

Infants Expect Some Degree of Positive and Negative Reciprocity between Strangers



Open Access This file is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. In the cases where the authors are anonymous, such as is the case for the reports of anonymous peer reviewers, author attribution should be to 'Anonymous Referee' followed by a clear attribution to the source work. The images or other third party material in this file are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit

<http://creativecommons.org/licenses/by/4.0/>.

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

This is a wonderful set of experiments designed to assess toddlers' expectations of reciprocity. The control condition(s) rule out some important alternative explanations. Overall, the methods and analyses and literature review are all excellent. The implications of these studies certainly will generate a number of interesting/important follow-ups.

I thought the adult data in the supplementary file could be briefly mentioned in the Discussion where the authors mentioned future work looking at tasks that might be more costly to perform. It seemed they could acknowledge the finding that adults might judge negatively when an agent doesn't help (when it costs little to do so).

I wonder whether the authors considered that their findings might be revealing something specific about the behavior of groups and expectations of how group members might interact rather than this being about individuals. Specifically, while the authors noted that no intentional cues to group membership were provided, is it possible that children assume that people coordinating action together, appearing together, turn taking, might be a cue that they're part of some group? The control conditions don't have such cues since agent3 comes in after one agent departs and in the other control task, agent2 doesn't enter until after agent1 left. Might the low-level novelty bias be there for reasoning about behavior of group members?

Have the authors considered another way of ruling out the low-level novelty bias might be to look at evaluative non-social behaviors and non-evaluative social behaviors? I imagine reciprocity expectations might be specific to (evaluative) social behaviors but perhaps not to other kinds of behaviors. Thus, you might not get an increase in looking-time in these other behavior types when there is a mismatch between action 1 and action 2.

Have the authors considered whether children of this age might have different expectations of reciprocity depending on who the agents are? Ie, if one is shown to be more dominant than the other? It could also be interesting to consider whether they have different expectations of reciprocity depending on whether the positive/negative behavior initially performed was intentional or accidental.

In some of the earlier literature on moral reasoning, researchers suggested that children only evaluate the outcome of an action when determining if it is good/bad, not the intention. Here, it suggests that children are judging bad outcomes differently depending on whether the agent(2 or 3) was aware of what agent1 was doing (thinking specifically about the control groups). If I'm correct, it might be worth acknowledging that the present data add to the body of the literature that suggests children's reasoning is more sophisticated than previously thought.

I could see some value in considering some of these ideas in the Discussion to highlight a broader set of questions this research opens the door to.

Great set of experiments!

Reviewer #2 (Remarks to the Author):

An interesting set of studies focusing on the question of whether 15-month-olds engage in positive and negative direct reciprocity. The set of conditions included is largely convincing, and the inclusion of different control conditions as well as the extension to withholding behavior as opposed to just engaging in behavior (as the authors note, consistent with partner choice perspectives) lend strength to the paper.

The authors could do more to discuss their findings in light of inconsistent findings with older children, most notably the several studies in a paper the authors do not cite or consider, Cherynak et al., 2019, Psych Science. In that paper the critical finding relevant to this work is that children between about age 3 and 7 do not engage in direct positive reciprocity on their tasks; rather, they engage in indirect or upstream reciprocity, in that after being a benefactor of a positive action they are no more likely to direct their own positive action towards their benefactor than towards

another individual. With the inclusion of the new-agent condition the authors of the present paper do have something that looks like evidence of direct reciprocity and a lack of indirect or upstream reciprocity. What do the authors make of this difference and the apparent u-shaped curve it applies? Of course there are important differences beyond age of participants, such as the Chernyak paper being first person, many task differences, etc. But it is an interesting discrepancy that is worthy of some discussion. In particular one might wonder if the new-agent had been present in the earlier part of the study whether infants would still engage in direct rather than indirect reciprocity.

I also note that this paper would be stronger if the authors had included an inconsistent condition in Exp 3; at present all evidence for reciprocity in the paper comes from comparisons with a single condition of $N = 40$ from Exp 2 (which is the comparison condition in both Exp 2 and 3); it is a shame not to have an internal replication in a work reporting a major new finding in infancy, especially given some differences with results found with older children (as discussed above).

