**Peer Review and Communication History**

**MS Title**: Children Expect Leaders to Oust Intruders, Refrain from Unprovoked Aggression, but Not to Be Generally Prosocial

**Author Names**: Ashley J Thomas, Silvia Navarro Hernandez, Emily Sumner, Barbara W. Sarnecka

**Submitted:** April 6, 2021

**Editor First Decision**: Revise & Resubmit

July 29, 2021

Dear Ashley Thomas,

I have now received all reviews of your manuscript, “Children expect leaders to oust intruders, to refrain from aggression, but do not expect leaders to be generally more prosocial” from two qualified researchers. I also independently read the manuscript before consulting these reviews. I agree that your manuscript has important strengths and also that there are some issues that need to be addressed. I therefore encourage you to submit a revised version for further consideration at Collabra: Psychology.

The reviewers did an great job in their reviews. I will highlight issues I think are particularly salient here. In your resubmission, please include a document with a point-by-point response to both the points I list here and the reviewers’ comments, outlining each change made in your manuscript or providing a suitable rebuttal.

The topic of your research is clearly interesting and important, but I think you could do more in the introduction to explain how it related to theory and what your research questions are. I assume since the study is not preregistered that you are not testing hypotheses? In any case, more discussion of how the research relates to theory is needed, along with more explicit research questions.

In addition, more work is needed to justify your choices related to the methods. Why do some of the aggressive actions involve ingroup and outgroup members while others involve aggression within the group? I also wondered what it means that participants were still included who failed the manipulation check. Were niceness and the friendship question also manipulation checks? What role do these questions play in regard to your research questions?

I also agree with the reviewer who suggested that you include more explanation of how Bayes factors are calculated and how they can be interpreted for readers who are less familiar. Relatedly, it would be helpful to know what your priors were and where they came from. If you expected the leader to be judged differently than the others, isn’t chance 50/50?

In summary, I think this is a promising manuscript and, I hope you will revise it for further consideration at Collabra: Psychology. I look forward to receiving your revision.

Please ensure that your revised files adhere to our author guidelines, and that the files are fully copyedited/proofed prior to upload. Please also ensure that all copyright permissions have been obtained. This is the last opportunity for major editing, therefore please fully check your file prior to re-submission.

If you have any questions or difficulties during this process, please contact the editorial office at editorialoffice@collabra.org.

We hope you can submit your revision within the next six weeks. If you cannot make this deadline, please let us know as early as possible.

Sincerely,

Clare Conry-Murray

**Reviewer 1**

**Open response questions**

Please write your review here. The author(s) will see this review. Your identity will not be revealed to the authors unless you also include your name (i.e., sign your review) in this box. It is up to you whether to reveal your identity or not, either is fine.

The current work provides a systematic study of children’s expectations regarding leaders’ behaviors in novel social groups. The studies add to a growing field examining children’s intutitions about social power dynamics in societies. Though I enjoyed learning about the findings of this research, I also had several concerns, which I outline below.

• In the introduction, when setting up the distinction between dominance and prestige, it is important that the authors refer to recent work conducted by Enright and colleagues (<https://www.tandfonline.com/doi/abs/10.1080/15248372.2020.1797745>).
• Are dominance and prestige easily separable in real-life situations, as the authors seem to suggest they are in the introduction? I can think of few powerful people who don’t utilize both strategies in different contexts. Are the authors equating dominance with malevolence and prestige with benevolence? It would help to clarify these definitions early in the paper so that the authors’ claims about different (but related) constructs are clear.
• How is the power information conveyed in the protocols? Is the crown the only clue? If so, then are these studies merely measuring children’s tendency to associate these behaviors with kings, rather than “leaders” in general? The authors bring up this possibility briefly in the discussion, but perhaps more elaboration is needed.
• I really liked the contrast set up between non-human primates and children in the discussion on the use of aggression by dominant individuals. However, related to the previous point I made, I am not convinced that we can conclude from these data that kids think dominant individuals are not necessarily aggressive. Perhaps kids understand that dominant individuals, like kings and other leaders in our society are able to employ others to commit aggressive acts on their behalf. Similarly, kids might understand that leaders can employ others’ use of force to enforce norms (e.g., judges, police forces, etc). So, I wonder to what extent these findings can be generalized to children’s understanding of power dynamics in our society.
• Was there any overlap in participants of Study 1, 2 and 3? If not, I find that it might be difficult to make some of the interpretations that the authors are making. For example, on p. 21, the authors state: “These responses differed from children’s responses to the stories in Studies 1 and 2 where the leader ousted an intruder who had already transgressed. In those Studies, children overwhelmingly chose the leader as the individual who ejected an outsider. However, it seems that this was not because children necessarily expect leaders to protect subordinates, but because they expect them to expel anti-social intruders.” It seems important that authors provide such interpretations using more speculative or tentative language, given that the different findings are found in different samples, for which different sampling biases might have been at play.

Other misc. issues:
• There are typos and missing punctuations throughout the paper.
• On p. 4 there’s this example: “Of course, dominance and prestige cannot always be clearly distinguished. Take, for example, the relationship between an academic advisor and advisee. While the advisee may perceive their advisor as someone who provides benefits through knowledge and experience, they also can inflict material harm, for example by taking away resources such as income.” This doesn’t seem to be the best example as it oversimplifies how a graduate student is hired in academic institutions. Perhaps a more direct and intutive example can be thought of.
• This might just be a personal preference so the authors can certainly choose to ignore it, but it’s nice when demographics like race/ethnicity are in a table so they can be quickly glanced at.

