177/1948550614533134

Social Psychological and
Personality Science
2014, Vol. 5(7) 796-804

DOI: 10.1177/1948550614533134
spps.sagepub.com
SAGE

Worth Keeping but Not Exceeding: Asymmetric Consequences of Breaking Versus Exceeding Promises

Ayelet Gneezy[1] and Nicholas Epley[2]

Abstract

Promises are social contracts that can be broken, kept, or exceeded. Breaking one's promise is evaluated more negatively than keeping one's promise. Does expending more effort to exceed a promise lead to equivalently more positive evaluations? Although linear in their outcomes, we expected an asymmetry in evaluations of broken, kept, and exceeded promises. Whereas breaking one's promise is obviously negative compared to keeping a promise, we predicted that exceeding one's promise would not be evaluated more positively than merely keeping a promise. Three sets of experiments involving hypothetical, recalled, and actual promises support these predictions. A final experiment suggests this asymmetry comes from overvaluing kept promises rather than undervaluing exceeded promises. We suggest this pattern may reflect a general tendency in social systems to discourage selfishness and reward cooperation. Breaking one's promise is costly, but exceeding it does not appear worth the effort.

Keywords

judgment and decision making, social cognition, social judgment, interpersonal processes, impression formation

It is often said that promises are easier to make than to keep. Woodrow Wilson promised to stay out of World War I, Roosevelt promised the same for World War II, and George Bush Sr. promised "no new taxes." All were promises broken while still in office. Although notorious for breaking promises, politicians are not alone (Buehler, Griffin, & Ross, 1994; Robinson & Rousseau, 1994). Contractors may promise to complete projects on time and on budget but end up doing neither. Employers may promise working conditions that go unfulfilled. And friends may promise to help us exercise, clean the apartment, and pick up groceries, only to leave us hungry while exercising alone in a dirty apartment.

Clearly, not all promises are broken. Promises can be kept, or even exceeded. Bush could have reduced taxes, contractors could finish faster and cheaper, and employers could provide even better working conditions (Lester, Turnley, Bloodgood, & Bolino, 2002).

This research examines the social consequences of exceeding versus breaking promises, compared to keeping promises. We define a promise as a commitment to perform some specific action made by one person to another. Promises are different than having a belief or expectation about another's behavior because promises are inherently interpersonal, whereas a belief or expectation is intrapersonal. Existing research demonstrates many negative consequences from broken versus kept promises, including reduced trust, diminished satisfaction, and increased turnover (see Rousseau, 1995, for a review). However, existing research does not measure whether putting in additional effort to exceed a promise produces symmetrically more positive consequences than merely keeping a promise. Does doing more than one promised lead to even more positive evaluations than simply keeping one's promise?

We predicted it would not. In particular, we predicted that exceeding one's promise would not be evaluated more positively than merely keeping one's promise. This predicted asymmetry could emerge for two reasons.

First, keeping a promise could be *overvalued* compared to a linear relationship between outcomes and evaluations. Overvaluing kept promises is consistent with their contractual nature, where keeping a promise provides relational value of trust and reliability beyond the outcome's objective value. Indeed, social systems that place a premium on fairness may promote mutually beneficial cooperation within groups (Bowles & Gintis, 2003; Cosmides, 1989; Gneezy & Fessler, 2012; Tooby & Cosmides, 1996).

Second, exceeding a promise could be relatively *undervalued* compared to a linear relationship between outcomes and evaluations. This undervaluation is consistent with the gain/loss

[1] University of California San Diego, San Diego, CA, USA
[2] University of Chicago, Chicago, IL, USA

Corresponding Author:
Ayelet Gneezy, Rady School of Management, 3W119 Wells-Fargo Hall, 9500 Gilman Drive, La Jolla, CA 92093-0553, USA.
Email: agneezy@ucsd.edu or epley@chicagobooth.edu

asymmetry predicted by Prospect Theory (Kahneman & Tversky, 1979), in which gains from reference points (such as a promise) produce relatively little impact compared to losses from that reference point (Thaler, 1985). It is also consistent with the general tendency for negative outcomes to produce stronger psychological consequences than positive outcomes (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001; Rozin & Royzman, 2001). This may be particularly true in moral domains where a negative action, such as lying, is seen as more diagnostic of a person's character than a negative action is in competence (or ability) domains, such as performing poorly on a test (Martijn, Spears, Van der Pligt, & Jakobs, 1992; Skowronski & Carlston, 1987).

