

Cueing implicit commitment

Francesca Bonalumi¹, Margherita Isella², John Michael^{1,3}

¹Department of Cognitive Science, Central European University, Budapest, Hungary

²Department of Philosophy, Vita-Salute San Raffaele, Milan, Italy

³Department of Philosophy, University of Warwick, UK

Forthcoming in *Review of Philosophy and Psychology*

Abstract: Despite the importance of commitment for distinctively human forms of sociality, it remains unclear how people prioritize and evaluate their own and others' commitments - especially implicit commitments. Across two sets of online studies, we found evidence in support of the hypothesis that people's judgments and attitudes about implicit commitments are governed by an implicit sense of commitment, which is modulated by cues to others' expectations, and by cues to the costs others have invested on the basis of those expectations.

1. Introduction

The phenomenon of commitment is a cornerstone of human social life. Commitments make individuals' behavior predictable in the face of fluctuations in their desires and interests, thereby facilitating the planning and coordination of joint actions involving multiple agents (Clark, 2006; Michael & Pacherie, 2015). Moreover, by stabilizing expectations about individuals' future behavior, commitments can also help to support cooperation. As such, the origin and stability of everyday social exchanges and institutions such as marriage, scientific collaboration, and employment depend upon the credibility of commitments. Speech acts such as promises and vows, as well as complex social institutions such as contracts, allow the creation of explicit commitments – i.e., commitments whose terms are clearly understood and accepted by all parties. But even

when commitments are not made explicit, they can nevertheless support the same important social functions. Indeed, philosophers such as Margaret Gilbert and Michael Bratman have recently emphasized the role of implicit commitments in joint actions, based on the idea that joint actions are characterized by the existence of a shared goal – the achievement of which is what all parties implicitly commit to¹ (Bratman, 1993; Gilbert, 1990). Despite the importance of implicit commitment for distinctively human forms of sociality, it remains unclear how people identify, prioritize and assess their own and others' commitments.

Imagine, for example, that two colleagues, Polly and Pam, are in the habit of meeting and chatting together on the balcony of their office building every afternoon during the coffee break (adapted from Gilbert 2006). Even if they have never agreed explicitly to engage in this routine, they may over time come to feel much the same as they would if an explicit commitment were in place. As a result, if Pam finds herself confronted with some other important obligation or enticing alternative on one occasion, she may hesitate before breaking the routine she shares with Polly. What factors will influence her judgment as to whether it is acceptable to break with the routine? And what factors will shape Pam's response if Polly does fail to show up? Following Michael, Sebanz & Knoblich (2016), we hypothesize that people's judgments and attitudes about such situations are governed by a *sense of commitment*, which is modulated by various cues that another agent expects one to perform a particular action, such as the history of repeated interaction, and cues that another agent may have invested effort or other costs on the basis of that expectation.

This hypothesis builds upon prior research on the role of expectations in demanding and motivating prosocial behavior such as maintaining promises or abiding by tacit rules. MacCormick and Raz (1978) and Scanlon (1998) hold that promises have normative force in situations when the promiser leads the promisee to form certain expectations about her (the promiser's) future behavior. In another highly influential contribution made in the

¹ With substantial differences: while according to Bratman commitment is not a necessary aspect of shared intentionality, but a characteristic consequence of it, Gilbert holds commitment to be a core aspect of shared intentionality: by sharing a goal, subjects are implicitly agreeing to be part of a plural subject of the shared goal.

context of an analysis of how social practices are established and become self-reinforcing, Lewis introduces the idea of a 'presumptive reason', according to which one ought to fulfill others' preferences when it is the case that one is reasonably expected to do so (1969, pp. 97-98; cf. Bicchieri, 2005). Building upon this idea, Sugden (2000) claims that one is normatively expected to perform a certain course of action X when such a presumptive reason is present, and that one is typically motivated to perform X by means of an aversion to frustrating others' reasonable expectations. Sugden also suggests that this aversion mirrors the emergence of a feeling of resentment towards those who have frustrated one's own expectations.

More recently, some empirical research has begun to test these ideas, and in particular to probe the cognitive and motivational mechanisms leading people to feel committed and to act accordingly, and to expect the same of others as well. For example, studies using game-theoretic paradigms have shown that people's expectations have a positive impact on the behavior of their partners. For instance, Heintz and colleagues (2015) found that participants playing the role of dictator in a dictator game made more prosocial choices when they *explicitly* received information about the recipients' expectations -- provided the expectations were reasonable (Cf. also Dana et al. 2006; Ockenfels & Werner, 2012).

However, when there is no explicit information about others' expectations, how can people become aware of them? Addressing this question, Michael, Sebanz & Knoblich (2016) argue that a partner's investment of effort or other costs in a joint activity may provide an *implicit* cue to that partner's expectations about one's contribution to the joint activity -- i.e. if the partner were not expecting one to remain committed and to do one's part, then she would be unlikely to invest effort or other costs. Moreover, a partner's investment of effort also provides a cue that the joint activity is of value to her, implying that she may be particularly disappointed or annoyed if one did not remain committed and do one's part. This line of reasoning is also motivated by previous findings suggesting that the cost invested by one agent in order to allow a partner to obtain rewards has an influence on the choices made by the partner (Charness & Rabin 2010). More recently, Székely & Michael (2018) also found that the perception of a partner's investment of effort

in a joint activity led participants to remain engaged longer despite increasing boredom². In a 2-player version of the classic snake game which became increasingly boring over the course of each round, participants persisted longer when they were given cues of their partner's highly effortful contribution to the game compared to when they were given cues of a partner's low investment of effort.

While Székely & Michael's (2018) finding is consistent with the hypothesis that the perception of a partner's investment of effort led participants to persist longer out of a sense of commitment, alternative explanations are also possible. For example, the perception of a partner's effort might have led participants to infer that the task was particularly worthwhile. Alternatively, the perception of another agent investing effort may have primed them to exert effort as well, irrespective of any sense of commitment to another agent. To address these open questions, we designed a pair of experiments (Study 1a and Study 1b) to probe participants' normative judgments and affective responses to a scenario in which (as in Székely & Michael's 2018 study) one agent is relying on a second agent who is presented with a temptation to disengage. However, whereas Székely & Michael focused on the agent who was presented with the temptation (i.e. they were investigating the effect of a sense of commitment upon this agent's motivation), we opted to focus on the other side of the relation. In fact, the effect of a sense of commitment implies that while one agent would feel motivated to do what she committed to doing, the partner will feel more entitled to expect it to happen, and to blame more the first agent if she fails to do it. Thus, we presented participants with vignettes describing a scenario in which one agent had a high level of expectation (generated by investing a higher degree of effort into a joint activity, i.e. the High Cost Condition) or a low level of expectation (generated by investing a lower degree of effort, i.e. the Low Cost Condition), and a second agent failed to remain committed, and operationalized the sense of commitment in terms of the degree to which participants made negative normative and non-normative judgments about the second agent's violation.