**Rating scale questions**

|  | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- | --- |
| The study/studies in this manuscript have strong construct validity (good measures and/or manipulations of the constructs the authors wish to study). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  | ✔ |  |
| The study/studies in this manuscript have strong statistical validity (appropriate statistical tests, assumptions are clear and reasonable, no statistical errors, appropriate statistical inferences, etc.). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  |  | ✔ |
| The study/studies in this manuscript have strong internal validity (any causal claims or implications are well-justified, alternative explanations are thoroughly considered, etc.). (Choose “Neutral” if this is not an empirical manuscript, or no causal claims are made or even vaguely implied.) |  |  |  | ✔ |  |
| The study/studies in this manuscript have strong external validity (authors appropriately constrain their conclusions based on the limits of the generalizability of their findings to other contexts (including from lab to real world), other populations, other stimuli or measures, etc.) |  |  | ✔ |  |  |

**Reviewer 2**

**Open response questions**

Please write your review here. The author(s) will see this review. Your identity will not be revealed to the authors unless you also include your name (i.e., sign your review) in this box. It is up to you whether to reveal your identity or not, either is fine.

Using hypothetical vignettes, this manuscript examined the behaviors young children expect of “leaders” across three related studies. They found that children expected leaders to be neither antisocial (randomly aggressive) nor prosocial towards subordinates, but to provide certain benefits to the group (expelling antisocial intruders). Overall, I thought this series of studies were well designed and used appropriate statistical procedures to test concrete and novel questions. The authors did a particularly nice job anticipating questions/potential flaws in the methodology (such as the leader sitting down versus standing), summarizing the overall findings, and providing relevant future directions for readers.

My comments are relatively minor, and mostly pertain to the presentation of the results. I detail them below.

1. Please specify the general geographic location where the sample was collected (e.g., a large metropolitan city in the Northeast United States). The authors note that their research question was about children in “North America,” but this is still a quite broad area, including multiple countries and states with somewhat different political leadership styles. Also, I recognize that it is unlikely that the authors were able to collect socioeconomic information about the children’s families, but if possible, it would be helpful to provide a general snapshot of the SES range of families that typically attend the children’s museum where the sample was recruited. Either way, the role of SES (in addition to race/culture) may be an important feature to discuss in the future directions section as lower SES families often have a more hierarchical/authoritarian structure, which could impact expectations about leadership.
2. The use of Bayes Factors is a strength of the study, as it is a particularly appropriate way of testing the authors’ (proportion-based) hypotheses. However, this procedure and general rules of thumb for the values’ meaning/effect size/significance are likely less well known to readers (including myself) than frequentist methods. The authors often reference terms like “weak evidence,” “positive evidence,” or “strong evidence” but it is unclear where these terms come from. It would therefore be helpful if the authors would provide some general interpretative guidelines for what different size bayes factors typically mean in terms of their strength of supporting the null versus the alternative hypothesis (this could be done as part of the analytic approach section). This would particularly help when trying to understand how BF values can represent “inconclusive” evidence (neither in favor of the null or the alternative hypothesis).
3. Relatedly, the authors are quite inconsistent about providing an accompanying p value along with the BF values they list. Please be sure to provide such a value each time.
4. It was very helpful when the authors stated what the null hypothesis and alternative hypothesis represent for the main ProportionBF tests in the Analysis Approach section. Please also provide a similar statement of what the “null” and “alt” hypothesis represent for the generalTestBF tests of demographic differences. This is especially important because the authors frequently report a BF factor “in favor of the null (or altertative) hypothesis”.
5. Why did the authors choose to dichotomize age rather than examine it as a continuous predictor? Typically (in frequentist methods) categorizing continuous predictors reduces power and increases Type I error. Was age left as a continuous predictor for the generalTestBF (comparing models with and without the age predictor) and merely dichotomized later to examine effects in different age groups? If so, it would be somewhat less concerning, because dichotomizing would merely be used as a tool for probing the effect, but this use should still be specified in the analysis plan and the choice to dichotomize age at the split between 5 and 6 year olds justified.
6. I think that the BF factor listed at the top of p. 12 for 6-8 year olds on story 1b (BF=43.74) may be incorrect. The proportion in favor of the leader is higher for this group than it is for the 3-5 year olds, but the BF factor is much smaller (43.74 vs >1000). Also Figure 2 depicts this proportion as having three stars (BF > 1000).
7. There are no results of a gender difference test for the “Who is the nicest?” and “Who would you rather be friends with?” questions in Study 1. Please provide these results.
8. I don’t fully understand how stories 3b and 3f serve as “understanding checks” for stories 3a and 3e. While these stories are useful questions to help clarify children’s thinking/reasons for their answers to 3a and 3e, it doesn’t seem like the answer to 3b and 3f really tell you whether children understood the questions in 3a and 3e. I would recommend reframing the discussion of these stories as “checking for” or “clarifying” potential reasoning behind children’s answers to 3a and 3e.
9. In the first paragraph of the results and discussion section for Study 3 (p. 21), which is presumably for Story 3a (although this is not made explicit with a header) the authors reference Figure 6. Figure 6 appears to be the picture aligned with Story 3b rather than Story 3a based on its caption. I found this somewhat confusing. I would recommend that the authors make figure 6 the image/caption from Story 3a instead.
10. The “Gülgoz and Gelman” citation on page 25 is missing its parenthetical date.
11. In order to help clarify how the Gülgoz and Gelman findings support the age difference in the current study, please clarify what was meant by “malevolent power” and “benevolent power” in that study.
12. I generally liked the way the authors framed the limitations of the study in terms of future directions rather than limitations, but it would be good to explicitly state that the findings may not generalize to youth outside of the cultural and demographic features of the current sample (as part of the note that expectations about leaders vary across cultures).