Reviewer #3 (Remarks to the Author):

This manuscript describes experiments with 15 month old infants, testing their expectations of direct reciprocity. This is a valuable contribution to the literature on children's early understanding of the foundations of cooperation. Indeed, I was somewhat surprised that direct reciprocity has not been tested in infants given that indirect reciprocity has. That said, there are several areas of the manuscript that require revision and clarification.

In the introduction, the authors should describe the difference between direct and indirect reciprocity and why it is important. In prior findings, infant studies such as this one, Hamlin, Wynn, Bloom & Mahajan (2011), should be included. A more recent study has also found that preschoolers more readily engage in direct negative reciprocity than positive reciprocity even though they are capable of both (Chernyak et al., 2019).

The set of helper/hinderer studies described in lines 116-122 is third party but is not about reciprocity and so seems out of place in this lead up to the current study.

The presentation of the three experiments was a bit puzzling. Experiment 1 seems like a preliminary test as opposed to an actual experiment. Only familiarization trials are used, no test trials, and a null result is predicted. I understand that the authors needed to establish that infants were not more surprised by stealing than giving at baseline, but this should be moved to a supplement.

Experiments 2 and 3 provide the main results. I must admit that I was surprised at the choice of actions for negative and positive reciprocity. By 18 months, children are highly sensitive to property destruction but "cleaning up" seems a far less dramatic positive act. The results here suggest that infants are not surprised by property destruction but this strikes me as odd, but this might be a discussion point.

In Experiment 2, the split control condition requires more theoretical motivation. Using a new agent is an important control because infants may recognize generalized reciprocity as opposed to direct reciprocity. The no-knowledge group serves a different function, but the explanations for these controls should be set up before we are in the methods section. Otherwise these controls seem orthogonal to the main research question which suggests the main contrast should be with the consistent trials that appear as a separate experiment.

In Experiment 3, another subgroup split occurs but this one I do not understand. 20 infants get the consistent version, which acts as a critical control for Experiment 2. But the other 20 infants are in a no-engagement group. Here, agent 2 could help the person who helped them but opts not to or does not harm the person (their property) who harmed them. There are no differences between these groups, but I think there should be, at least for a mature understanding of reciprocity. Failing to help a person who helped you is pretty mean. I'm puzzled by the prediction that no engagement (particularly in the positive case) would be consistent with responding positively. I'm also not sure why one would add these variations that are intended to produce null results. Bayes factors to prove the null are fun but we design experiments to find differences. I suggest the authors revisit their logic here and consider whether they are underpowered to find differences between these sub groups.

The discussion section contains several references to social distance that were not part of the introduction and don't really help explain the results. Infants could just as well have a more categorical view of the social world; this study can't tell us either way. I would like to see the authors place the results in contrast to a mature understanding of reciprocity. They seem to show some version of direct reciprocity here, but it is certainly not the version available to 5 year olds.

March 28, 2024

Dear Reviewers,

Thank you for your constructive and thoughtful feedback on our manuscript. In response to your comments, we have made three main changes, while keeping in mind the strict word limits imposed by the journal.

--We have slightly expanded the introduction section of Experiment 1, in response to comments by R1 and R3, to better describe the motivation for this experiment and the conclusions that can be drawn from it.

--We have slightly expanded the introduction and discussion sections of Experiment 2, in response to comments by R1 and R3, to make clearer our design choices and the conclusions we draw from the experiment.

--We have considerably expanded the Discussion. First, both R2 and R3 were concerned about the apparent discrepancy between our results and those of Chernyak et al. (2019) on positive direct reciprocity. We first point out that these results are also inconsistent with most of the prior results on reciprocity in preschoolers we describe in the Introduction. We then point out that our results offer a plausible explanation for this discrepancy, and suggest a way of testing this explanation. Second, R1 and R3 provided suggestions for possible discussion topics, and we have incorporated several of them in our revised Discussion.