We tested the asymmetry prediction in three sets of experiments (Experiments 1a–3). A final experiment tested the mechanism for our observed asymmetry: Whether it occurs because keeping a promise is overvalued or because exceeding promises are undervalued (Experiment 4). Because promises arise in varying contexts that cannot be captured in any single paradigm, we adopted a multimethod approach utilizing hypothetical (Experiments 1a and 1b), remembered (Experiments 2a, 2b, and 4), and actual promises (Experiment 3). Each method has strengths and weaknesses. Hypothetical surveys examine contexts that cannot be recreated experimentally but utilize imagination rather than actual behavior. Memory is notoriously imperfect, but it benefits from ecological validity by examining actual events from real life. Experiments are the gold standard for causality, but a constructed laboratory context may also raise concerns about ecological validity. Testing our hypotheses using all of these methods increases confidence in the robustness of any observed results and addresses weaknesses from one methodology that is not contained in another.[1]

Experiments 1a and 1b: Imagined Promises

Participants in Experiments 1a and 1b read hypothetical scenarios in which they were asked to imagine receiving a promise that was exceeded, kept as promised, or broken. We used these scenarios to investigate how promise outcomes were likely to influence promise receivers' happiness and trust in the promise maker (1a), and the impressions that promise receivers would form of promise makers (1b).

Method

Experiment 1a Procedure

Sixty-two University of California San Diego (UCSD) undergraduates read a one-paragraph scenario in which another student promised to help them by reading their term paper and giving them comprehensive feedback. After reading details about the circumstance, participants in the broken promise condition read that "your friend returned the paper . . . with comments that were actually far less extensive than promised . . . and instead included only a general uninformative comment at the top . . . " Participants in the kept promise condition instead read that "Your friend returned the paper . . . with comments that were exactly as promised . . . " Participants in the exceeded condition read that "your friend returned the paper . . . with comments that were actually far more extensive than promised . . . catching problems with style, grammar, and the general flow . . . [and] also identified two major flaws . . . and offered very helpful alternative to strengthen the paper."

Participants predicted how happy they would be, how likely they were to trust the promise maker in the future, and how interested they would be to help that individual in the future (on 1–11 scales).

Experiment 1b Procedure

We approached 60 UCSD undergraduates in various locations across campus that agreed to participate in exchange for candy. Participants read the same scenario described in Experiment 1a and then predicted the impression they would form of promise makers on 27 personality traits (on 0–10 scales). Of these, nine traits were of interest, measuring receivers' perceptions of the promise maker's selfishness (unfair, unkind, and selfish), fairness (just, fair, and equitable), and generosity (charitable, generous, and kind; see Appendix for the complete list).

Results and Discussion

Experiment 1a

We averaged participants' responses to the three measures into an overall positivity composite ($\alpha = .95$). Not surprisingly, participants evaluated breaking a promise more negatively ($M = 4.40$, standard deviation [SD] = 1.51) than keeping a promise ($M = 9.70$, $SD = 1.27$), $t(59) = 11.90$, $p < .01$, $d = 3.75$. More important, participants did not evaluate exceeding a promise ($M = 9.74$, $SD = 1.57$) more positively than keeping it, $t(59) = .08$, *ns*, $d = .03$.

Experiment 1b

We averaged participants' responses to the three selfish ($\alpha = .84$), three fair ($\alpha = .86$), and three generous ($\alpha = .88$) items into a composite for each trait type. As shown in Figure 1, participants rated promise makers as more selfish, less fair, and less generous after breaking a promise than after keeping it, ts(57) = 7.10, 4.05, and 4.58, respectively, ps < .01, ds = 5.31, 3.01, and 3.07. More important, they rated promise makers as equally (un)selfish, fair, and generous when they exceeded versus kept their promise, ts(57) = .03, .13, and 1.52, respectively, ps > .1, ds = .02, .08, and 1.10 (see Figure 1).

We believe these two experiments suggest a psychological asymmetry between breaking and exceeding a promise. Whereas breaking a promise is evaluated negatively, exceeding a promise is evaluated no more positively than merely keeping a promise. One alternative explanation, however, is that this asymmetry is an experimental artifact produced by a ceiling effect on evaluations in the kept and exceeded conditions.

We address this potential artifact in three ways here and revisit this issue later where applicable. First, we conducted additional analyses of all experiments using both linear and

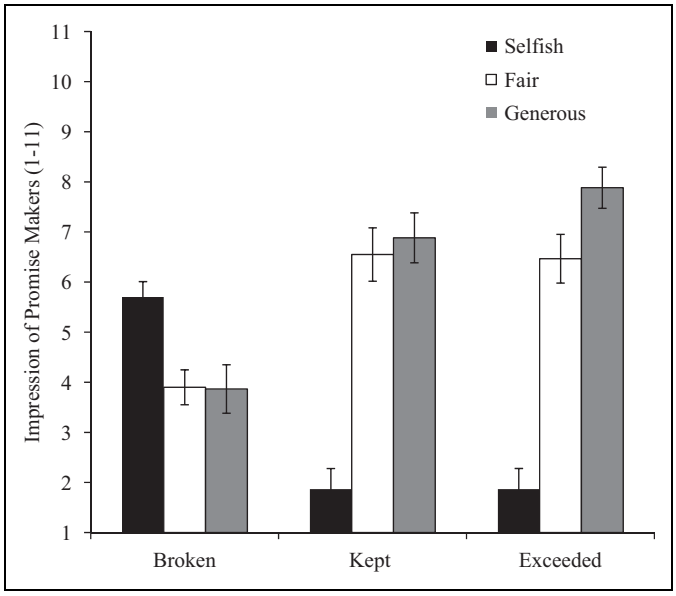

Figure 1. Promise receivers' impressions of promise makers following a hypothetical broken, kept, or exceeded promise (Experiment 1b).