**Rating scale questions**

|  | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- | --- |
| The study/studies in this manuscript have strong construct validity (good measures and/or manipulations of the constructs the authors wish to study). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  |  | ✔ |
| The study/studies in this manuscript have strong statistical validity (appropriate statistical tests, assumptions are clear and reasonable, no statistical errors, appropriate statistical inferences, etc.). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  |  | ✔ |
| The study/studies in this manuscript have strong internal validity (any causal claims or implications are well-justified, alternative explanations are thoroughly considered, etc.). (Choose “Neutral” if this is not an empirical manuscript, or no causal claims are made or even vaguely implied.) |  |  |  |  | ✔ |
| The study/studies in this manuscript have strong external validity (authors appropriately constrain their conclusions based on the limits of the generalizability of their findings to other contexts (including from lab to real world), other populations, other stimuli or measures, etc.) |  |  |  | ✔ |  |

**Author Response**
Nov 3, 2021

Dear Clare Conry-Murray,

Thank you for inviting us to resubmit the Manuscript originally titled, “Children expect leaders to oust intruders, to refrain from aggression, but do not expect leaders to be generally more prosocial.” (Now titled, “Children expect leaders to oust intruders, refrain from unprovoked aggression, but not to be generally prosocial”). We have made substantial changes to the manuscript and responded to the reviews below.

Thank you very much!

Ashley J. Thomas

I have now received all reviews of your manuscript, "Children expect leaders to oust intruders, to refrain from aggression, but do not expect leaders to be generally more prosocial" from two qualified researchers. I also independently read the manuscript before consulting these reviews. I agree that your manuscript has important strengths and also that there are some issues that need to be addressed. I therefore encourage you to submit a revised version for further consideration at Collabra: Psychology.

The reviewers did an great job in their reviews. I will highlight issues I think are particularly salient here. In your resubmission, please include a document with a point-by-point response to both the points I list here and the reviewers' comments, outlining each change made in your manuscript or providing a suitable rebuttal.

The topic of your research is clearly interesting and important, but I think you could do more in the introduction to explain how it related to theory and what your research questions are. I assume since the study is not preregistered that you are not testing hypotheses?

These experiments were not pre-registered because our lab did not practice pre-registration at the time that these experiments were conducted (starting in 2016).

In any case, more discussion of how the research relates to theory is needed, along with more explicit research questions.

Thanks for this suggestion. we have updated the following paragraph to be more specific about how our paper is theoretically important and what specific questions they answer.

In the current study, we ask whether children (visitors to a science museum in Southern California), have expectations about how high-ranking individuals (hereafter, ‘leaders’[[1]](#footnote-1)) will treat other individuals. While much prior work in developmental psychology has focused on how children understand others in terms of their traits [(Heyman & Gelman, 1998; Kiley Hamlin & Steckler, 2015)](https://paperpile.com/c/CYhU7Z/8gjb%2BYL3X) or group membership [(Dunham, 2018; Gelman & Taylor, n.d.)](https://paperpile.com/c/CYhU7Z/hlbJ%2BUPKy), less work has focused on children’s understanding of relationships [(Afshordi & Liberman, 2020; Olson & Spelke, 2008)](https://paperpile.com/c/CYhU7Z/ZGqM%2BGEzR), including whether they use existing relationships to predict behavior. Thus, the first question the current studies set out to test is whether children use an existing hierarchical relationship to predict behavior. Second, these studies investigate whether children expect leaders to act in ways that agree more with dominance or prestige-based hierarchies. That is, without information about whether a high-ranking person is high-ranking because of dominance or prestige, would children nonetheless expect them to act more in line with one of these models? Thus, the actions we use in the current study were chosen with the distinction between dominance and prestige in mind, however as noted above, these types of social rank cannot always be cleanly disambiguated. In line with prestige-based hierarchies we specifically ask, do children expect leaders to provide certain benefits? If so, what types of benefits? In line with dominance-based hierarchies, we ask, do they expect leaders to be responsible for unprovoked aggression? Thus these studies can shed light on whether children are biased to see leaders, at least in this context, in line with one of these two models.

We also added the following to the introduction to Study 3:

In Study 3, we investigated several follow-up questions arising from Studies 1 and 2. For example, to what extent do children expect leaders to be protective? Do children expect leaders to put themselves in harm’s way to protect others? Do children see leaders as generally more prosocial than non-leaders? We included conditions to probe this question because if children did expect leaders to be generally prosocial, then children’s expectations of leaders may not stem from an understanding of specific roles based on relative status (i.e., the leader’s job is to protect others), but instead from an inference about traits of leaders (i.e., ‘leaders are nice people’).

In addition, more work is needed to justify your choices related to the methods. Why do some of the aggressive actions involve ingroup and outgroup members while others involve aggression within the group?

Thank you, we added the following to explain these design choices:

We used in-group aggression because in dominance-based hierarchies, dominant individuals often use random aggression toward in-group members to maintain rank.

Here we used out-group aggression as a relatively straightforward case of protection.

I also wondered what it means that participants were still included who failed the manipulation check.

Because we did not decide to exclude participants before data collection, we thought it was more straightforward and involved fewer experimenter degrees of freedom, if we included as much data as possible in the final sample.

Were niceness and the friendship question also manipulation checks? What role do these questions play in regard to your research questions?

We now include a longer explanation as to why we included those questions (page 11):

After the test trials came several control questions. We showed children an illustration of a new, novel group with four characters and asked (1) *Who is in charge?* (2) *Who do you think is the nicest?* and (3) *Who would you most want to be friends with?* The first question served as a manipulation check, to confirm that children saw the crowned character as a leader. The second and third questions checked whether children had an overall bias for or against choosing the leader.

I also agree with the reviewer who suggested that you include more explanation of how Bayes factors are calculated and how they can be interpreted for readers who are less familiar. Relatedly, it would be helpful to know what your priors were and where they came from. If you expected the leader to be judged differently than the others, isn't chance 50/50?

The priors were default priors, we include that in the manuscript now (page 11). Chance isn’t 50/50 because if they were randomly choosing a character, they should have chosen the leader 1/3 of the time in cases where there were 3 characters and ¼ of the time when there were four characters. Moreover, this analysis was chosen before analysis was carried out.

In summary, I think this is a promising manuscript and, I hope you will revise it for further consideration at Collabra: Psychology. I look forward to receiving your revision.

Please ensure that your revised files adhere to our author guidelines, and that the files are fully copyedited/proofed prior to upload. Please also ensure that all copyright permissions have been obtained. This is the last opportunity for major editing, therefore please fully check your file prior to re-submission.