We believe these changes have considerably added to the clarity, contribution, and potential impact of our research, and we are grateful to the reviewers for their help. Below are point-by-point responses to each reviewer's suggestions.

R1

This is a wonderful set of experiments designed to assess toddlers' expectations of reciprocity. The control condition(s) rule out some important alternative explanations. Overall, the methods and analyses and literature review are all excellent. The implications of these studies certainly will generate a number of interesting/important follow-ups.

Many thanks for these positive and encouraging comments.

I thought the adult data in the supplementary file could be briefly mentioned in the Discussion where the authors mentioned future work looking at tasks that might be more costly to perform. It seemed they could acknowledge the finding that adults might judge negatively when an agent doesn't help (when it costs little to do so).

Thank you for this suggestion, we now include this point in the Discussion.

I wonder whether the authors considered that their findings might be revealing something specific about the behavior of groups and expectations of how group members might interact

rather than this being about individuals. Specifically, while the authors noted that no intentional cues to group membership were provided, is it possible that children assume that people coordinating action together, appearing together, turn taking, might be a cue that they're part of some group? The control conditions don't have such cues since agent3 comes in after one agent departs and in the other control task, agent2 doesn't enter until after agent1 left. Might the low-level novelty bias be there for reasoning about behavior of group members?

It is unlikely that infants viewed agent1 and agent2 as members of the same group. First, previous research indicates that infants need clear and explicit cues to signal group membership (e.g., marked similarities in labels or outfits), and no such clues were available in our experiments. We have now clarified this point in Experiment 1. Second, recall that in the familiarization trials of Experiments 1, 2, and 3, infants consistently viewed agent1's theft of agent2's cookie as acceptable; based on previous research, such an action would be unacceptable with an ingroup member. Third, recall that in Experiment 3, infants in the no-engagement subgroup also viewed it as acceptable for agent2 to ignore agent1's need for help; based on previous research, such a decision would again be viewed as unacceptable with an ingroup member. This last finding makes it clear that the fact that both agents again appeared together in the test trial did not provide additional information that led infants to view the two agents as ingroup members.

Have the authors considered another way of ruling out the low-level novelty bias might be to look at evaluative non-social behaviors and non-evaluative social behaviors? I imagine reciprocity expectations might be specific to (evaluative) social behaviors but perhaps not to other kinds of behaviors. Thus, you might not get an increase in looking-time in these other behavior types when there is a mismatch between action 1 and action 2.

For the purposes of this initial body of research, the new-agent and no-knowledge subgroups were better choices for controls because they allowed us to evaluate the novelty interpretation that infants in the inconsistent group were simply responding to a change in the valence of the evaluative social actions shown in the familiarization and test trials, irrespective of their targets (give/harm, steal/help). Infants responded to this change in valence in the inconsistent group but not in the control group, ruling out this alternative interpretation.

The more basic sort of control group R1 has in mind is one we have used many times in the past. Here one keeps the test trials the same but tweaks the familiarization trials instead, so that reciprocity no longer applies. This sort of control does not fully address the novelty interpretation described above, because a critic could always argue that this interpretation might apply only when evaluative social behaviors occur in both the familiarization and test trials. Nonetheless, a non-evaluative control is useful in showing that, in line with our chosen interpretation of the inconsistent group, a relevant modification of the familiarization trials can make the effect disappear in the test trials. We now describe a possible non-evaluative control in the Discussion, when outlining future directions for our research.

Have the authors considered whether children of this age might have different expectations of reciprocity depending on who the agents are? Ie, if one is shown to be more dominant than the other? It could also be interesting to consider whether they have different expectations of

reciprocity depending on whether the positive/negative behavior initially performed was intentional or accidental.

Thank you for these suggestions, we have now incorporated the first in the Discussion. (There is ample evidence in the developmental literature that infants can distinguish between intentional and accidental actions, so this suggestion would seem less novel to developmental researchers).