Tobit regressions. Tobit regressions (or censured regressions) treat observations at the ceiling of a measure as equivalent to that observation or greater. Tobit regressions are designed specifically to address ceiling effects, testing the effect of an independent variable on a dependent variable if the measure's ceiling were removed. In all experiments (see Table 1), the linear and Tobit regressions yield nearly identical results, inconsistent with a ceiling effect artifact.

Second, not all measures yield the same percentage of observations at ceiling (see Table 1). In Experiment 1a, for instance, 30–35% of observations were at ceiling in the kept and exceeded conditions. In Experiment 1b, only 5–10% of observations were at ceiling in these conditions. However, both experiments show the same patterns. If our results were constrained by a ceiling effect, then results should vary as average responses move further from the extreme point of a measure. But they do not.

Finally, a ceiling effect suggests that evaluations are restricted in some conditions more than in others. If so, then the variance in evaluations should be smaller in the conditions restricted by a ceiling effect compared to conditions that are not restricted by a ceiling effect. However, an inspection of the *SD*s in Experiments 1a and 1b, as well as in all of the experiments that follow, shows that the variance in evaluations is not systematically smaller in the kept and exceeded conditions than in the broken conditions. This is inconsistent with a ceiling effect constraining evaluations in some conditions.

An artifact based on ceiling effects does not appear to explain the results of Experiments 1a and 1b, nor the results of the experiments that follow. Instead there appears to be an asymmetry in evaluations of broken versus exceeded promises. Breaking a promise is evaluated negatively compared to keeping a promise, but exceeding one's promise does not yield markedly more positive evaluations than merely keeping a promise.

Experiments 2a and 2b—Recalled Promises

Experiments 2a and 2b asked participants to recall promises. Recalled promises have more ecological validity than hypothetical promises, because they are memories of actual events rather than imagined scenarios. Recalled promises are also important because our memories of how others treated us may influence decision making more than (forgotten) real-time evaluations (Wirtz, Kruger, Scollon, & Diener, 2003).

Participants in Experiment 2a recalled three promises: one broken, one kept, and one exceeded. Experiment 2b provided a replication using a between-participants' manipulation of the promise outcome, and an additional measure of how much effort participants thought promise makers expended.

Method

Experiment 2a Procedure

Fifty-nine University of Chicago undergraduates recalled promises someone made to them that were broken, kept, and exceeded. For each event, participants described its details, then reported their relationship with the promise maker, then reported their happiness with the person's behavior (1 = *not at all happy*, 11 = *extremely happy*), and finally reported how difficult it was to recall the event (1 = *not at all*, 11 = *extremely*).

Experiment 2b Procedure

Forty-five University of Chicago master of business administration students completed a procedure similar to Experiment 2a, with three exceptions. First, participants recalled only one promise (kept, broken, or exceeded). Second, after writing down the details of the event, participants reported both how happy they were at the time of the event (−5 = *extremely unhappy*, 5 = *extremely happy*) and how happy they were thinking back on that event (−5 = *extremely unhappy*, 5 = *extremely happy*). Finally, participants reported how much effort they thought promise makers invested in keeping the promise (0 = *none at all*, 10 = *very much*).

Results and Discussion

Experiment 2a

Two participants did not recall broken and exceeded promises, and three additional participants did not recall an exceeded promise. Analyzing only those who recalled all three events ($N = 54$), participants reported feeling significantly less happy when recalling broken versus kept promises ($M_{\text{Broken}} = 3.04$, $SD = 1.53$; $M_{\text{Kept}} = 8.04$, $SD = 2.01$), $t(53) = 15.34$, $p < .01$, $d = 5.78$. Participants also recalled being happier, although to a lesser extent (only 13.3% of the size), when remembering exceeded ($M = 8.81$, $SD = 2.43$) versus kept promises, $t(53) < 2.29$, $p < .05$, $d = .77$. Notice that relatively few observations were at ceiling in the kept condition (9.5%, see Table 1), and that the variance in responses was directionally larger in the

Table 1. Regression Analyses and Percentage of Responses at Ceiling for Experiments 1a–4.