If you have any questions or difficulties during this process, please contact the editorial office at editorialoffice@collabra.org.

We hope you can submit your revision within the next six weeks. If you cannot make this deadline, please let us know as early as possible.

Sincerely,

Clare Conry-Murray

**Reviewer #1**

The current work provides a systematic study of children’s expectations regarding leaders’ behaviors in novel social groups. The studies add to a growing field examining children’s intutitions about social power dynamics in societies.Though I enjoyed learning about the findings of this research, I also had several concerns, which I outline below.

• In the introduction, when setting up the distinction between dominance and prestige, it is important that the authors refer to recent work conducted by Enright and colleagues (<https://www.tandfonline.com/doi/abs/10.1080/15248372.2020.1797745>).

Thank you for the suggestion! We added this.

However, children do not seem to distinguish dominance from prestige in their preferences and resource allocations (Enright et al, 2020)

• Are dominance and prestige easily separable in real-life situations, as the authors seem to suggest they are in the introduction? I can think of few powerful people who don’t utilize both strategies in different contexts. Are the authors equating dominance with malevolence and prestige with benevolence? It would help to clarify these definitions early in the paper so that the authors’ claims about different (but related) constructs are clear.

No, as we stated in our original submission, we don’t think dominance and prestige are always easily separable, we have added more to the paper to expand.

Of course, dominance and prestige cannot always be clearly distinguished. Take, for example, the relationship between an employer and an employee. While the employee may perceive their boss as someone who provides benefits in the form of guidance and resources, the employer can inflict harm on the employee by taking away income or opportunities. Because each person’s rank is defined in relation to other people, the ultimate determinants of ‘dominance’ and ‘prestige’ are the perceptions of the lower-ranking people in the relationship and their expectations about how the higher-ranking people will treat them.

The actions of dominant and prestigious individuals’ also influence perceptions of them: A dominant individual who is especially aggressive may instill fear in subordinates, whereas a prestigious individual who is notably non-aggressive may instill trust in subordinates [(Fiske, 1992; Fiske & Rai, 2014)](https://paperpile.com/c/CYhU7Z/yin8m%2BWcGX).

Thus, the actions we use in the current study were chosen with the distinction between dominance and prestige in mind, however as noted above, these types of social rank cannot always be cleanly disambiguated.

• How is the power information conveyed in the protocols? Is the crown the only clue? If so, then are these studies merely measuring children’s tendency to associate these behaviors with kings, rather than “leaders” in general? The authors bring up this possibility briefly in the discussion, but perhaps more elaboration is needed.

This may be true, but we asked children across all studies, ‘who is in charge?/Who is the boss’. This measure has been used in several other studies to measure children’s knowledge of status difference.

We added some more explanations and qualifiers to statements summing up the results:

We illustrated the leader this way because sitting higher than others, or wearing something to make oneself appear larger or brighter is a common cue of rank across cultures (e.g., [(Fiske, 1992)](https://paperpile.com/c/CYhU7Z/WcGX)).During pilot-testing, children consistently identified the crowned, throned character as the “leader.”.

Children in this study did have certain expectations about the behavior of leaders, at least in this context in which rank was cued using a crown and throne

These data suggest that children expect leaders, at least in this context, to provide specific benefits such as expelling hostile intruders, but not to be more prosocial than other people overall.

And include the original paragraph that points this out:

Future research could explore several questions raised by the limitations of these studies. The first is whether the expectations we found in this study would also be found if social rank was presented in other ways. Here, rank was manipulated by having the leader look like a prototypical king or queen: The character wore a crown, and in Studies 1 and 3 they sat on a throne. We chose these cues because children recognize them and easily identify the crowned, throned character as being ‘in charge.’ These children’s intuitions about royalty undoubtedly are influenced by fictional depictions, such as those in movies and books. Future research could test whether children have similar intuitions about high-status people whose rank is not associated with royalty.

• I really liked the contrast set up between non-human primates and children in the discussion on the use of aggression by dominant individuals. However, related to the previous point I made, I am not convinced that we can conclude from these data that kids think dominant individuals are not necessarily aggressive. Perhaps kids understand that dominant individuals, like kings and other leaders in our society are able to employ others to commit aggressive acts on their behalf. Similarly, kids might understand that leaders can employ others’ use of force to enforce norms (e.g., judges, police forces, etc). So, I wonder to what extent these findings can be generalized to children’s understanding of power dynamics in our society.

We thank the reviewer for this comment because it helped us clarify what we meant. In our original draft we did not intend to make the argument that children do not expect leaders to be aggressive, we meant to make the argument that they do not expect them to commit *random* or unprovoked acts of aggression.

We made the following changes:

**Title:** Children expect leaders to oust intruders, refrain from unprovoked aggression, but not to be generally prosocial

**Research Highlights:**

Children have expectations about how high-ranking individuals will treat low ranking individuals.

Younger children do not expect high-ranking individuals to commit unprovoked acts of aggression.

Older children expect high-ranking individuals to be less likely to commit unprovoked acts of aggression.

Children of all ages expect leaders to provide at least one benefit: protection from hostile outsiders

Dominant individuals in both human hierarchies and hierarchies in other species often maintain their rank through threat of violence. For example, in some non-human primate species, high-ranking individuals act like school-yard bullies, committing unprovoked acts of aggression which helps them maintain their social rank [(Silk, 2002)](https://paperpile.com/c/CYhU7Z/tCXia).

In line with dominance-based hierarchies, we ask, do they expect leaders to be responsible for unprovoked aggression? Thus these studies can shed light on whether children are biased to see leaders, at least in this context, in line with one of these two models.

It is not yet known whether children have general expectations about how high-ranking individuals will treat subordinates, including whether they expect them to be more or less likely to commit unprovoked acts of aggression

We start by asking the following questions: Do children expect leaders to commit unprovoked acts of aggression?

We used in-group aggression because in dominance-based hierarchies, dominant individuals often use unprovoked aggression toward in-group members to maintain rank.