² Indeed, if it is the case that such cues typically track others' expectations, then people may respond to them by increasing their commitment to joint activities even in cases in which they do not in fact reflect a partner's expectations.

We reasoned that if participants made more negative normative judgments and reported more negative emotional attitudes in response to the High Cost condition, this would be difficult to account for in terms of the aforementioned alternative explanations of Székely and Michael's (2018) finding. Indeed, the priming of the partner's effort and the value of an action to an agent can imply an emotional reaction but does not in itself imply any obligation that she has to any other agent to perform the action. This additional normative measure we added would therefore provide further support for the hypothesis that a partner's investment of effort in a joint activity triggers a sense of commitment to that joint activity. We opted for operationalizing commitment using a 6-point Likert scale for the following reason: if the sense of commitment is modulated by cues of another agent's expectations, rather than by a norm-violation per se, we should expect that participants' judgments would vary between conditions in a graded manner rather than in a binary manner.

As a further test of the hypothesis that the sense of commitment is modulated by various cues that another agent expects one to perform a particular action, such as the history of repeated interaction, we also carried out a second pair of studies (2a and 2b). Studies 2a and 2b were designed to probe participants' normative evaluations and affective attitudes in response to scenarios in which one agent failed to remain engaged to a joint activity toward which her partner had either a high level of expectation (due to having shared a long history of repeated interaction; High Repetition Condition) or a low level of expectation (due to having shared only a brief history of repeated interaction; Low Repetition Condition). We reasoned that a long history of repeated interaction is likely to establish a high degree of expectation of continued interaction, and that the scenario described in the High Repetition condition would be likely to elicit more negative normative judgments and emotional responses than the scenario described in the Low Repetition condition. This line of reasoning is motivated by previous research showing that cooperation in social dilemmas such as the prisoners' dilemma can be boosted if participants experience a history of successful coordination -- i.e. in the context of behavioral economics paradigms such as the stag hunt (Rusch & Luetge, 2016) or a pure coordination game (Guala & Mittone, 2010). Unlike these previous studies, however, the current study focused on the perspective of the agent whose expectation was

disappointed. Moreover, our paradigm enabled us to investigate people's attitudes and judgments about everyday scenarios with a high degree of ecological validity.

2. Cost and Commitment

The first pair of studies we conducted was designed to test the hypothesis that an agent's sense of commitment to an interaction is enhanced by her or his partner's investment in the interaction. To this end, we presented participants with vignettes describing everyday situations in which an implicit commitment between two agents was violated. We operationalized the sense of commitment with a normative measure (i.e. a normative question prompting a moral judgment about whether an apology was appropriate), with a non-normative, intuitive measure (i.e. an affective question asking whether the situation triggered a feeling of annoyance), and with an indirect question about how much time the participant herself would be willing to invest to honor the implicit commitment in the scenario described in the vignette.

a. Study 1a

Methods

We used Amazon M-Turk to implement a web-based paradigm with a between-subjects design. Since each participant gave only one judgment for each test question, and since online experiments produce greater variability than lab-based experiments, we expected a high variability in our dependent variables. We therefore opted for a large sample size: 200 participants (2 conditions, 100 per group). We included data from those participants who had already begun the experiment when M-Turk registered that this number had been reached. Our data set therefore comprised 260 adults (124 in *High Cost Condition* and 136 in *Low Cost Condition*) using Amazon M-Turk. (110 female; $M_{age} = 33.62$ years, $SD = 10.53$). No participant was discarded, since none failed the comprehension question. The methods used were in accordance with the international ethical requirements of psychological research and approved by the EPKEB in Hungary. All participants gave their informed consent by ticking a box prior to the experiment.

Participants were asked to read a vignette describing a hypothetical situation involving a repeated joint activity that gets interrupted. Subjects were randomly assigned to one of two between-subjects conditions (*High Cost, Low Cost*). We manipulated the magnitude of costs that an agent invested in order to be able to maintain the joint activity with the other agent. In the High Cost condition, the scenario reads as follows:

High Cost: You and Pam used to work in the same office on the 5th floor, until you were moved to a 1st floor office one year ago. Every day for the past three years, you and Pam have spent your afternoon coffee break sitting out on the 5th floor balcony and chatting, though you never agreed to start doing this. After you moved to the new office down on the 1st floor, you nevertheless continued to walk up to the same balcony on the 5th floor to spend the coffee break with Pam, even though the balcony is five flights of stairs up from your new office. The sequence is broken when one day you walk all the way up the five flights of stairs and wait for Pam during the coffee break, but she doesn't turn up.

In the Low Cost condition, the vignette differs insofar as the new office is around the corner rather than down on the first floor (See Appendix for the full vignette). After reading the vignette, participants were asked to respond to the following questions, which were presented in this order:

Normative Question: *"On a scale from 0 to 5, to what extent would you agree that Pam owes you an apology?" (0= Disagree strongly; 5= Agree strongly).*

Affective Question: *"If Pam did not apologize or offer any explanation, how annoyed would you be on a scale from 0 to 5?" (0 = not at all annoyed; 5 = highly annoyed)*

Control Question: *"In the scenario described above, where is it that you and Pam spend the coffee break?" (on the balcony, in the lounge, in the cafeteria)*

Indirect Question: *"Now imagine that you're Pam. The reason why you cannot make it is that, while running an errand in town, you learn that your favorite spa is offering free admission until 4 pm. It is currently 2:30 pm. You would like to write a text message to your colleague back at the office to let her know that you won't be coming today, but you notice that your phone is out of batteries. You plug it in to charge in the car. How long*

would you be willing to wait in the parking lot for the phone to charge before going in to the spa, in order to be able to send a text message to your colleague?” (not at all, 1 minute, 5 minutes, 10 minutes, 15 minutes, 20 minutes, 25 minutes, 30 minutes)

The normative question was designed to tap participants' explicit moral evaluations of the scenario. We predicted that they would more strongly agree that an apology was in order in the High Cost condition than in the Low Cost condition. The affective question was designed to tap participants' more intuitive, emotional reactions to the commitment violation described in the scenario. We predicted that participants would indicate a higher level of annoyance if no apology or explanation were forthcoming in the High Cost condition. The control question was designed to filter out participants who had not read the vignette with sufficient care to retain the critical information presented therein. The indirect question was intended to tap participants' appraisal of the commitment indirectly, namely by measuring the opportunity cost they themselves would be willing to pay in order to uphold the commitment. We predicted that participants would indicate a willingness to wait longer in the High Cost condition than in the Low Cost condition.

Results and Discussion

For the normative question, participants gave higher estimates in the High Cost condition ($M = 2.38$, $SD = 1.32$) than in the Low Cost condition ($M = 1.87$, $SD = 1.38$), $t(258) = 3.007$, $p = .003$, Cohen's $d = 0.37$ (small effect size). These results were confirmed by additional nonparametric tests, Mann-Whitney $U = 6728.500$, $p = .003$ (**See Figure 1a**).