In some of the earlier literature on moral reasoning, researchers suggested that children only evaluate the outcome of an action when determining if it is good/bad, not the intention. Here, it suggests that children are judging bad outcomes differently depending on whether the agent (2 or 3) was aware of what agent1 was doing (thinking specifically about the control groups). If I'm correct, it might be worth acknowledging that the present data add to the body of the literature that suggests children's reasoning is more sophisticated than previously thought.

There is already ample evidence in the developmental literature that when evaluating an agent's action toward someone else, infants attend to whether the agent does or does not possess knowledge relevant to this interaction. We now refer to a few of these findings when introducing the no-knowledge subgroup in Experiment 2.

I could see some value in considering some of these ideas in the Discussion to highlight a broader set of questions this research opens the door to.

We entirely agree, and thank R3 for these many helpful suggestions, which have considerably enriched our Discussion.

Great set of experiments!

Thank you again for your support of our research!

R2

An interesting set of studies focusing on the question of whether 15-month-olds engage in positive and negative direct reciprocity. The set of conditions included is largely convincing, and the inclusion of different control conditions as well as the extension to withholding behavior as opposed to just engaging in behavior (as the authors note, consistent with partner choice perspectives) lend strength to the paper.

Thank you for these positive comments. We were particularly pleased that R2 liked our discussion of how our findings tie in with the partner-choice approach, which is less commonly discussed than it should be in the developmental literature.

The authors could do more to discuss their findings in light of inconsistent findings with older children, most notably the several studies in a paper the authors do not cite or consider, [Cherynak et al., 2019](#), Psych Science. In that paper the critical finding relevant to this work is that children between about age 3 and 7 do not engage in direct positive reciprocity on their

tasks; rather, they engage in indirect or upstream reciprocity, in that after being a benefactor of a positive action they are no more likely to direct their own positive action towards their benefactor than towards another individual. With the inclusion of the new-agent condition the authors of the present paper do have something that looks like evidence of direct reciprocity and a lack of indirect or upstream reciprocity. What do the authors make of this difference and the apparent u-shaped curve it applies? Of course there are important differences beyond age of participants, such as the Chernyak paper being first person, many task differences, etc. But it is an interesting discrepancy that is worthy of some discussion. In particular one might wonder if the new-agent had been present in the earlier part of the study whether infants would still engage in direct rather than indirect reciprocity.

I also note that this paper would be stronger if the authors had included an inconsistent condition in Exp 3; at present all evidence for reciprocity in the paper comes from comparisons with a single condition of $N = 40$ from Exp 2 (which is the comparison condition in both Exp 2 and 3); it is a shame not to have an internal replication in a work reporting a major new finding in infancy, especially given some differences with results found with older children (as discussed above).

Thank you for bringing up the work of Chernyak et al. (2019), which we agree appears to offer a striking counterpoint to our results. We actually had included a discussion of this work in an earlier version of our manuscript, but we had to delete it to stay within the word limits of the journal. However, encouraged by R2's comments, we now include a lengthy discussion of this research in our Discussion. As noted earlier, we first point out that their findings on positive direct reciprocity are inconsistent not only with our own findings but also with most of the developmental findings we discuss in the Introduction. Second, we point out that our results offer a possible explanation for this discrepancy, and we suggest a way of testing this explanation.

Because the results of our inconsistent group include 40 infants (sample sizes of 16 to 24 participants per group are common in violation-of-expectation tasks, so our sample went well beyond such numbers); because our results are consistent with the prior results on reciprocity in preschoolers that we review in the Introduction; and because, at least from our perspective, our results are also consistent with those of Chernyak et al. (2019), we believe that we are presenting sufficient evidence overall to support our conclusions. We therefore chose to leave it to future research to confirm and expand our findings, which already rely on 160 infants.

R3

This manuscript describes experiments with 15 month old infants, testing their expectations of direct reciprocity. This is a valuable contribution to the literature on children's early understanding of the foundations of cooperation. Indeed, I was somewhat surprised that direct reciprocity has not been tested in infants given that indirect reciprocity has. That said, there are several areas of the manuscript that require revision and clarification.

Thank you for these positive comments.