These findings are interesting considering how dominance hierarchies often work in non-human primates. In many primate species, dominant individuals commit unprovoked aggressive acts on subordinates, which is thought to help the dominant individual maintain rank without engaging in more costly fights [(Silk, 2002)](https://paperpile.com/c/CYhU7Z/tCXia).

For the children in this study, it seems that being ‘in charge,’ did not mean being morelikelyto commit unprovoked aggression.

In our study, only older children thought that leaders were less likely to commit unprovoked acts of aggression.

We take these results to mean that although it might be easier for younger children to identify malevolent leaders as being in charge, they do not necessarily expect leaders to be randomly aggressive.

• Was there any overlap in participants of Study 1, 2 and 3? If not, I find that it might be difficult to make some of the interpretations that the authors are making. For example, on p. 21, the authors state: “These responses differed from children’s responses to the stories in Studies 1 and 2 where the leader ousted an intruder who had already transgressed. In those Studies, children overwhelmingly chose the leader as the individual who ejected an outsider. However, it seems that this was not because children necessarily expect leaders to protect subordinates, but because they expect them to expel anti-social intruders.” It seems important that authors provide such interpretations using more speculative or tentative language, given that the different findings are found in different samples, for which different sampling biases might have been at play.

We have added the following:

However, it should be noted that while these children were recruited from the same museum using the same methods, they were not the same children so these comparisons should be interpreted with caution.

Other misc. issues:
• There are typos and missing punctuations throughout the paper.

We have worked to address this.

• On p. 4 there’s this example: “Of course, dominance and prestige cannot always be clearly distinguished. Take, for example, the relationship between an academic advisor and advisee. While the advisee may perceive their advisor as someone who provides benefits through knowledge and experience, they also can inflict material harm, for example by taking away resources such as income.” This doesn’t seem to be the best example as it oversimplifies how a graduate student is hired in academic institutions. Perhaps a more direct and intutive example can be thought of.

Take, for example, the relationship between an employer and an employee. While the employee may perceive their boss as someone who provides benefits in the form of guidance and resources, the employer can inflict harm on the employee by taking away income or opportunities

• This might just be a personal preference so the authors can certainly choose to ignore it, but it’s nice when demographics like race/ethnicity are in a table so they can be quickly glanced at.

Thanks for this advice, we now include tables

**Reviewer #2**

Using hypothetical vignettes, this manuscript examined the behaviors young children expect of “leaders” across three related studies. They found that children expected leaders to be neither antisocial (randomly aggressive) nor prosocial towards subordinates, but to provide certain benefits to the group (expelling antisocial intruders). Overall, I thought this series of studies were well designed and used appropriate statistical procedures to test concrete and novel questions. The authors did a particularly nice job anticipating questions/potential flaws in the methodology (such as the leader sitting down versus standing), summarizing the overall findings, and providing relevant future directions for readers.

Thank you!

My comments are relatively minor, and mostly pertain to the presentation of the results. I detail them below.

Please specify the general geographic location where the sample was collected (e.g., a large metropolitan city in the Northeast United States). The authors note that their research question was about children in “North America,” but this is still a quite broad area, including multiple countries and states with somewhat different political leadership styles. Also, I recognize that it is unlikely that the authors were able to collect socioeconomic information about the children’s families, but if possible, it would be helpful to provide a general snapshot of the SES range of families that typically attend the children’s museum where the sample was recruited. Either way, the role of SES (in addition to race/culture) may be an important feature to discuss in the future directions section as lower SES families often have a more hierarchical/authoritarian structure, which could impact expectations about leadership.

Thank you for pointing this out, we added a supplementary materials section that summarizes more of the demographic information about the children we tested as well as the following:

In the current study, we ask whether children (visitors to a science museum in Southern California), have expectations about how high-ranking individuals (hereafter, ‘leaders’[[2]](#footnote-2)) will treat other individuals.

In Study 1, we tested 195 children between the ages of 3 and 9 years old who were visitors to a children’s museum in Orange County, Southern California.

The use of Bayes Factors is a strength of the study, as it is a particularly appropriate way of testing the authors’ (proportion-based) hypotheses. However, this procedure and general rules of thumb for the values’ meaning/effect size/significance are likely less well known to readers (including myself) than frequentist methods. The authors often reference terms like “weak evidence,” “positive evidence,” or “strong evidence” but it is unclear where these terms come from. It would therefore be helpful if the authors would provide some general interpretative guidelines for what different size bayes factors typically mean in terms of their strength of supporting the null versus the alternative hypothesis (this could be done as part of the analytic approach section). This would particularly help when trying to understand how BF values can represent “inconclusive” evidence (neither in favor of the null or the alternative hypothesis).

We have added the following:

Following recommendations by [(Kass & Raftery, 1995)](https://paperpile.com/c/CYhU7Z/7e2K), we consider Bayes Factors of over 10 as strong evidence, over 30 as very strong evidence, and over 100 as decisive evidence., Bayes Factors between 1/3 and 3 are considered weak, inconclusive or anecdotal.

Relatedly, the authors are quite inconsistent about providing an accompanying p value along with the BF values they list. Please be sure to provide such a value each time.

We have updated this, we originally did not include p-values when we found inconclusive or weak evidence, or evidence for the null since p-values cannot provide evidence for the null

It was very helpful when the authors stated what the null hypothesis and alternative hypothesis represent for the main ProportionBF tests in the Analysis Approach section. Please also provide a similar statement of what the “null” and “alt” hypothesis represent for the generalTestBF tests of demographic differences. This is especially important because the authors frequently report a BF factor “in favor of the null (or altertative) hypothesis”.

This has been updated throughout the manuscript.

Why did the authors choose to dichotomize age rather than examine it as a continuous predictor? Typically (in frequentist methods) categorizing continuous predictors reduces power and increases Type I error. Was age left as a continuous predictor for the generalTestBF (comparing models with and without the age predictor) and merely dichotomized later to examine effects in different age groups? If so, it would be somewhat less concerning, because dichotomizing would merely be used as a tool for probing the effect, but this use should still be specified in the analysis plan and the choice to dichotomize age at the split between 5 and 6 year olds justified.