Similarly, for the affective question, participants gave higher estimates in the High Cost condition ($M = 2.28$, $SD = 1.20$) than in the Low Cost condition ($M = 1.94$, $SD = 1.37$), $t(258) = 2.121$, $p = .035$, Cohen's $d = 0.26$ (small effect size). These results were confirmed by additional nonparametric tests, Mann-Whitney $U = 7169.000$, $p = .032$ (**See Figure 1b**).

Responses to the indirect question revealed a numerical difference in the same direction, with participants giving higher estimates in the High Cost condition ($M = 4.01$, $SD = 4.15$) than in the Low Cost condition ($M = 3.14$, $SD = 3.47$), but this difference did not reach statistical significance, $t(241) = 1.820$, $p = .070$. Levene's test revealed a violation of the equality of variance assumption, $p = .007$.

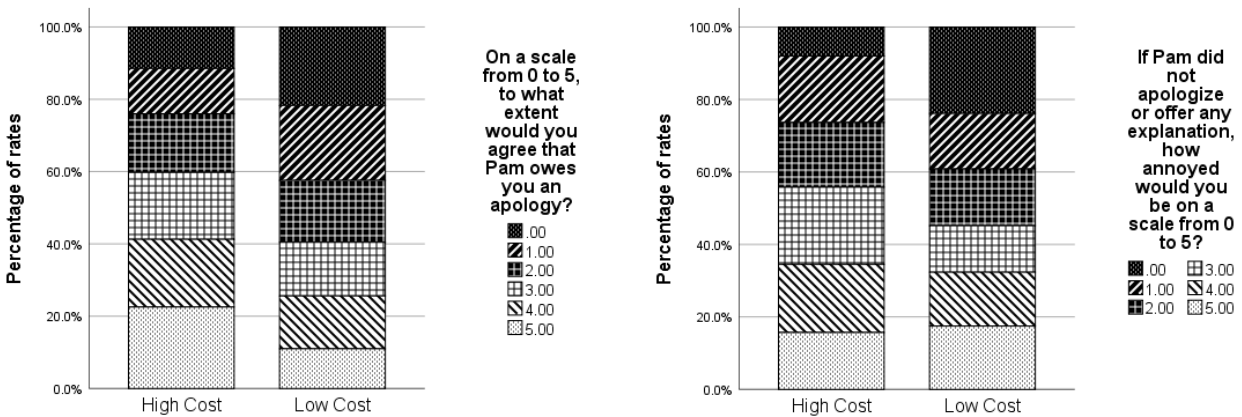


Fig. 1a and 1b. Percentage of responses to the normative (a) and the non-normative question (b). White background bars indicate a mild-to-strong agreement, whereas black-background bars indicate a mild-to-strong disagreement with the statement. In other words, the stronger the agreement, the higher the perception of the commitment being violated.

It is worth noting that responses to both the normative and the affective questions tend to cluster around the middle of the scale rather than towards the two extremes. For the normative question, responses tended to be just below the midpoint both in the Low Cost condition ($M = 1.87$, $SD = 1.38$, $Mdn = 2.00$), and in the High Cost condition ($M = 2.38$, $SD = 1.32$, $Mdn = 3.00$). For the affective question, responses again tended to be just below the midpoint both in the Low Cost condition ($M = 1.94$, $SD = 1.37$, $Mdn = 2.00$) and in the High Cost condition ($M = 2.28$, $SD = 1.20$, $Mdn = 2.00$). The findings from Study 1a were consistent with our predictions, providing support for the hypothesis that people's sense of commitment to a joint activity can be enhanced as a function of their partner's investment of effort in the joint activity. In order to ensure that our findings were not due

to any incidental features of the scenario, we ran a replication study using a different scenario, and predicted the same pattern of results.

b. Study 1b

Methods

As in Study 1a, we used Amazon M-Turk to implement a web-based paradigm with a between-subjects design, aiming for a sample size of 200 participants (2 conditions, 100 per group). We again included data from those participants who had already begun the experiment when M-Turk registered that this number had been reached. Our data set therefore comprised 205 adults. After discarding the data from participants who failed the control question or failed to complete the questionnaire ($N = 5$), the final sample included 200 participants (105 female; $M_{age} = 38.18$ years, $SD = 11.85$) -- 94 in *High Cost Condition* and 106 in *Low Cost Condition*. The research was carried out in accordance with the international ethical requirements of psychological research and approved by the EPKEB in XXX. All participants gave their informed consent by ticking a box prior to the experiment.

The procedure was identical to the procedure of Study 1a, except that we implemented a different scenario and different control questions (See Appendix for the full vignette).

The questions were presented to the participants in a randomized order, except for the indirect question, which was always presented last.

Results and Discussion

The results of Study 1a were replicated. For the normative question, participants again gave higher estimates in the High Cost condition ($M = 2.56$, $SD = 1.46$) than in the Low Cost condition ($M = 1.79$, $SD = 1.39$), $t(198) = 3.828$, $p < .001$, Cohen's $d = 0.54$ (medium effect size). These results were confirmed by additional nonparametric tests, Mann-Whitney $U = 3484.000$, $p < .001$ (see **Figure 2a**). For the affective question, participants

again gave higher estimates in the High Cost condition ($M = 2.50$, $SD = 1.41$) than in the Low Cost condition ($M = 1.68$, $SD = 1.28$), $t(198) = 4.317$, $p < .001$, Cohen's $d = 0.61$ (medium effect size). These results were confirmed by additional nonparametric tests, Mann-Whitney $U = 3318.000$, $p < .001$ (see **Figure 2b**).

Responses to the indirect question showed the expected pattern, with participants giving higher estimates in the High Cost condition ($M = 4.29$, $SD = 2.73$) than in the Low Cost condition ($M = 4.05$, $SD = 2.85$), but there was again no statistically significant difference between the two conditions, $t(198) = .606$, $p = .545$.