Experiment	Comparison	Linear Regression	Tobit Regression	% at Ceiling	
1a	Broken–Kept	−5.30 ($p < .01$)	−5.63 ($p < .01$)	Kept	31.6
	Kept–Exceed	0.04 ($p = .93$)	0.06 ($p = .54$, *ns*)	Exceeded	33.3
1b Fairness	Broken–Kept	−2.65 ($p < .01$)	−2.74 ($p < .01$)	Kept	10
	Kept–Exceed	−0.08 ($p > .1$)	−0.13 ($p > .1$)	Exceeded	5
Generosity	Broken–Kept	−3.02 ($p < .01$)	−3.11 ($p < .01$)	Kept	5
	Kept–Exceed	1.00 ($p > .1$)	1.13 ($p > .1$)	Exceed	15
2a	Broken–Kept	−5.00 ($p < .01$)	−5.34 ($p < .01$)	Kept	9.3
	Kept–Exceed	0.78 ($p = .05$)	1.19 ($p < .05$)	Exceeded	31.5
2b	Broken–Kept	−6.27 ($p < .01$)	−6.73 ($p < .01$)	Kept	33.3
	Kept–Exceed	−.40 ($p = .45$, *ns*)	−.49 ($p = .75$, *ns*)	Exceeded	27.7
3	Broken–Kept	−1.54 ($p < .01$)	−2.60 ($p < .01$)	Kept	28
	Kept–Exceed	.10 ($p = .88$, *ns*)	.37 ($p = .63$, *ns*)	Exceeded	29.33
4 Expectations	Broken–Kept	−4.08 ($p < .01$)	−4.47 ($p < .01$)	Kept	23.5
	Kept–Exceed	1.92 ($p < .01$)	2.76 ($p < .01$)	Exceeded	70
Promises	Broken–Kept	−6.87 ($p < .01$)	−8.89 ($p < .01$)	Kept	35.7
	Kept–Exceed	−0.26 ($p > .1$)	0.23 ($p > .1$)	Exceeded	64.3

exceeded promises condition, again suggesting that a ceiling effect is not artifactually suppressing evaluations in the kept and exceeded promises conditions. Including all participants in the analyses does not alter these results meaningfully.

Experiment 2b

Participants' recalled happiness at the time of the event and thinking back on the event were highly correlated ($r = .79$), and so we averaged them into a composite for the following analyses. Participants were less happy recalling broken ($M = -2.62$, $SD = 1.35$) versus kept promises ($M = 3.65$, $SD = 1.74$), $t(42) = 10.21$, $p < .01$, $d = 3.15$. More important, participants were no happier recalling exceeded ($M = 3.25$, $SD = 1.70$) versus kept promises, $t(42) = .70$, *ns*, $d = .23$. Figure 2 shows this same pattern for both reported happiness measures.

Participants' estimates of promise makers' effort followed the same pattern: participants believed promise makers expended less effort in broken versus kept promises (Ms = 1.67 vs. 6.82, respectively), $t(42) = 5.64$, $p < .01$, $d = 2.25$, but did not expend significantly more effort in exceeded ($M = 7.69$) versus kept promises, $t(42) = 1.02$, *ns* $d = .32$. If exceeding a promise required more effort than keeping it, then promise receivers failed to recognize it.

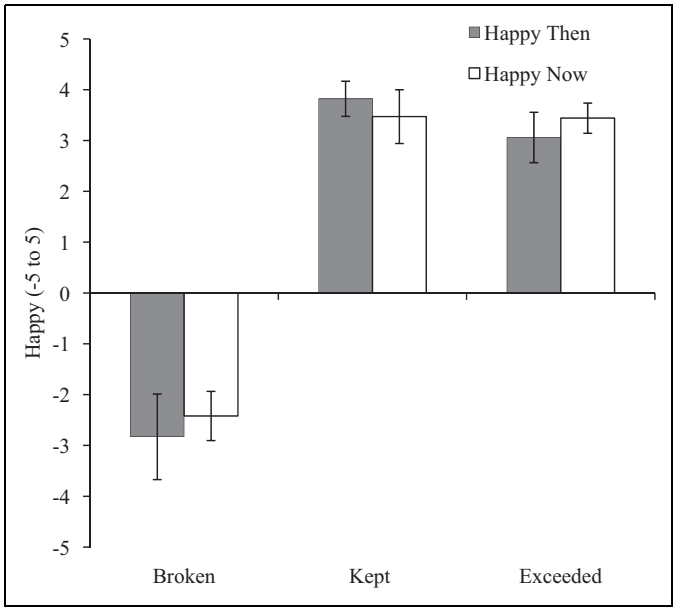

Figure 2. Promise receivers' recalled happiness at the time of, and now thinking back on, a broken, kept, and exceeded promise (Experiment 2a).

Experiment 3—Actual Promises

Imagination and memory allowed us in Experiments 1a–2b to measure evaluations of events that we could not replicate in an actual experiment, but imagination is not reality and memory is prone to distortions, calling into question the generalizability of these experiments to actual promises that might arise in social interactions. Although all existing studies of promise-keeping that we are aware of rely on either imagined or recalled promises (Conway & Briner, 2002), we sought in Experiment 3 to create a real promise that is broken, kept, or exceeded under experimental conditions.

In Experiment 3, promise makers were instructed to make a promise to another participant that they were subsequently instructed to break, keep, or exceed in a clearly specified way. This design allowed for experimental control over the promises' outcomes. In addition, this experimental method allowed us to control the actual effort required under each promise outcome and to ensure that the objective cost associated with breaking a promise equaled the objective gain from exceeding it. Finally, this method allows clear causal inferences about the effects of a promise's outcome on evaluations.