Based on pilot data we decided to split children up into two age groups. However, we include this analysis in the SM now.

I think that the BF factor listed at the top of p. 12 for 6-8 year olds on story 1b (BF=43.74) may be incorrect. The proportion in favor of the leader is higher for this group than it is for the 3-5 year olds, but the BF factor is much smaller (43.74 vs >1000). Also Figure 2 depicts this proportion as having three stars (BF > 1000).

Thank you! This has been updated.

There are no results of a gender difference test for the “Who is the nicest?” and “Who would you rather be friends with?” questions in Study 1. Please provide these results.

Thank you, these have been added

I don’t fully understand how stories 3b and 3f serve as “understanding checks” for stories 3a and 3e. While these stories are useful questions to help clarify children’s thinking/reasons for their answers to 3a and 3e, it doesn’t seem like the answer to 3b and 3f really tell you whether children understood the questions in 3a and 3e. I would recommend reframing the discussion of these stories as “checking for” or “clarifying” potential reasoning behind children’s answers to 3a and 3e.

We thank the reviewer for expressing their confusion. We originally hypothesized that children would choose the leader in the first conditions but not the latter. Thus, two similar stories could have shown that children answered in the opposite way. But this is not what we found. We included this hypothesis in the manuscript now.

The next story, Story 3b, was used to probe children’s reasoning for their answers in Story 3a. In Story 3b, children were told that a character *wanted* to help someone who was being attacked but was too scared to do so (*Someone wants to help the Blip, but they’re too scared. Which one do you think was too scared to help?*). Here we predicted that children would be less likely to choose the leader than a non-leader, since leaders are generally seen as stronger or braver than others.

Story 3f was designed to probe the reasons behind children’s answers in Story 3e, here someone *wanted* to enforce the norm but did not (*Look! One of the Rookas is wearing the wrong shirt! One of the Rookas didn’t think it was ok but didn’t say anything. Which one do you think didn’t say anything?*). Again, we predicted that children would choose leaders more often in Story 3f and choose non-leaders more often in Story 3d. However, the results of these last two conditions were inconclusive.

In the first paragraph of the results and discussion section for Study 3 (p. 21), which is presumably for Story 3a (although this is not made explicit with a header) the authors reference Figure 6. Figure 6 appears to be the picture aligned with Story 3b rather than Story 3a based on its caption. I found this somewhat confusing. I would recommend that the authors make figure 6 the image/caption from Story 3a instead.

Thank you, this is fixed now.

The “Gülgoz and Gelman” citation on page 25 is missing its parenthetical date.

Thank you we have fixed this now

In order to help clarify how the Gülgoz and Gelman findings support the age difference in the current study, please clarify what was meant by “malevolent power” and “benevolent power” in that study.

We changed the following passage to make this more clear:

The difference we found between younger and older children’s expectations aligns with developmental changes found in other studies. Gülgoz and Gelman (2016) found that younger children did not acknowledge benevolent power (interactions that benefitted the lower-ranking individual) as a cue of who was ‘in charge’, while older children did. In our study, only older children thought that leaders were less likely to commit unprovoked acts of aggression. Thus there may be a general trend in which older children see leaders as more benevolent.

I generally liked the way the authors framed the limitations of the study in terms of future directions rather than limitations, but it would be good to explicitly state that the findings may not generalize to youth outside of the cultural and demographic features of the current sample (as part of the note that expectations about leaders vary across cultures).

Thanks! We added the following to make it clearer that this work is limited. We also added more demographic information to the SM about the participants.

Future research could also ask how children explain the actions of leaders and non-leaders. For example, children might think that ifa leader is aggressive, the aggression must provide a benefit to the group. If so, this would provide evidence for sophisticated reasoning about the roles of leaders in social groups. Another important direction for future research will be to learn how children’s own experiences with social hierarchy affect how they expect leaders to act. There is a hierarchical element to children’s relationships with peers, siblings, teachers, and parents, but in each of those relationships higher-ranking individuals likely act in different ways. Moreover, since expectations about leaders vary across cultural contexts, one major limitation of this work is that the population we tested was specific to families visiting a museum in one geographic area. These findings may not generalize across cultures or populations. Future work could investigate how children in different cultural contexts, considered both on large and small scales, expect leaders to act. In as much as there are culturally variable views on leadership, we might expect children’s views on leadership to become more culturally divergent with age, as children gain more experience with their specific culture.

**Editor First Decision**: Revise & Resubmit

Jan 10, 2022

Dear Ashley Thomas,

I have now read your revised manuscript. I appreciate your careful attention to the concerns the reviewers and I raised. I am happy to provisionally accept your manuscript for submission. However, I found a few things I would like you to address.

Required revisions:

1. Please examine whether your conclusions would change if you excluded participants who did not identify the person with the crown as the leader before data analysis. You may add this discussion in supplementary online materials, and just refer to it briefly in the manuscript.
2. Please add to the section on limitations to address the suggestion from Reviewer 2 that Study 3 had two elements that varied from the other studies, and future studies may want to examine each element separately.

I look forward to receiving your final revision and accepting it for publication in Collabra: Psychology.

Please ensure that your revised files adhere to our author guidelines, and that the files are fully copyedited/proofed prior to upload. Please also ensure that all copyright permissions have been obtained. This is the last opportunity for major editing, therefore please fully check your file prior to re-submission.

If you have any questions or difficulties during this process, please contact the editorial office at editorialoffice@collabra.org.

We hope you can submit your revision within the next six weeks. If you cannot make this deadline, please let us know as early as possible.

Sincerely,

Clare Conry-Murray

**Reviewer 1**

**Open response questions**

Please write your review here. The author(s) will see this review. Your identity will not be revealed to the authors unless you also include your name (i.e., sign your review) in this box. It is up to you whether to reveal your identity or not, either is fine.