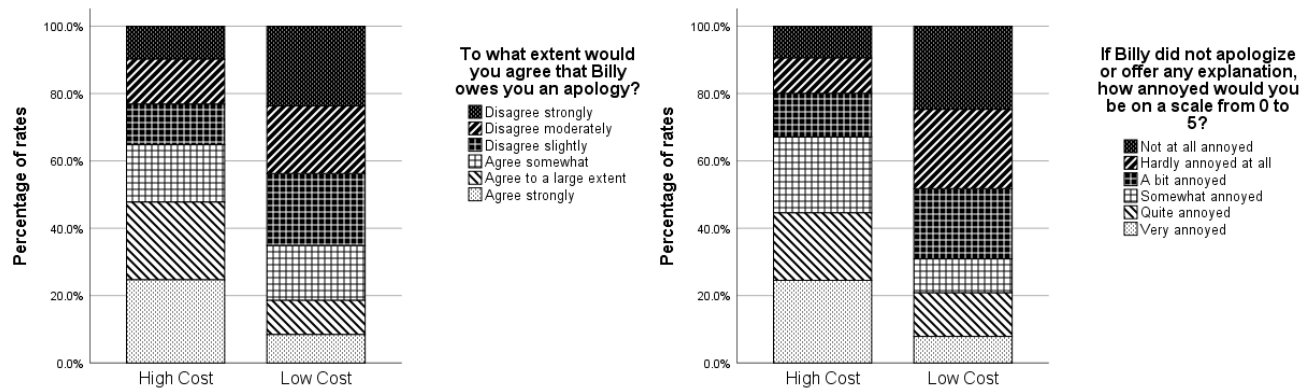


Fig. 2a and 2b. Percentage of responses to the normative and the non-normative question. White background bars indicate a mild-to-strong agreement, whereas black background bars indicate a mild-to-strong disagreement with the statement. In other words, the stronger the agreement, the higher the perception that a commitment had been violated.

We again found that responses to both the normative question and the affective question tended to cluster around the middle of the scale rather than towards the two extremes (see **Figure 3**). Indeed, for the normative question, responses tended to be around the midpoint both in the Low Cost condition ($M = 1.81$, $SD = 1.40$, $Mdn = 2.00$) and in the High Cost condition ($M = 2.57$, $SD = 1.45$, $Mdn = 3.00$). For the affective question, responses tended to be around the midpoint both in the Low Cost condition ($M = 1.70$, $SD = 1.29$, $Mdn = 2.00$), and in the High Cost condition ($M = 2.52$, $SD = 1.41$, $Mdn = 2.00$).

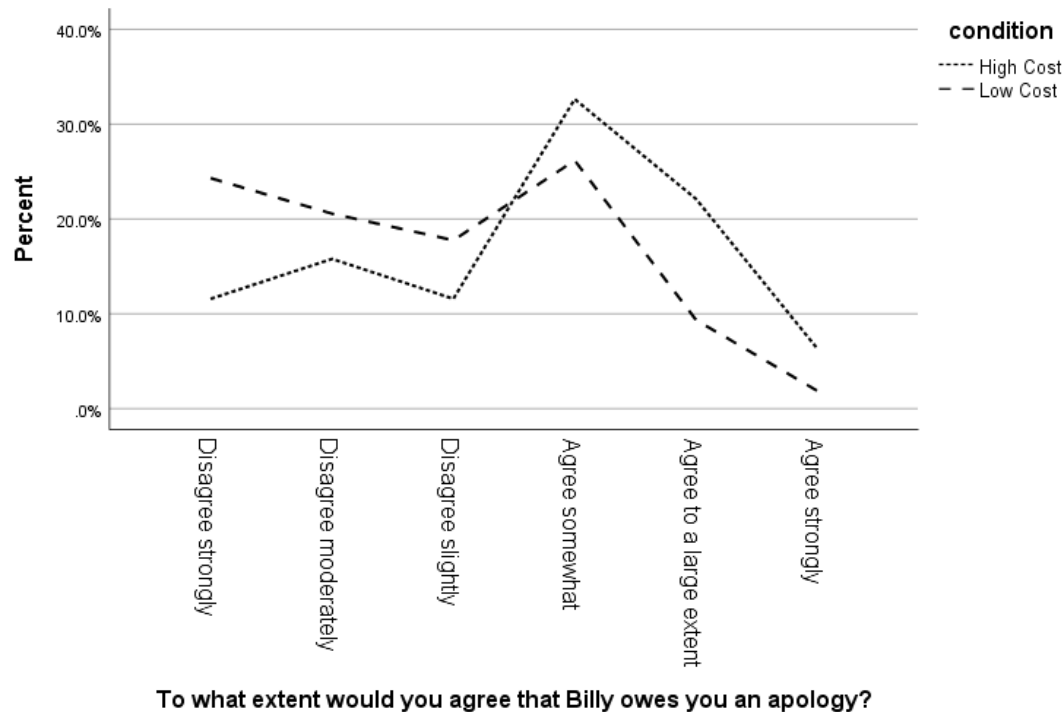


Fig. 3. Distribution of responses to the normative question. Although in the Low Cost condition there is a significantly higher percentage of responses at the lower end of the scale than in the High Cost condition, we can see that the largest number of participants in both groups give responses just above the midpoint of the scale.

The findings from Study 1b replicate those from Study 1a in a different scenario, which constitutes strong evidence for our hypothesis that one's sense of commitment to a joint activity can be enhanced as a function of her partner's investment of effort in the joint activity.

3. Repetition and Commitment

The second pair of studies was designed to test the hypothesis that the repetition of a joint activity can enhance people's sense of commitment to that joint activity. To this end, we presented participants with vignettes describing everyday situations in which an implicit commitment between two agents was violated. We again operationalized the sense of commitment with both normative and non-normative measures (i.e. with the normative, the affective and the indirect question), as we did in Studies 1a and 1b. We

marked in bold those parts of the text that implemented the manipulation (i.e. the phrases ‘three years’ and ‘three days’) in order to ensure that participants would not fail to notice these apparently minor details which might be overlooked by a casual reader.

a. Study 2a

Methods

As in Studies 1a and 1b, we used Amazon M-Turk to implement a web-based paradigm with a between-subjects design, and again aimed for a sample size of 200 participants (2 conditions, 100 per group). As in the previous studies, we included data from those participants who had already begun the experiment when M-Turk registered that this number had been reached. Our data set therefore comprised 210 adults. After discarding the data from participants who failed one or more control questions ($N = 14$), the dataset included 196 data from participants, 97 in the *High Repetition Condition* and 99 in the *Low Repetition Condition* (109 female; $M_{age} = 37.74$ years, $SD = 11.62$). The research was carried out in accordance with the international ethical requirements of psychological research and approved by the EPKEB in Hungary. All participants gave their informed consent by ticking a box prior to the experiment.

The procedure employed was the same as Studies 1a and 1b. The scenario reads as follows:

High Repetition: You and Pam work in the same office building. **Every day for the past 3 years**, you and Pam have spent your coffee break sitting out on the balcony and chatting, though you never agreed to start doing this. The sequence is broken when one day you walk up to the balcony and wait for Pam during the coffee break, but she doesn't turn up. This is surprising given that it hasn't happened in the past 3 years.

In the Low Repetition condition, the vignette differs insofar as the coffee break routine was initiated only three days rather than three years earlier (See Appendix for the full vignette).