Method

Procedure

One hundred and twenty University of Chicago undergraduates ($N = 60$) participated in dyads. One participant in each dyad was randomly assigned to be a promise receiver and the other a promise maker. Promise receivers learned that they would solve 40 puzzles (counting zeros in a table of numbers) and be paid per puzzle solved. They learned that they would not have enough time to solve all of the puzzles, but that they would be paired up with another participant (the promise maker) who would decide whether or not to help them complete more puzzles.

Promise makers learned that they could choose to help the other participant solve tables. If willing to help, they were told they would promise the other participant to solve 10 puzzles. We informed all participants that the promise makers' help would only benefit the promise receivers. All promise makers agreed to help, walked to the other participant's room, and verbally promised to solve 10 tables following a script provided by the experimenter.

To manipulate promise keeping, the experimenter then instructed promise makers to solve 10 tables (as promised), 5 tables, or 15 tables. Promise makers worked for 7 min, during which promise receivers completed a personality inventory (the Five Factor Model; John, Naumann, & Soto, 2008). Next, the experimenter delivered 5, 10, or 15 (pre)-solved tables to promise receivers (with the remaining unsolved), ostensibly completed by the promise maker.The experimenter said, ''while you were filling out the personality test, your partner solved 5 tables, instead of the 10 tables he or she promised to solve (or 10 tables, exactly as he or she promised; or 15 tables, instead of the 10 tables he or she promised to solve).''

Promise receivers then completed a survey measuring the effort they thought promise makers expended in keeping the promise (1 = *very little effort*, 11 = *extreme effort*), to what extent they thought promise makers intended to keep their promise (1 = *not at all*, 11 = *very much*), how happy they were (−5 = *extremely unhappy*, 5 = *extremely happy*), and how grateful they were for their help (1 = *not at all grateful*, 11 = *very much grateful*). Promise receivers were then allowed to solve as many tables as they wanted. Just before being paid, promise receivers completed the same survey measures again to test whether their final evaluations depended on their earnings.

Results and Discussion

We reverse scored and transformed happiness ratings to a 1–11 scale. Reported happiness and gratitude were highly correlated both before and after being paid ($r = .79$ and .80, respectively). We analyze before and after ratings separately below to test for possible main effects or interactions.

A 2(Evaluation: before vs. after payment) × 3(Promise: broken, kept, exceeded) analysis of variance (ANOVA) on overall positivity yielded only a significant main effect for promise, $F(2, 114) = 6.37$, $p < .01$, $\eta^2 = .10$, with no main effect of evaluation or interaction (*p*s > .1). Again, promise receivers evaluated a broken promise less positively ($M = 8.32$, $SD = 2.58$) than a kept promise ($M = 9.88$, $SD = 1.80$), $t(57) = 2.34$, $p < .05$, $d = .74$, but did not evaluate exceeded promises more positively ($M = 9.77$, $SD = 2.05$) than kept promises, *ns*, $d = .06$.

The same ANOVA on perceived effort yielded only a significant main effect for promise, $F(2, 114) = 20.06$, $p < .01$, $\eta^2 = .26$, with no main effect of evaluation or interaction (*p*s > .1). Participants believed promise makers expended less effort when the promise was broken ($M = 6.76$, $SD = 2.13$) versus kept ($M = 9.55$, $SD = 1.68$), $t(57) = 4.99$, $p < .01$, $d = 1.45$, but did not believe promise makers expended more effort when the promise was exceeded ($M = 9.50$, $SD = 1.37$) versus kept, $t(57) = .09$, *ns*, $d = .03$.

Finally, the same ANOVA on intent to keep the promise yielded only a significant main effect for Promise, $F(2, 114) = 7.57$, $p < .01$, $\eta^2 = .11$, with no main effect of evaluation or interaction (*p*s > .1). Participants assumed a stronger intention to keep the promise when it was kept ($M = 10.18$, $SD = 1.23$) versus broken ($M = 8.31$, $SD = 2.09$), $t(57) = 2.96$, $p < .01$, $d = 1.12$, but did not attribute stronger intentions when a promise was exceeded ($M = 9.45$, $SD = 2.45$) versus kept, $t(57) = 1.14$, *ns*, $d = .40$.

This experiment replicated an asymmetry in evaluations of broken versus exceeded promises using actual promises. Unlike the preceding experiments, we observed this asymmetry despite the obvious (and symmetrical) increase in effort required of promise makers to exceed a promise than to keep it. Exceeding a promise clearly required more effort and produced more benefit than merely keeping a promise, and yet promise receivers evaluated them equally.