This is an excellent revision. The authors have done a wonderful job addressing my suggestions/concerns as well as others’. Thank you.

Selin Gülgöz

**Rating scale questions**

|  | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- | --- |
| The study/studies in this manuscript have strong construct validity (good measures and/or manipulations of the constructs the authors wish to study). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  |  | ✔ |
| The study/studies in this manuscript have strong statistical validity (appropriate statistical tests, assumptions are clear and reasonable, no statistical errors, appropriate statistical inferences, etc.). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  |  | ✔ |
| The study/studies in this manuscript have strong internal validity (any causal claims or implications are well-justified, alternative explanations are thoroughly considered, etc.). (Choose “Neutral” if this is not an empirical manuscript, or no causal claims are made or even vaguely implied.) |  |  |  |  | ✔ |
| The study/studies in this manuscript have strong external validity (authors appropriately constrain their conclusions based on the limits of the generalizability of their findings to other contexts (including from lab to real world), other populations, other stimuli or measures, etc.) |  |  |  |  | ✔ |

**Reviewer 2**

**Open response questions**

Please write your review here. The author(s) will see this review. Your identity will not be revealed to the authors unless you also include your name (i.e., sign your review) in this box. It is up to you whether to reveal your identity or not, either is fine.

I think the authors did a good job of responding to my comments and those of the other reviewer. Many elements of the study (especially the description of the Bayes results) are now clearer.
With this in mind, I did notice two other issues that didn’t catch my attention the first time around but became more obvious after reading the editor’s comments and on this new read through – both of these are significant.

1. Like the editor, I am concerned that the authors included children who failed the manipulation check. I appreciate the author’s argument that they had not decided, a priori, to exclude children who failed the manipulation check and so included all children in the analyses to avoid excess “researcher degrees of freedom”. However, I think that the inclusion of children who did not view the crowned character as the “leader” undermines the findings – those children picking or not picking the leader on all the rest of the questions has a very different meaning regarding their expectations of a leader. This becomes particularly relevant any time that an age group difference was found, as the age groups differed in their likelihood of correctly identifying the “leader.” I think the paper would be methodologically stronger and more trustworthy were the authors to present findings with those children removed. At minimum, I would like to see evidence (perhaps in the SM) that the authors verified that findings remained essentially the same when analyses were run on only those children who passed the manipulation check.
2. I believe the authors need to devote more and clearer attention to the fact that TWO key elements were changed in story 3a (compared to stories 1b/d and 2b/d). Specifically, it is not just that the character intervened BEFORE the intruder transgressed (which is the element emphasized in the description of the story on page 22), but also that they HELPED THE VICTIM rather than EXPELLED/PUNISHED THE INTRUDER. This is a really key difference and should be emphasized in the description of story 3a. It does not necessarily change the interpretation of the question – indeed the authors seem to have focused on this second element as the key reason for different responses to questions 1/2b&d and 3a (“it seems that this was not because children necessarily expect leaders to protect others, but because they expect them to expel anti-social intruders”). However, it is a bit unclear whether differences in child answers for story 3a (compared to 1/2b&d) were due to the preemptive nature of the response (before the transgression took place) or the focus of the response (helping ingroup member vs punishing/expelling outgroup member). I think this complication is at least as relevant as the fact that it is a different sample in terms of impacting the authors’ ability to make direct comparisons/conclusions across the studies.

This confound also impacts the authors’ ability to accurately draw the conclusion on page 24: “Together with Story 3a, these results suggest that children think of leaders as being braver than other characters but not necessarily more likely to intervene before a transgression has taken place.” It may be that children do not view leaders as more likely to help others out/be prosocial rather than that they are not more likely to intervene before (vs after) a transgression. Indeed, as stories 3c and 3d showed that children viewed leaders as less likely to be prosocial than others, the results of story 3a (leaders not more or less likely to intervene) likely involve a combination of the protection/punishment element from stories 1/2b&d (leaders being MORE likely to defend ingroup members from outgroup members by expelling problematic outgroup members) and the prosocial element from stories 3c/d (leaders being LESS likely to just randomly help out ingroup members when no outgroup threat is involved). I think that this complication really needs explicit attention.

Again, the authors DO attend to this difference and draw this conclusion in some regards (e.g., page 27: “One possibility is that the intuition driving children’s choices in Studies 1 and 2 (where children identified the leader as the one who probably ousted an antisocial intruder) was not the intuition that leaders protect their subordinates but rather that leaders protect their group, or that they punish anti-social individuals”) but I think that the multiple changes to story 3a (that result in an unclear ability to determine which change in the story resulted in the changed response) need to be more explicitly addressed both in the description of story 3a and in the limitations section.

My other notes are quite minor:
Introduction:

1. Can the authors elaborate a bit more on the statement “Interestingly, older children do not seem to distinguish dominance from prestige in their preferences and resource allocations (Enright et al., 2020)” on page six? What age older children don’t make these distinctions, and do they make any of the distinctions previously described that younger children draw?
2. I like the authors’ decision to specify the questions they asked on page 4 in order to better link these questions to the distinction between dominance and prestige based hierarchies (“In line with prestige-based hierarchies we specifically ask, do children expect leaders to provide certain benefits? If so, what types of benefits? In line with dominance-based hierarchies, we ask, do they expect leaders to be responsible for unprovoked aggression?”). However, doing so makes your discussion of the questions you asked on page 7 now feel a bit redundant (“We start by asking the following questions: Do children expect leaders to commit unprovoked acts of aggression? Do children expect leaders to provide protection?”). I would encourage the authors to find a way to reference those earlier questions or merely note that they were previously specified when discussing them again on page 7. Doing so would help avoid confusion as to whether the page 7 questions are a different set of questions or just a reiteration of previous information.