Again, participants were asked to respond to normative and non-normative questions. In light of participants' feedback to a pilot version of the study, we opted to introduce a milder

normative measure than that used in Studies 1a and 1b. Specifically, we asked participants to evaluate whether the partner who had violated the implicit commitment owed them an explanation (rather than an apology). Also, we opted for an implicit question that was tailored to the manipulation of repetition rather than costs -- i.e. rather than probing participants' willingness to pay a cost to honor the commitment (as in Studies 1a and 1b), we asked about their willingness to resume the routine. The questions were presented to the participant in the following order:

Normative Question: *"On a scale from 0 to 5, to what extent would you agree that Pam owes you an explanation?" (0= Disagree strongly; 5= Agree strongly)*

Affective Question: *"If Pam did not apologize or offer any explanation, how annoyed would you be on a scale from 0 to 5?" (0 = not at all annoyed; 5 = highly annoyed)*

Implicit Question *"How interested would you be in spending your coffee break with Pam the next day?" (Not at all interested, Hardly interested at all, A bit interested, Somewhat interested, Quite interested, Highly interested)*

Control Question: *"In the scenario, where is it that you and Pam spend the coffee break?" (On the balcony, At the cafeteria, In the lounge)*

As in the previous studies, the normative question was designed to tap participants' explicit moral evaluations of the scenario. We predicted that they would more strongly agree that an explanation was in order in the High Repetition condition than in the Low Repetition condition. The affective question was designed to tap participants' more intuitive, emotional reactions to the commitment violation described in the scenario. We predicted that participants would indicate a higher level of annoyance if no apology or explanation were forthcoming in the High Repetition condition. The control question was designed to filter out participants who had not read the vignette with sufficient care to retain the critical information presented therein. The implicit question was intended to tap participants' implicit appraisal of the commitment, namely by measuring their willingness to restore the routine if they were in the position of the individual described in the scenario. Despite the fact that no straightforward prediction logically follows from our theory, we

reasoned that participants would indicate a lower willingness to restore the routine in the High Repetition condition than in the Low Repetition condition, as the violation of the commitment in the former condition would have more serious consequences: this would be the case because people might deploy partner choice strategies following the violation of a commitment. If so, then the more severe the violation, the more serious would be the consequences for the violator.

Results and discussion

For the normative question, participants gave higher estimates in the High Repetition condition ($M = 3.13$, $SD = 1.54$) than in the Low Repetition condition ($M = 2.10$, $SD = 1.34$), $t(189) = 5.014$, $p < .001$, Cohen's $d = 0.73$ (medium effect size). Since the sample failed Levene's Test for equality of variance ($p = .018$), we also conducted nonparametric tests, which yielded consistent results, Mann-Whitney $U = 2890.000$, $p < .001$ (**see Figure 4a**).

For the affective question, participants again gave higher estimates in the High Repetition condition ($M = 3.03$, $SD = 1.66$) than in the Low Repetition condition ($M = 1.87$, $SD = 1.17$), $t(172) = 5.659$, $p < .001$, Cohen's $d = 0.86$ (large effect size). Since the sample failed Levene's Test for equality of variance, ($p < .001$), we again conducted nonparametric tests, which again yielded consistent results, Mann-Whitney $U = 2841.500$, $p < .001$ (**see Figure 4b**). These results confirm our prediction, providing support for the hypothesis that a joint activity which has been repeated over a longer period of time elicits a stronger sense of commitment than a joint activity that has been repeated only over a short period of time.

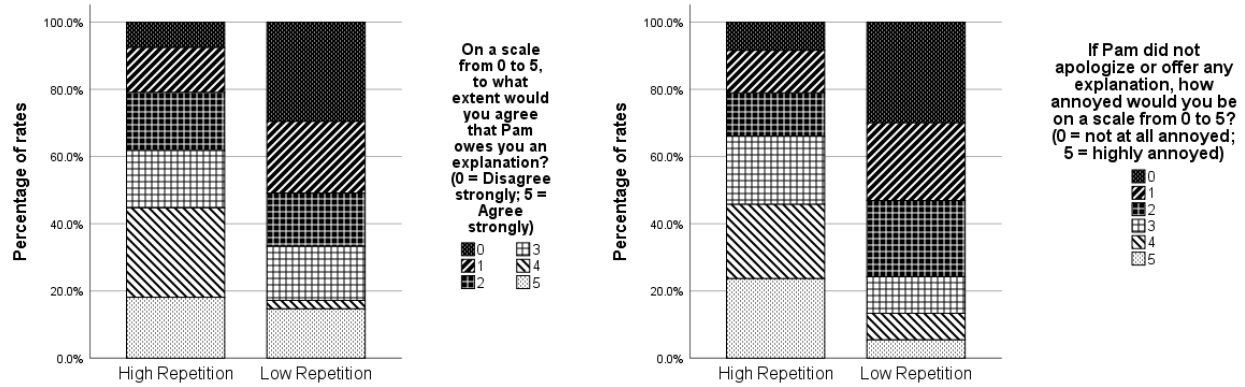


Fig. 4a and 4b. Percentages of responses to the normative and to non-normative question. White background bars indicate a mild-to-strong agreement, whereas black background bars indicate a mild-to-strong disagreement with the statement: in other words, the stronger the agreement, the higher the perception that a commitment has been violated.

The opposite pattern of results was found for responses to the implicit question. Participants indicated a higher degree of willingness to restore the previous routine after a commitment violation in the High Repetition Condition ($M = 5.02$, $SD = .85$) than in the Low Repetition condition ($M = 4.45$, $SD = 1.03$), $t(189) = 4.186$, $p < .001$, Cohen's $d = 0.64$ (medium effect size). The sample failed the Levene's Test for equality of variance, $p = .005$. Nonetheless, this pattern of results is confirmed by a nonparametric test, Mann-Whitney $U = 3302.500$, $p < .001$ (**see Figure 5**). Although these findings are not consistent with our prediction, we believe that they can be explained by hypothesizing that a longer history of interaction gives rise to a more stable sense of commitment, which continues to bind the two partners even after minor violations such as the ones described in both scenarios.

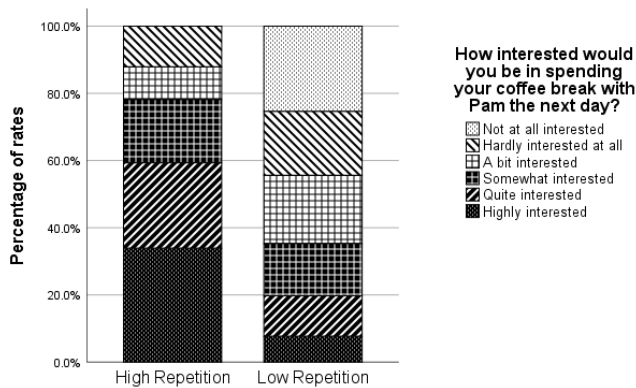


Fig. 5. Percentage of responses to the implicit question. White background bars indicate a mild-to-strong disinterest, whereas black background bars indicate a mild-to-strong interest in restoring the previous routine: in other words, the stronger the interest, the greater the perception that a commitment is in place.