Experiment 4—Promises Versus Reference Points

The asymmetry observed in our experiments could emerge through at least two mechanisms. First, *kept* promises could be evaluated more positively than a perfectly linear relationship between outcomes and evaluations would produce. Promises on this account serve as social contracts, in which the value of *keeping* a promise goes beyond the objective benefit provided, thereby increasing its perceived value beyond its objective benefit. Second, *exceeding* a promise could be evaluated less positively than a linear relationship would produce. Promises could serve as reference points such that gains from a reference point have less impact on evaluations than losses from a reference point (Kahneman & Tversky, 1979). On this account, keeping a promise is evaluated in line with its objective outcome but exceeding a promise is undervalued compared to objective benefit.

We tested these mechanisms by comparing promises with expectations. Although both promises and expectations produce a reference point in judgment, they differ because a promise is interpersonal—a commitment made from one

person to the other—whereas an expectation is intrapersonal—a belief held by only one person (i.e., the "receiver"). Promises are therefore social contracts whereas expectations are merely reference points of comparison for evaluating outcomes. We therefore predicted that expectations would produce a relatively linear relationship between outcomes and evaluations, providing a benchmark of comparison for the asymmetry observed between promised outcomes and evaluations. If keeping one's promise is overvalued, then a kept promise should be evaluated more favorably than a met expectation. If, however, exceeding one's promise is undervalued, then an exceeded promise should be evaluated less positively than an exceeded expectation.

Method

Procedure

UCSD undergraduates ($N = 143$) were randomly assigned to recall a time when someone broke, kept, or exceeded a promise made to them, *or* a time they expected someone to do something that the other person fell short of, met, or exceeded. Participants assigned to the promises conditions received instructions similar to those in Experiment 2a and 2b—to recall a promise someone made to them that was broken (someone did less than promised), kept (someone did exactly as promised), or exceeded (someone did more than promised). Participants in the expectations condition, in contrast, were asked to recall an instance when they expected someone to do something and the person did less than expected, did exactly what was expected, or did more than expected. After describing the event in writing, participants reported how happy (-5 = *extremely unhappy*, 5 = *extremely happy*) and how disappointed (1 = *not at all disappointed*, 11 = *extremely disappointed*) they were with the person's behavior.

Results and Discussion

We reverse scored and transformed disappointment ratings to a -5 to 5 scale, and then averaged them with happiness ratings to create a positivity composite ($r = .84$). A 2 (Context: expectation vs. promise) $\times$ 3 (Outcome: broken, kept/met, exceeded) ANOVA on this composite revealed a significant effect for outcome, $F(2, 137) = 117.93$, $p < .01$, $\eta^2 = .63$, qualified by the predicted interaction, $F(2, 137) = 5.20$, $p < .01$, $\eta^2 = .07$.

Participants' responses to expectations were linearly related to the outcome but responses to promises were not (Figure 3). When recalling expectations, participants were more positive when another person met versus fell short of their expectation, $t(69) = 6.24$, $p < .01$, $d = 1.53$, and were also more positive when expectations were exceeded versus met, $t(69) = 2.99$, $p < .01$, $d = 1.04$. When considering promises, however, an asymmetry emerged: participants were less positive when a promise was broken than when kept, $t(68) = 10.77$, $p < .01$, $d = 4.68$, but were not more positive when a promise was exceeded versus kept, $t(68) = .45$, *ns*, $d = .14$. More important, participants were more positive when a promise was kept than when an expectation was met, $t(42) = 3.19$, $p < .01$, $d = 1.10$, the only instance where evaluations of promises and expectations diverged.

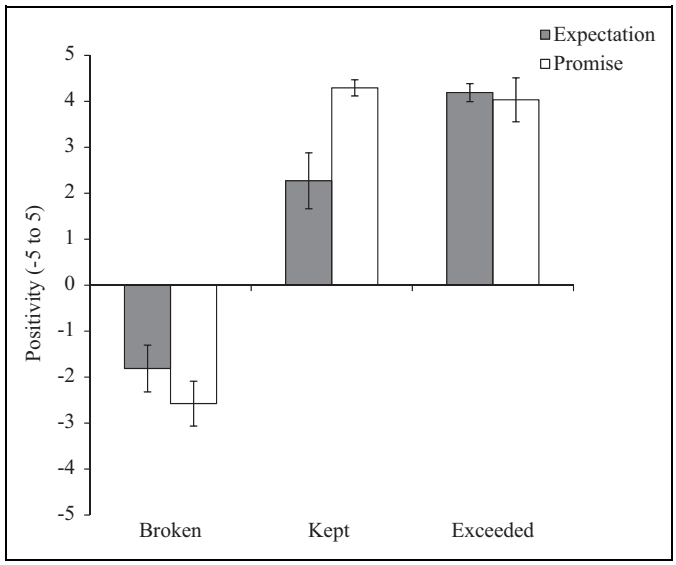

Figure 3. Promise receivers' evaluations of broken, kept, or exceed promises versus expectations (Experiment 4).

These results are inconsistent with the reference point mechanism derived from Prospect Theory, and instead support the social contract mechanism in which keeping promises is relatively overvalued. Promise receivers appear to place a premium on keeping one's promise, a premium that comes from fulfilling a social contract, with no added value from exceeding one's promise.