In the introduction, the authors should describe the difference between direct and indirect reciprocity and why it is important. In prior findings, infant studies such as this one, Hamlin, Wynn, Bloom & Mahajan (2011), should be included. A more recent study has also found that preschoolers more readily engage in direct negative reciprocity than positive reciprocity even though they are capable of both (Chernyak et al., 2019).

We agree with R3's comments, here and in the preceding paragraph, that there has been much more research interest in indirect than in direct reciprocity in infants and toddlers so far. However, we were concerned that delving into indirect reciprocity in the Introduction would make it too broad, given all we had to cover and the journal's word limits. In the Discussion, we do describe one of our previous studies on indirect reciprocity, more specifically third-party punishment, that echoes the present findings. We now add that readers can refer to that paper for a review of the extensive literature on this topic. The Hamlin et al. (2011) manuscript is included in that review, as it presents evidence that infants are more likely to take a treat away from a hinderer vs. a helper, and to give a treat to a helper vs. a hinderer.

As we noted in our response to R2, we also agree with R3 that including a discussion of the Chernyak et al. (2019) experiments in the Discussion is an excellent idea, and we described above how we have done so.

The set of helper\hinderer studies described in lines 116-122 is third party but is not about reciprocity and so seems out of place in this lead up to the current study.

We respectfully disagree with R3 on this point. The climber's decision to approach either the nice puppet who has helped it climb the hill or the mean puppet who knocked it to the bottom of the hill could arguably involve direct reciprocity. The logic of the design is similar to that of other prior studies on reciprocity we describe, even if the authors of these studies themselves do not highlight how their findings might potentially bear on reciprocity. We should also note that we are not the first to make such claims; see for example the review by Kuhlmeier et al. (2014) cited in our manuscript.

The presentation of the three experiments was a bit puzzling. Experiment 1 seems like a preliminary test as opposed to an actual experiment. Only familiarization trials are used, no test trials, and a null result is predicted. I understand that the authors needed to establish that infants were not more surprised by stealing than giving at baseline, but this should be moved to a supplement.

R3 is correct that we could have presented Experiment 1 in the Supplementary Online Materials (SOM)—or indeed even skipped it altogether and hoped that the results of the familiarization trials in Experiments 2 and 3 would turn out as expected. However, we believe that the more step-by-step approach we promote here has both *theoretical* and *methodological* value, and for this reason, we chose to keep Experiment 1 in the main text.

From the theoretical standpoint, we were particularly interested in whether infants have expectations about reciprocity between strangers, so it seemed essential to show that our young participants (a) perceived agent1 and agent2 as neutral outgroups or strangers and, as in prior

research, (b) perceived agent1's moderate positive and negative actions (i.e., giving and stealing cookies) toward neutral outgroup agent2 as acceptable. Experiment 1 was thus crucial for establishing that we were in the right ballpark, so to speak, and we have expanded the introduction and discussion sections of Experiment 1 to make this point clearer.

From a methodological standpoint, because our VOE method used looking times as its main measure, it was also important to show that infants looked equally at our give and steal events. Longer looking times at either event would have made it more difficult to interpret longer test looking times in the inconsistent group, or at least, made it more difficult to isolate the contribution of direct reciprocity to these looking times.

Experiments 2 and 3 provide the main results. I must admit that I was surprised at the choice of actions for negative and positive reciprocity. By 18 months, children are highly sensitive to property destruction but “cleaning up” seems a far less dramatic positive act.

We were surprised that R3 construed agent2's action of storing the decorated paper agent1 left behind when she was interrupted by the bell as “cleaning up”. Simply putting things away or discarding them are actions that could be initiated without any reference to reciprocity. However, our data suggest that infants did interpret agent2's storing action as implicating reciprocity. Agent1 clearly liked decorating her papers with her stickers, she did these actions repeatedly in the pretest and test trials, and she smilingly admired each decorated paper before storing it into her special black-and-white box. After agent1 stole agent2's cookies, infants in the inconsistent condition viewed agent2's action of storing the decorated paper as unacceptable; after agent1 gave agent2 cookies, this same action was viewed as acceptable. These results suggest that infants did construe agent2's storing action as a helpful one, intended to help agent1 by completing the sequence of actions interrupted by the bell. We have added the word “treasure” when initially describing the box, to emphasize that it was more like a special box a child might use for precious possessions or keepsakes than a bin for discarding litter. Hopefully, our figures also help convey that agent1 is storing her prized decorated papers for safekeeping into her special box, rather than discarding them.