As in the previous studies, we found that responses did not cluster at the extreme ends of the scale, but tended to be distributed homogeneously across the scale (i.e. distributions were not skewed). For the normative question, responses tended to be right around the midpoint both in the Low Cost condition ($M = 2.22$, $SD = 1.40$, $Mdn = 2.00$), and in the High Cost condition ($M = 3.18$, $SD = 1.54$, $Mdn = 3.00$). For the affective questions, responses tended to be around the midpoint both in the Low Cost condition ($M = 2.01$, $SD = 1.30$, $Mdn = 2.00$), and in the High Cost condition ($M = 3.07$, $SD = 1.66$, $Mdn = 3.00$).

As previously designed, we ran a replication study with a different scenario, and we predicted the same pattern of results.

b. Study 2b

Methods

As in the previous studies, we used Amazon M-Turk to implement a web-based paradigm with a between-subjects design, and again aimed for a sample size of 200 participants (2 conditions, 100 per group). We again included data from those participants who had already begun the experiment when M-Turk registered that this number had been reached. Our data set therefore comprised 203 adults. After discarding the data from participants who failed the comprehension question ($N = 12$), the sample included 191 participants, 90 in *High Repetition Condition* and 101 in *Low Repetition Condition* (112 female; $M_{age} = 40.49$ years, $SD = 13.38$). The procedure was identical to Study 2a. The

research was carried out in accordance with the international ethical requirements of psychological research and approved by the EPKEB in Hungary. All participants gave their informed consent by ticking a box prior to the experiment.

The procedure was identical to the procedure of Study 2a, except that we implemented a different vignette and different control questions (See Appendix for the full vignette).

The questions were presented to the participants in a randomized order.

Results and Discussion

The results of the previous study were replicated. For the normative question, participants gave higher estimates in the High Repetition condition ($M = 2.29$, $SD = 1.51$) than in the Low Repetition condition ($M = 1.39$, $SD = 1.44$), $t(189) = 4.236$, $p < .001$, Cohen's $d = 0.62$ (medium effect size). This pattern of result is confirmed by nonparametric tests, both for the normative measure, Mann-Whitney $U = 2995.500$, $p < .001$ (see **Figure 6a**). For the affective question, participants again gave higher estimates in the High Repetition condition ($M = 1.93$, $SD = 1.44$) than in the Low Repetition condition ($M = 1.29$, $SD = 1.42$), $t(189) = 3.110$, $p = .002$, Cohen's $d = 0.45$ (medium effect size). This pattern of results is confirmed by nonparametric tests, Mann-Whitney $U = 3370.500$, $p = .002$ (see **Figure 6b**).

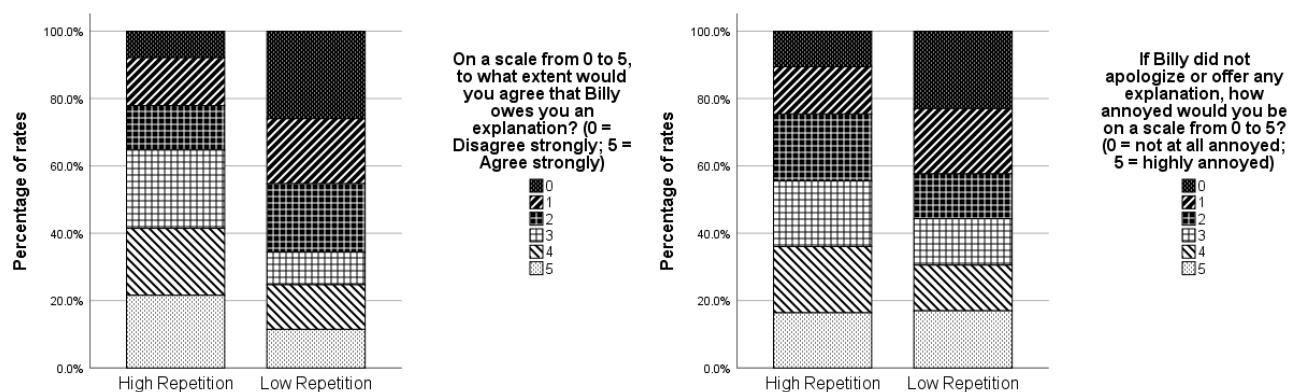


Fig. 6a and 6b. Percentage of responses to the normative and non-normative questions. White background bars indicate a mild-to-strong agreement, whereas black background bars indicate a mild-to-strong disagreement with the statement: in other words, the stronger the agreement, the higher the perception that a commitment has been violated.

As in Study 2a, responses to the implicit question exhibited the opposite pattern to what we had predicted. Participants reported being more willing to restore the previous routine after a commitment had been violated following a longer repeated interaction, giving higher estimates in the High Repetition condition ($M = 5.04$, $SD = .96$) than in the Low Repetition condition ($M = 4.44$, $SD = 1.10$), $t(189) = 4.020$, $p < .001$, Cohen's $d = 0.58$ (medium effect size). The sample failed the Levene's Test for equality of variance, $p = .008$. Nonetheless, this pattern of results was confirmed by a nonparametric test, Mann-Whitney $U = 3073.000$, $p < .001$ (see Figure 7).

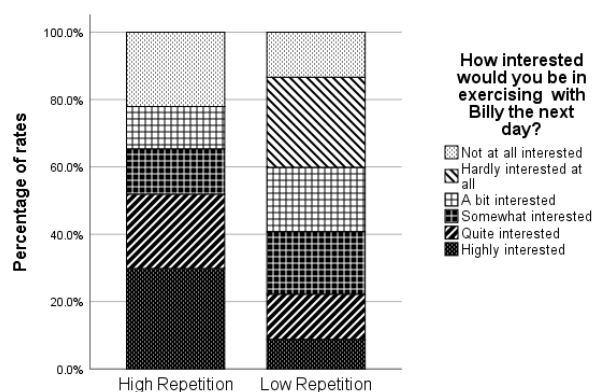


Fig. 7. Percentage of responses to the implicit question. White background bars indicate a mild-to-strong disinterest, whereas black background bars indicate a mild-to-strong interest in restoring the previous routine: in other words, the stronger the interest, the greater the perception that a commitment is in place.

As in the previous set of studies, we again found that responses to both the normative question and the affective question tended to cluster around the middle of the scale rather than towards the two extremes (see Figure 8). For the normative question, responses tended to be around the midpoint both in the Low Repetition condition ($M = 1.53$, $SD = 1.54$, $Mdn = 1.00$) and in the High Repetition condition ($M = 2.31$, $SD = 1.50$, $Mdn = 3.00$). For the affective question, responses tended to be around the midpoint both in the Low Repetition condition ($M = 1.43$, $SD = 1.52$, $Mdn = 1.00$) and in the High Repetition condition ($M = 1.97$, $SD = 1.45$, $Mdn = 2.00$).