General Discussion

Four sets of experiments demonstrate an asymmetric evaluation of promises. Although broken promises are evaluated negatively, exceeded promises are not evaluated more positively than merely keeping one's promise. Promise receivers consistently failed to recognize the additional effort required to exceed a promise but recognized a lack of effort when breaking a promise, even when effort and benefits associated with exceeding versus breaking a promise were clearly symmetrical.

This pattern does not seem to reflect a simple ceiling effect on evaluations in the kept and exceeded conditions. Tobit regressions yield the same results as linear regressions, the percentage of evaluations at ceiling varies across our experiments but the same pattern emerges consistently, and the variance in evaluations is not systematically smaller in the exceeded promises condition than in the broken promises condition.

Instead, Experiment 4 suggests the asymmetry in evaluations is caused by overvaluing kept promises, suggesting the contractual nature of a promise provides value above and beyond a promise's tangible outcome. This premium on keeping one's promise may reflect a broader social system meant to sustain cooperation for everyone's benefit, one highly sensitive to signals of trustworthiness (Almenberg, Dreber, Apicella, & Rand, 2011). Indeed, existing research suggests that social systems tend to be hypersensitive to signals of dishonesty and unfairness but do not necessarily reward generosity (Alexander, 1987; Keysar, Converse, Wang, & Epley, 2008; Klein & Epley, 2014; Tooby & Cosmides, 1996).

We believe this observed "fairness premium" may reflect a more general pattern in social judgment. Consider a follow-up experiment testing this possibility. We asked 50 MTurk participants to rate the desirability of each trait used in Experiment 1b on scales ranging from 1 (*extremely undesirable*) to 10 (*extremely desirable*). We used the same nine traits to create measures of selfishness (unfair, unkind, and selfish; $\alpha = .91$), fairness (just, fair, and equitable; $\alpha = .71$), and generosity (charitable, generous, kind; $\alpha = .81$). Consistent with a fairness premium, participants evaluated selfish traits ($M = 3.63$, $SD = 2.37$) as significantly less desirable than fairness traits ($M = 7.55$, $SD = 1.59$), $t(49) = 8.83$, $p < .01$, $d = 4.43$, but did not rate generosity traits as more desirable ($M = 7.62$, $SD = 1.57$) than fairness traits, $t(49) = .35$, *ns*, $d = .09$. Generous traits were not evaluated more positively than fair traits, but selfish traits were evaluated more negatively than fair traits. We know of one other result consistent with these findings. In an experiment involving a *dictator game* (Almenberg et al., 2011), participants could punish or reward another person for being selfish, fair, or generous to another participant. Results showed that participants punished others for being selfish, but rewarded them equally for being fair or generous. Placing a premium on fairness, but no additional benefit for generosity, may represent a more general pattern of behavior in a social system designed to enable cooperation among unrelated individuals.

Although our experiments assessed promises between individuals, we expect this asymmetry extends to promises involving more abstract relationships, such as between customers and companies or employees and employers (Conway & Briner, 2002). Businesses may work hard to exceed their promises to customers or employees (Lester et al., 2002), but our research suggests that this hard work may not produce the desired consequences beyond those obtained by simply keeping promises. As an initial test of this possibility, we asked [university] undergraduates to imagine that they had bought tickets to a concert from an online company. They imagined purchasing tickets for Row 10, but then either received worse tickets than promised (Row 11, 13, or 15), better tickets than promised (Row 9, 7, or 5), or exactly what was promised (Row 10). Participants then indicated how satisfied they would be, how likely they were to recommend the company to a friend, and how likely they were to use the same seller in the future ($\alpha = .97$). Once again, participants were more negative when they received worse tickets than promised (regardless of how much worse; $M = 2.05$, $SD = .74$) than when they got exactly what was promised ($M = 7.97$, $SD = 1.34$), $t(86) = 16.32$, $p < .01$. Participants were no more positive, however, when they received better tickets than promised (regardless of how much better) compared to exactly what was promised. In fact, they were somewhat more negative ($M = 7.20$, $SD = 2.02$), $t(86) = 2.03$, $p < .05$.

Promises can be hard to keep, and promise makers should spend their effort keeping them wisely. The results of our experiments suggest that it is wise to invest effort in keeping a promise because breaking it can be costly, but it may be unwise to invest additional effort to exceed one's promises. When companies, friends, or coworkers put forth the effort to keep a promise, their effort is likely to be rewarded. But when they expend extra effort in order to exceed those promises, their effort appears likely to be overlooked.

Appendix

Table A1. Full List of Traits, Experiment 1b.