The results here suggest that infants are not surprised by property destruction but this strikes me as odd, but this might be a discussion point.

Our results are in line with prior research. A moderate negative action is deemed acceptable when aimed at non-ingroups, and this has been shown with a wide range of actions including not only stealing cookies, as we show in Experiment 1, but also hindering someone's goal, knocking down a tower they have built, crumbling a drawing they have made, flipping over a puzzle they have just completely, discarding a piece they need to complete a tower or a puzzle, fighting over a toy by pulling it back and forth while growling, and so on. The action of tearing up a sticker seemed to us in keeping with these other negative actions, and our data suggest that infants agreed. In particular, tearing up agent3's sticker was deemed acceptable in the new-agent subgroup, just as stealing and eating agent2's cookies was deemed acceptable in the familiarization trials. This prior research is cited in the Introduction to Experiment 1, and we now give a few more examples of moderate negative actions that infants and toddlers view as acceptable with non-ingroups, to make this point clearer.

In Experiment 2, the split control condition requires more theoretical motivation. Using a new agent is an important control because infants may recognize generalized reciprocity as opposed to direct reciprocity. The no-knowledge group serves a different function, but the explanations for these controls should be set up before we are in the methods section. Otherwise these controls seem orthogonal to the main research question which suggests the main contrast should be with the consistent trials that appear as a separate experiment.

Let us begin with the second point. The difficulty with simply comparing the inconsistent and consistent groups was that obtaining the predicted results (longer looking times in the inconsistent group, shorter looking times in the consistent group) would not help rule out the novelty interpretation that infants in the inconsistent group were simply detecting a change in the valence of the actions shown, irrespective of their targets (give/tear, steal/help). Shorter looking times in the consistent group (give/help, steal/tear) would be in line with this interpretation, because there was no longer a change in the valence of the actions shown. By comparing the inconsistent and control groups in Experiment 2, however, we could immediately address this alternative interpretation.

As to why we chose to use two different versions of the control group: Broadly speaking, using different controls with similar purposes and predicted results can help provide better support for the conclusions drawn. Controls are often somewhat imperfect. In our research, for example, a critic might suggest that infants in the new-agent subgroup produced shorter test looking times because they were distracted by the appearance of a new agent; or they might suggest that infants in the no-knowledge subgroup produced shorter looking times because they were distracted by seeing agent2 open her doors for the first time, or became confused because her doing so increased the initial phase of the test trial by a few seconds (from 44 to 49 s). Neither of these alternative interpretations was particularly plausible, but we believed that both control subgroups together would yield stronger support for our conclusions.

Finally, in a very short article, it was not possible to introduce this secondary design issue at the start of Experiment 2. To keep our line of reasoning clear and easy to follow, we had to focus on the issues central to our main hypothesis.

In Experiment 3, another subgroup split occurs but this one I do not understand. 20 infants get the consistent version, which acts as a critical control for Experiment 2. But the other 20 infants are in a no-engagement group. Here, agent 2 could help the person who helped them but opts not to or does not harm the person (their property) who harmed them. There are no differences between these groups, but I think there should be, at least for a mature understanding of reciprocity. Failing to help a person who helped you is pretty mean. I'm puzzled by the prediction that no engagement (particularly in the positive case) would be consistent with responding positively. I'm also not sure why one would add these variations that are intended to produce null results. Bayes factors to prove the null are fun but we design experiments to find differences. I suggest the authors revisit their logic here and consider whether they are underpowered to find differences between these subgroups.