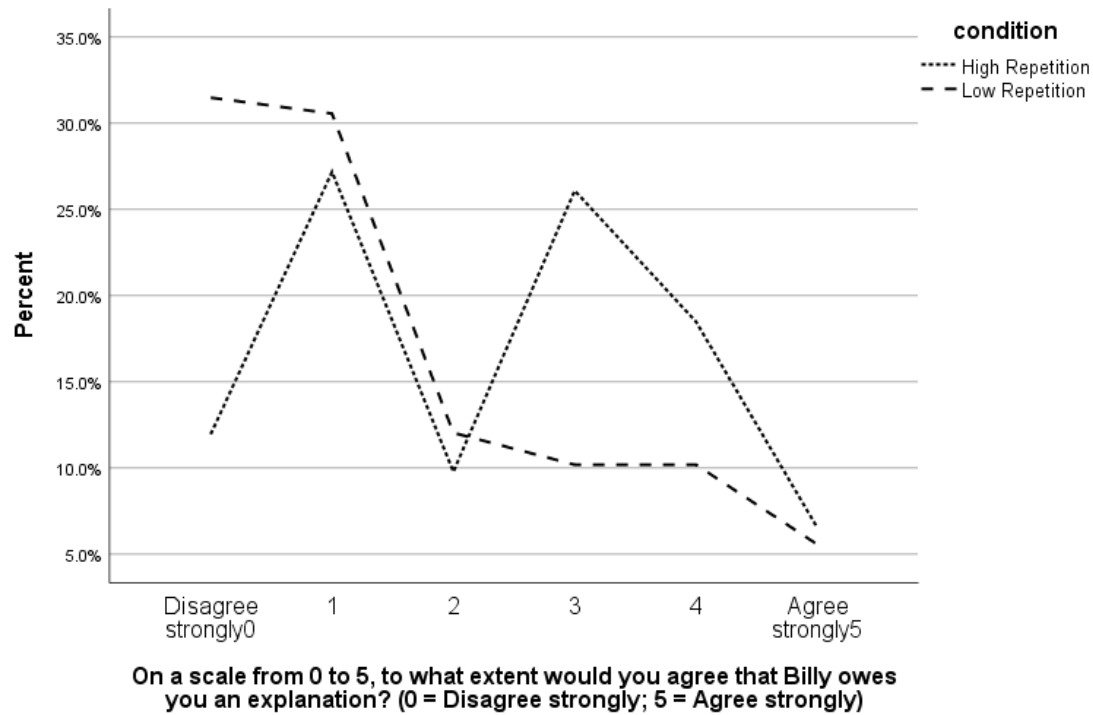


Fig. 8. Distribution of responses to the normative question. Although in the Low Repetition condition there is a significantly higher percentage of responses at the lower end of the scale, we can see that in the High Repetition condition the largest number of participants give responses right around the midpoint of the scale.

4. General discussion

In Studies 1a and 1b, we presented participants with vignettes describing a scenario in which one agent had either high expectations (generated by the investment of either a high degree of effort into a joint activity, i.e. the High Cost Condition) or low expectations (generated by a low degree of effort into a joint activity, i.e. the Low Cost Condition), and a second agent failed to remain committed. In line with our predictions, the results revealed that participants made more negative normative judgments and reported more negative emotional attitudes in response to the High Cost condition than the Low Cost Condition. Studies 2a and 2b were designed to probe participants' normative evaluations and affective attitudes in response to scenarios in which one agent failed to remain

engaged to a joint activity toward which her partner had either high expectations (generated by a longer history of repeating the routine, i.e. the High Repetition Condition) or low expectations (generated by a shorter history of repeating the routine, i.e. the Low Repetition Condition). Again, the results confirmed our predictions: the scenario described in the High Repetition condition elicited more negative normative judgments and emotional responses than the scenario described in the Low Repetition condition. Taken together, these results provide support of the hypothesis that people's judgments and attitudes about implicit commitments are governed by an implicit sense of commitment, which is modulated by cues to others' expectations, and by cues to the costs others have invested on the basis of those expectations.

Previous studies have shown that the opportunity cost paid by a partner incentivizes prosociality (Charness & Rabin 2010), and that a partner's apparent investment of effort costs in a joint task increases the people's persistence on the task (Székely & Michael 2018). These previous findings regarding the relevance of costs for implicit commitment, however, are also consistent with alternative explanations. The costs invested by a partner in order to engage in a joint task can also be interpreted as a cue to the value of the task itself, leading to higher persistence in the task. The same is true of another convergent line of research showing that participants with a history of successful coordination tend to behave more cooperatively when facing a social dilemma (Guala & Mittone, 2010; Rusche & Luetge, 2016), although it is tempting to interpret such findings as evidence that repeated coordinated interaction might signal reciprocal expectations, and that people may therefore be sensitive to such cues when reasoning about reciprocal commitments. By using both normative and non-normative measures, we were able to rule out alternative explanations. Specifically, our finding that participants were more likely to judge that an apology was in order in the conditions in which we had induced participants to perceive a higher degree of implicit commitment, a pattern consistent with the non-normative, emotional responses, and that cannot be explained by appealing to an increase in the perceived value of the task. In other words, the fact that responses to the normative and the non-normative questions provided a consistent picture suggests that people were not simply expressing their frustration with the outcome presented in the experiment or their disappointment about having missed out on a valuable activity, but

that costs and repetition are two factors that are reliably interpreted as cues to others' expectations. Both the normative and the non-normative (affective) questions reliably elicited higher estimates in the High Cost/ High Repetition conditions. This clearly supports the hypothesis that these two factors enhance people's sense of commitment in joint activity.

In Studies 1a and 1b, investigating the role of costs, responses to our indirect question manifest the same trend, but the difference did not reach statistical significance. This may be because our measure was too weak to pick up on participants' willingness to pay a cost in order to maintain the commitment, or because it was too unrealistic -- participants might have enough familiarity with charging phones to assess that four minutes should be enough to be able to send a message. In Studies 2a and 2b, investigating the role of repetition, the implicit measure yielded the *opposite* results to what we had predicted. Our rationale in formulating that question was that the longer the history of repeated interaction, the greater the disapproval of a violation of the routine. This, we predicted, would lead participants to be less inclined to resume the previous routine following a violation. What this rationale did not take into account, however, is that the longer history of interaction may also give rise to a more stable sense of commitment, which may continue to bind the two partners even after a minor violation. Thus, although the observed results did not confirm our prediction, we believe that they are indeed highly consistent with our hypothesis.

Our results provide further empirical evidence in support of some influential theories of social norms according to which people ought to fulfil others' preferences when they are *reasonably* expected to do so (Lewis, 1969; Bicchieri, 2005). The notion of reasonable expectation is at the core of Lewis' Presumption Reason: Agent A's expectation that agent B will perform an action X is reasonable if A has well-grounded reasons to believe that A will do X. According to Sudgen, this moral principle rests upon features of human psychology that enable a motivation to abide by it, such as an aversion to disappoint others' reasonable expectations (Sudgen, 2000). And indeed, it has been found that people exhibit an aversion to disappointing others' expectations when those expectations have been made explicit, but only when these expectations were not unreasonable (Heintz et al. 2015). Our results provide further empirical evidence in support of these

theories of social norms, demonstrating that people judge there to be an obligation to fulfil others' reasonable expectations even when these expectations have not been made explicit (but have been implicitly cued).