Alert	Capable	**Fair**	Efficient	Friendly	Brilliant
Kind	**Selfish**	Practical	**Generous**	Imaginative	Skillful
Unkind	Progressive	Artistic	**Charitable**	Lucky	Energetic
Open-minded	Humble	**Equitable**	Sophisticated	**Unfair**	Wise
Sympathetic	**Just**	Talented			

Acknowledgments

We thank Jasmine Kwong, Entzu Lin, Chau Ngo, Silvia Saccardo, and Beverly Tam for assistance conducting this research; Katherine Burson and Karsten Hansen for data analysis assistance; Tal Eyal, Tom Gilovich, Elizabeth Keenan, Alex Imas, and John Skowronski for helpful comments. Experiments 2a and 2b were included in Ayelet Gneezy's PhD dissertation.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Note

1. In all experiments, we collected evaluations from the perspective of both promise receivers and promise makers. The promise maker's perspective tested a distinct hypothesis about whether promise makers would underestimate the negative consequences of breaking a promise. All but one experiment supports this hypothesis significantly. Although interesting, the promise makers' evaluations are not central to our hypotheses about an asymmetry in evaluations for promise receivers. Because they neither qualify nor add to the results from promise receivers, we present the promise makers' procedures and results in the Online Supplemental Material.

Supplemental Material

The online data supplements are available at http://spp.sagepub.com/supplemental

References

Alexander, R. D. (1987). *The biology of moral systems*. London, England: Aldine.

Almenberg, J., Dreber, A., Apicella, C. L., & Rand, D. G. (2011). Third party reward and punishment: group size, efficiency and public goods. *Psychology and punishment* (pp. 73–92). Hauppauge, NY: Nova Publishers.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*, 323–370.

Bowles, S., & Gintis, H. (2003). Origins of human cooperation. In P. Hammerstein (Ed.), *Genetic and cultural evolution of cooperation* (pp. 429–443). Cambridge, MA: MIT Press.

Buehler, R., Griffin, D., & Ross, M. (1994). Exploring the ''planning fallacy'': Why people underestimate their task completion times. *Journal of Personality and Social Psychology*, *67*, 366–381.

Conway, N., & Briner, R. B. (2002). A daily diary study of affective responses to psychological contract breach and exceeded promises. *Journal of Organizational Behavior*, *23*, 287–302.

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition*, *31*, 187–276.

Gneezy, A., & Fessler, D. T. (2012). Conflict, sticks and carrots: War increases prosocial punishments and rewards. *Proceedings of the Royal Society B*, *279*, 219–223.

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big-Five trait taxonomy: History, measurement, and conceptual issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (pp. 114–158). New York, NY: Guilford Press.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decisions under risk. *Econometrica*, *47*, 313–327.

Keysar, B., Converse, B. A., Wang, J., & Epley, N. (2008). Reciprocity is not give and take: Asymmetric reciprocity to positive and negative acts. *Psychological Science*, *19*, 1280–1286.

Klein, N., & Epley, N. (2014). *The topography of generosity: Asymmetric evaluations of prosocial actions*. Manuscript submitted for publication. University of Chicago, Chicago, IL.

Lester, S. W., Turnley, W. H., Bloodgood, J. M., & Bolino, M. C. (2002). Not seeing eye to eye: Differences in supervisor and subordinate perceptions of and attributions for psychological contract breach. *Journal of Organizational Behavior*, *23*, 39–56.

Martijn, C., Spears, R., Van der Pligt, J., & Jakobs, E. (1992). Negativity and positivity effects in person perception and inference: Ability versus morality. *European Journal of Social Psychology*, *22*, 453–463.

Robinson, S. L., & Rousseau, D. M. (1994). Violating the psychological contract: Not the exception but the norm. *Journal of Organizational Behavior*, *21*, 525–546.

Rousseau, D. M. (1995). *Psychological contracts in organizations: Understanding written and unwritten agreements*. Thousand Oaks, CA: Sage.

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, *5*, 296–320.

Skowronski, J. J., & Carlston, D. E. (1987). Social judgment and social memory: The role of cue diagnosticity in negativity, positivity, and extremity biases. *Journal of Personality and Social Psychology*, *52*, 689–699.

Thaler, R. H. (1985). Mental accounting and consumer choice. *Marketing Science*, *4*, 199–214.

Tooby, J., & Cosmides, L. (1996). Friendship and the banker's paradox: Other pathways to the evolution of adaptations for altruism. In W. Runciman, J. Maynard Smith, & R. Dunbar (Eds.), *Evolution of social behavior patterns in primates and man. Proceedings of the British Academy* (Vol. 88, pp. 119–143).

Wirtz, D., Kruger, J., Scollon, C. N., & Diener, E. (2003). What to do on spring break? The role of predicted, online, and remembered experience on future choice. *Psychological Science*, *14*, 520–524.

Author Biographies

Ayelet Gneezy is an Associate professor of behavioral sciences and marketing at UCSD. Her research addresses a wide variety of questions pertaining to consumer behavior such as behavioral pricing and social preferences. In her research, professor Gneezy often collaborates with firms and integrates field experiments to test her predictions.

Nicholas Epley is the John T. Keller Professor of Behavioral Science at the University of Chicago Booth School of Business. His research is focused on the experimental study of social cognition, perspective taking, and intuitive human judgment.