R3's reaction to the no-engagement subgroup is particularly interesting because it is similar to

the one we observed in our adult sample as a whole (see SOM): They rated agent2's decision not to help agent1 after she had given her a cookie in the plausible range, but still as less plausible than a decision to help her, perhaps because it would have cost agent2 so little to help out in this case (as we discuss in the manuscript, SOM, and in our response above to R1). However, our data suggest that infants in the no-engagement subgroup had a more categorical, black-and-white response to these events. Inspection of Table S1 indicates that infants in the no-engagement subgroup showed some of the lowest looking test times in our experiments, very much in line with their responses in the in-kind subgroup. As R3 suggests, this may point to a less mature view of reciprocity, and we now mention this possibility in the Discussion.

As to why we thought infants were likely to respond as they did, and hence pulled the in-kind and no-engagement subgroups together: We have found in the past that infants expect an agent not to ignore someone's need for help only when the helper and the helpee are clearly marked as belonging to the same social group. We thought it likely that infants would view the two agents as somewhat closer after agent1 gave agent2 cookies; but it seemed unlikely that they would now view the two to be as close as ingroup members, so we expected infants to produce their more typical, more categorical response (ok to help, ok not to help). Of course, this no-engagement hypothesis was also consistent with partner-choice accounts more generally. If we were all forced to become ingroups with any stranger who approached us with an unrequested, unneeded positive action, there would no longer be room for personal choice as to whom we want to affiliate with.

The discussion section contains several references to social distance that were not part of the introduction and don't really help explain the results. Infants could just as well have a more categorical view of the social world; this study can't tell us either way. I would like to see the authors place the results in contrast to a mature understanding of reciprocity. They seem to show some version of direct reciprocity here, but it is certainly not the version available to 5 year olds.

We have three comments here. First, we do discuss social distance in the Introduction, as well as the possible effects of slight decreases (give condition) and increases (steal condition) in social distance in interactions between strangers. Please see the fourth paragraph, where we speculate about a mechanism (in line with the seminal work of Trivers and others) that could help anchor an early expectation of reciprocity between strangers. We simply return to this same point in the Discussion, as we now remind readers.

Second, we now offer a fuller Discussion of our results with many suggestions for future research, including changes that might occur with development in young children's expectation of reciprocity (e.g., even if you don't want to become friends with her, you could still help her, especially when it costs you so little to do so).

Finally, if our proposed interpretation of the failure by Chernyak et al. (2019) to observe exact positive direct reciprocity in children under 7 years of age is correct, then it would mean that these children's performance was actually not all that different from that of our infants. Specifically, children did not view a positive gesture, in a situation where it was unlikely to constitute an overture of friendship, as a gesture that needed to be reciprocated, nor did they

expect their own positive gesture in this context to be reciprocated. Of interest is whether they would still expect a modicum of reciprocity, as our infants did (e.g., you don't have to give that girl avatar a sticker back, but it would be rather mean to steal from her as opposed to one of the other avatars).

Concluding Remarks

As our responses above hopefully convey, we are grateful to all three reviewers for their suggestions and comments, and we believe our manuscript has been greatly improved by the review process. We hope you will agree.

With best wishes,

Kyong-sun Jin

REVIEWERS' COMMENTS

Reviewer #1 (Remarks to the Author):

i think the authors did a fine job responding to and integrating the feedback of the reviewers. i think this manuscript will certainly make a clear contribution to the literature. great job.

Reviewer #2 (Remarks to the Author):

Overall the authors have done a solid job responding to the reviews . The expansions to the paper, in particular the discussion section, are useful and help clarify the importance of the work.

I was specifically asked to comment on whether and how the authors responded to the question about power I raised in my prior review, i.e. that results hinge on one study with two subconditions of $N = 20$ each, so a total of $N = 40$, which control condition in Study 2 and the main conditions of Study 3 are all compared to; there is also the more general issue that several hypotheses in the paper involve predictions of null results. With respect to the latter of these issues the authors do employ Bayes Factors, which help somewhat in quantifying support for the null.