Our findings also have important implications for theorizing about the relationship between implicit and explicit commitments (e.g., promises). According to an influential theory of promises (see, Scanlon, 1998), the moral ground for the norm that we ought to keep our promises (and, presumably, explicit commitments in general) is that promises generate expectations (i.e. promising to do X creates in the recipient the expectation that the speaker will do X). As shown by our studies, others' (reasonable) expectations also ground implicit commitments in people's moral judgments. Thus, it might be argued that explicit and implicit commitments share the same moral ground, i.e. that we ought to act in accordance with the expectations others have of us.

Also related to theoretical research on promises, our findings challenge the idea of promissory commitment as a binary notion, according to which either one is committed (i.e., if all conditions for promising are met), or one is not committed (Searle, 1969). This way of conceptualizing promises leaves little room for the idea that recipients' desires and expectations might modulate the promisor's sense of commitment in a graded manner. Since promise-breaking is considered to be a violation of a specific norm (i.e., a violation of the norm that one ought to keep one's promises; see Hume, 1739/1978), one might predict that if there was an expectation that the speaker would perform a certain action, violating a promise to perform that action would always be considered blameworthy (on both normative and affective measures) -- *independently of the magnitude of the expectation*. In contrast to this, we found that given a 6-point scale, participants' assessments of commitment were distributed at intermediate points along the scale rather than at opposite poles. These results foster the idea that for implicit commitments, people assess accountability in a graded manner. Future studies could investigate the effect of recipient's mental attitudes on normative and emotional measures of commitment violation when the commitment has been created by a promise, which may challenge the philosophical conception of promises as binary sources of commitment.

Finally, our findings open up several new avenues for additional further research on implicit commitment. For instance, they raise the question whether different kinds of costs (time, effort, money, etc) may elicit a sense of commitment in different ways, which may be reflected in different reparation strategies or in reactions other than moral disapproval. Moreover, while we focused on those costs agents pay to enter into or to carry out a joint activity, it would be interesting to investigate the effects of costs that agents pay as a consequence of commitment violations. Finally, it would be interesting to investigate whether people's responsiveness to cues such as those implemented in our studies has an impact on subsequent partner choice.

To sum up, our studies shed some light on the way people prioritize and evaluate commitments, showing that people are not only sensitive to others' expectations in judging whether commitments are in place, but that they even "sense" commitments when expectations are only *implicitly* cued (e.g., by the amount of costs that one agent is investing in the interaction and by the history of repeated interactions). This sensibility allows people to act together and respond to each others' expectations even in the absence of explicit agreements, promises, or contracts, and might even be at the basis of the norms that define these acts.

5. Acknowledgments

This work was supported by a Starting Grant from the European Research Council (nr 679092, SENSE OF COMMITMENT). We would like to thank Jinnie Ooi and Eszter Salamon for assistance with data collection, as well as Christophe Heintz, Thom Scott-Phillips and Barbora Siposova for helpful comments and discussion.

6. References

- Bicchieri, C. (2005). *The grammar of society: The nature and dynamics of social norms*. Cambridge, MA: Cambridge University Press.
- Bratman, M. (1993). Shared Intention. *Ethics*, 104, 97-113.
- Charness, G., & Rabin, M. (2010). Understanding Social Preferences with Simple Tests. *The Quarterly Journal of Economics*, 117(3), 817–869.

- Clark, H. H. (2006). Social actions, social commitments. In S.C. Levinson, N.J. Enfield (Eds.), *Roots of human sociality: Culture, cognition and interaction*, (pp. 126-150). New York: Bloomsbury.
- Dana, J., Cain, D. M., & Dawes, R. M. (2006). What you don't know won't hurt me: Costly (but quiet) exit in dictator games. *Organizational Behavior and Human Decision Processes*, 100(2), 193–201.
- Gilbert, M. (2006). *A Theory of Political Obligation*. Oxford: Oxford University Press.
- Guala, F. & Mittone, L. (2010) How History and Conventions Create Norms: An Experimental Study. *Journal of Economic Psychology*, 31, 749-756
- Heintz, C., Celse, J., Giardini, F., & Max, S. (2015). Facing expectations : Those that we prefer to fulfil and those that we disregard. *Judgment and Decision Making*, 10(5), 442–455.
- Hume, D. (1978). *A Treatise of Human Nature*. Oxford: Clarendon Press. (Original work published in 1739-1740).
- Lewis, D. (1969). *Convention: A Philosophical Study*. Cambridge, MA: Harvard University Press.
- MacCormick, N., & Raz, J. (1972). Voluntary Obligations and Normative Powers. *Proceedings of the Aristotelian Society, Supplementary Volumes*, 46(1972), 59–102.
- Michael, J., & Pacherie, E. (2015). On Commitments and Other Uncertainty Reduction Tools in Joint Action. *Journal of Social Ontology*, 1(1). <https://doi.org/10.1515/jso-2014-0021>
- Michael, J., Sebanz, N., & Knoblich, G. (2016). The Sense of Commitment: A Minimal Approach. *Frontiers in Psychology*, 6, 1968. <https://doi.org/10.3389/fpsyg.2015.01968>
- Ockenfels, A., & Werner, P. (2012). “Hiding behind a small cake” in a newspaper dictator game. *Journal of Economic Behavior & Organization*, 82(1), 82–85.
- Rusch, H., & Luetge, C. (2016). Spillovers from coordination to cooperation: Evidence for the interdependence hypothesis? *Evolutionary Behavioral Sciences*, 10(4), 284–296. Retrieved from <http://psycnet.apa.org/buy/2015-58974-001>
- Scanlon, T. M. (1998). *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.
- Searle, J. R. (1969). *Speech Acts: An essay in the philosophy of language*. Cambridge, MA: Cambridge University Press.
- Sugden R. (2000) The Motivating Power of Expectations. In J. Nida-Rümelin, W. Spohn (Eds.) *Rationality, Rules, and Structure. Theory and Decision Library* (pp. 103-129) Dordrecht: Springer.

Székely, M., & Michael, J. (2018). Investing in commitment: Persistence in a joint action is enhanced by the perception of a partner's effort. *Cognition*, 174, 37-42. DOI: 10.1016/j.cognition.2018.05.009

Zajonc, R.B. (1980). Feeling and thinking: preferences need no inferences. *American Psychologist*, 35, 151–175.

Zajonc, R.B. (2001). Mere exposure: a gateway to the subliminal. *Current Directions in Psychological Science*, 10, 224–228.