Analysis Approach:

1. Can the authors provide just a bit more description as to what results from the pilot study encouraged them to dichotomize age?
2. I very much appreciate the authors providing the alternative version of each analysis treating age as a continuous variable in the supplementary material. Unfortunately, I was not able to view this supplement. As there may be readers of the final version who (for whatever reason) also cannot view the supplement, it would be helpful to provide a very succinct description of whether the findings when age was continuous were generally consistent with the age differences found when age was dichotomized.

Typos:
There are a few locations where a parenthetical citation is set within a broader parenthetical statement, and that broader parenthetical statement is missing an end parenthesis (ideally the internal parenthetical statement should also use square brackets). Specifically:
Pg. 4 “(e.g., stronger or more aggressive (Pratto et al., 1994; van Vugt & Tybur, 2014)”
pg. 6. “(e.g., the subordinate asked if they could play, and the high-ranking character said yes; (Gülgöz & Gelman, 2016).”

**Rating scale questions**

|  | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- | --- |
| The study/studies in this manuscript have strong construct validity (good measures and/or manipulations of the constructs the authors wish to study). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  | ✔ |  |
| The study/studies in this manuscript have strong statistical validity (appropriate statistical tests, assumptions are clear and reasonable, no statistical errors, appropriate statistical inferences, etc.). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  |  | ✔ |
| The study/studies in this manuscript have strong internal validity (any causal claims or implications are well-justified, alternative explanations are thoroughly considered, etc.). (Choose “Neutral” if this is not an empirical manuscript, or no causal claims are made or even vaguely implied.) |  |  |  |  | ✔ |
| The study/studies in this manuscript have strong external validity (authors appropriately constrain their conclusions based on the limits of the generalizability of their findings to other contexts (including from lab to real world), other populations, other stimuli or measures, etc.) |  |  |  |  | ✔ |

**Author Response**
Mar 25, 2022

Dear Clare Conry-Murray,

Thank you for the opportunity to address these final concerns. Responses are below in read.

Best,

Ashley

I have now read your revised manuscript. I appreciate your careful attention to the concerns the reviewers and I raised. I am happy to provisionally accept your manuscript for submission. However, I found a few things I would like you to address.

Required revisions:

1. Please examine whether your conclusions would change if you excluded participants who did not identify the person with the crown as the leader before data analysis. You may add this discussion in supplementary online materials, and just refer to it briefly in the manuscript.

This has been added to the supplementary materials. The answer is no. Apologies that we missed this the first round! As mentioned in a previous correspondence we realized there was a flaw in our original analyses, so we updated our approach.

1. Please add to the section on limitations to address the suggestion from Reviewer 2 that Study 3 had two elements that varied from the other studies, and future studies may want to examine each element separately.

We now updated the paragraph in the discussion to say,

“There were two key changes from Story 1a to Story 3a: in Story 1a, the action included ousting an intruder, in Story 3a, the action involved intervention and ‘helping’, thus it is not clear which of these elements led children to answer differently across the two stories.”

“However, neither did children in our study seem to expect complete benevolence from leaders. When we told children that a character helped another character by putting themselves in harm’s way to protect others (Story 3a), children did not choose the leader. This was not because children did not understand this story: when told that someone wanted to help a subordinate but was too scared to do so, children chose the leader less often than the others (Story 3b). Thus, children did not judge that a leader would be more likely than other people to put themselves in harm’s way to protect a subordinate, and they did *not* attribute this behavior to the leader being afraid. There were two key changes across Stories 1a and 3a – In Story 1a, the action only included ousting an intruder, in Story 3a, the action involved intervention and ‘helping’, thus it is not clear which of these elements led children to answer differently across the two stories. One possibility is that the intuition driving children’s choices in Studies 1 and 2 (where children identified the leader as the one who ousted an antisocial intruder) was not the intuition that leaders protect their subordinates but rather that leaders protect their group, or that they punish anti-social individuals. “

I look forward to receiving your final revision and accepting it for publication in Collabra: Psychology.

Please ensure that your revised files adhere to our author guidelines, and that the files are fully copyedited/proofed prior to upload. Please also ensure that all copyright permissions have been obtained. This is the last opportunity for major editing, therefore please fully check your file prior to re-submission.

If you have any questions or difficulties during this process, please contact the editorial office at editorialoffice@collabra.org.

We hope you can submit your revision within the next six weeks. If you cannot make this deadline, please let us know as early as possible.

Sincerely,

Clare Conry-Murray

**Editor Final Decision:** Accept

Mar 29, 2022

Dear Dr. Thomas,

I have now had a chance to read over your manuscript “Children expect leaders to oust intruders, refrain from unprovoked aggression, but not to be generally prosocial”, along with the letter describing the changes you made. Thank you for your responsiveness to the concerns that the reviewers and I raised. I am happy to say that your paper is now officially accepted for publication in Collabra: Psychology. Congratulations on this excellent work, I think it will make an important contribution to the literature and I look forward to seeing it published! I hope your experiences with Collabra: Psychology have been positive and that you will continue to consider it as an outlet for your work.

As there are no further reviewer revisions to make, you do not have to complete any tasks at this point.

You will be receiving separate correspondence regarding any production and technical comments, data deposits, as well as publication charges. We work with the Copyright Clearance Center to process any applicable APC charges. Please note that your APC transaction must be completed before your article gets published.

You will have an opportunity to check the page proofs before we publish your article. Thank you again for publishing in Collabra: Psychology.

Sincerely,
Clare Conry-Murray

1. The scholarly literature includes many different terms for high- and low-ranking individuals. Here we use ‘leader’ for ease of reading. We do not mean leader to be a person who influences a group, as the term has sometimes been used elsewhere (Van Vugt, Smith 2019) here we use it as a shorthand for ‘high-ranking individual’. [↑](#footnote-ref-1)
2. The scholarly literature includes many different terms for high- and low-ranking individuals. Here we use ‘leader’ for ease of reading. We do not mean leader to be a person who influences a group, as the term has sometimes been used elsewhere (Van Vugt, Smith 2019) here we use it as a shorthand for ‘high-ranking individual’. [↑](#footnote-ref-2)