With respect to the former issue, the authors dismiss the concern by saying the sample size of $N = 40$ per condition is consistent with norms in infant research. That seems correct. However, major infant findings in top journals generally include built-in replications across studies, such as could have been accomplished with a fresh violation condition in Study 3 rather than comparing back to the same data in Study 2. I continue to think that would have made these findings more compelling, but of course I also understand the demands of running additional studies.

Reviewer #3 (Remarks to the Author):

This revised manuscript addresses the concerns I raised in my first review either in the responses or the manuscript. I appreciate that the authors make a limited claim about direct reciprocity given that infants don't show expectations of responding in kind.

One suggestion is that in the discussion of the Chernyak paper the authors neglect the key fact that children did show negative reciprocity. This negates their "no reciprocity" alternative interpretation based on not knowing each other's identities, etc. And I think it would be interesting to test whether the asymmetry between negative and positive reciprocity appears in infants when the costs are raised.

RESPONSE TO REVIEWERS

The reviewers' original comments were very helpful in suggesting ways to improve our manuscript. We have done our best to address their comments in our revision, and we are delighted and grateful that they were generally pleased with our revised manuscript.

Below are detailed responses to their final comments.

Reviewer #1:

I think the authors did a fine job responding to and integrating the feedback of the reviewers. I think this manuscript will certainly make a clear contribution to the literature. Great job.

We thank the reviewer for these positive comments and for their many previous suggestions, which helped improve our manuscript.

Reviewer #2:

Overall the authors have done a solid job responding to the reviews. The expansions to the paper, in particular the discussion section, are useful and help clarify the importance of the work.

We thank the reviewer for these positive comments and for their previous suggestions concerning the Discussion. We agree that our revised Discussion better integrates our findings with prior work in the developmental literature.

I was specifically asked to comment on whether and how the authors responded to the question about power I raised in my prior review, i.e. that results hinge on one study with two subconditions of $N = 20$ each, so a total of $N = 40$, which control condition in Study 2 and the main conditions of Study 3 are all compared to; there is also the more general issue that several hypotheses in the paper involve predictions of null results. With respect to the latter of these issues the authors do employ Bayes Factors, which help somewhat in quantifying support for the null.

With respect to the former issue, the authors dismiss the concern by saying the sample size of $N = 40$ per condition is consistent with norms in infant research. That seems correct. However, major infant findings in top journals generally include built-in replications across studies, such as could have been accomplished with a fresh violation condition in Study 3 rather than comparing back to the same data in Study 2. I continue to think that would have made these findings more compelling, but of course I also understand the demands of running additional studies.

The reviewer is correct that replicating the results of our inconsistent group with an additional 40 infants would make our conclusions even more compelling. However, given that (a) our project already involves 160 usable infants; (b) our data are internally highly consistent; and (c) our data are also highly consistent with the prior work with preschoolers and infants described in the Introduction, we believe that our project, as it stands, makes a solid and valuable contribution to the field that other researchers will be able to build on.

Reviewer #3:

This revised manuscript addresses the concerns I raised in my first review either in the responses or the manuscript. I appreciate that the authors make a limited claim about direct reciprocity given that infants don't show expectations of responding in kind.

We thank the reviewer for these positive comments and for their many prior suggestions and requests for clarification, which helped strengthen our manuscript.

One suggestion is that in the discussion of the Chernyak paper the authors neglect the key fact that children did show negative reciprocity. This negates their "no reciprocity" alternative interpretation based on not knowing each other's identities, etc. And I think it would be interesting to test whether the asymmetry between negative and positive reciprocity appears in infants when the costs are raised.

On reflection, we agreed with the reviewer that labeling our first alternative interpretation of the negative findings obtained by Chernyak et al. as the "no-reciprocity" interpretation could be misleading, as the children they tested did show evidence of direct negative reciprocity (a finding we acknowledge at the start of our discussion of their work: "our findings and those summarized above contrast with other results suggesting that although children exhibit direct negative reciprocity by at least 4 years of age, they do not exhibit direct positive reciprocity until several years later"). We therefore revised our text to be more neutral.

Conclusion

We again thank the reviewers for their help with our manuscript, which has been significantly improved through the review process.

With best wishes,

Kyong-sun Jin