**Peer Review and Communication History**

**MS Title**: How to Cancel Plans: A Mixed Methods Study of Strategy and Experience

**Author Names**: Sophia Caron, Jacqueline Thomas, Alaina Torres, Jeewon Oh, William J. Chopik

**Submitted:** Jun 2, 2022

**Editor First Decision**: Revise & Resubmit

Jul 27, 2022

Dear Chopik,

I have now received 2 reviews of your manuscript, “How to Cancel Plans: A Mixed Methods Study of Strategy and Experience”, from researchers with special expertise in friendship and rejection. I also independently read the manuscript before consulting these reviews. The reviewers had mostly positive reactions to your manuscript. 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 outstanding 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.

Major revisions:

1. Reviewer 1 noted several instances where question wording limits the conclusions we can draw from the data. I also had this reaction to some of the questions (e.g., participants are asked “How much notice do you think your friend should give when cancelling plans?” but the answer options go from “none at all” to “a great deal” without any specifics). I would also have been curious to know whether participants would want a friend to hang out with them when that friend would prefer to be doing something else (e.g., they got a better offer). And, whether participants would prefer that a friend who cancels for a better offer is honest, or vague (not necessarily lying) about the reason. And, whether they think friends/acquaintances are obligated to provide an excuse when canceling. These reflections aren’t particularly helpful post-data-collection, but they do lead me to wonder if the authors can do more to explain their wording and design decisions. And, perhaps they could be helpful if the authors decide to further pursue this line of research.
2. I appreciated Reviewer 2’s point question about how frequencies are to be interpreted. If something is cited infrequently as an “inappropriate excuse,” does that suggest it’s not that inappropriate. Some additional guidance for readers here could be helpful.

Minor revisions:

1. As both reviewers note, it would be informative to include more information on the two sub-samples.
2. The first heading in the results section is: “the affective and relational correlates of canceled plans” but this does not seem to be what is reported (nor does this seem to be analyzed anywhere in the MS).

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 see the instructions below for submitting 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,

Alexa Tullett

Reviewer 1

**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.) |  | ✔ |  |  |  |

**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.

Thank you for the opportunity to review the manuscript, “How to Cancel Plans: A Mixed Methods Study of Strategy and Experience”. This paper investigates in an exploratory manner the ways that people cancel plans with friends as well as their perceived impact. Its strengths include its large cross-sectional sample and examination of qualitative and quantitative data. I also found this paper to cover an interesting topic. However, I had a number of methodological and conceptual questions/concerns, which I outline.

1. A central issue is that it is difficult to interpret some of the results based on the survey questions and design, which affects both the internal and external validity of the findings. For example:
* The dependent variable changes across the main research questions outlined (e.g., ‘helpful’, ‘unforgiveable’, ‘good/bad’) as well as the key questions administered (e.g., upset, offended, and annoying). For the latter, these are arguably distinct emotional responses, and both the rationale and justification for including these variables is unclear. Was the intention to assess different indicators of ‘distress’? At a minimum, justification should be provided for incorporating these particular emotional states.
* I would have a difficult time answering these questions if I were a participant because the reference points for a lot of the questions are vague and unclear. Question 4 asks how much cancellation would ‘affect’ the friendship - affect how? Findings from Question 1 state that people preferred a ‘moderate amount of notice’, but limited clarification is provided on what exactly constitutes a moderate amount.
* Pg. 9: “It is worth noting that most participants were not particularly upset when any of these social actors had to cancel plans (i.e., the absolute ratings on the scales were low)”. Statements such as these are somewhat dubious. Likewise, subjective comparisons between the mean and median as described is not sufficient and more compelling metrics should be used to support conclusions such as “being cancelled on might be upsetting for some people, but it is not particularly upsetting for most people.”
1. More information needs to be provided on the final sample being comprised of separate subsamples of undergraduate and Mturk participants. Were there notable differences between these samples? For example, Mturk paritcipants were recruited “to increase the age and racial/ethnic diversity of the overall sample”. Given this, the sample demographics should also be reported separately somewhere. Also, please clarify whether the recruitment of the complementary Mturk sample was planned before any exploratory analyses had been done.
2. How were the open-ended responses distributed before initial coding of the 20% began? Were they randomly distributed? I see on the data file provided that responses are organized by data source, so I presume coding occurred almost exclusively for the Mturk subsample, which may have affected results.
3. Greater specific on the content coding approach should be provided (e.g., in a supplemental file). The authors describe what they did and reference (Morgan, 1993 Woike, 2007), but it is not entirely clear how they went about the various selection decisions that govern the general recommended steps described. Related to this, I was confused by some of the coding categories that did not seem like distinct categories. For example, what is the difference between Health versus Physically/emotionally not well? As for the example Important category (“Something important came up”), would these not fall under the Unexpected category (and similarly for Romance within Better Offer, Being Mean within Feelings)
4. What is the conceptual significance of this behavior compared to other related social psychological behaviors such as rejection communication? Incorporating aspects of this distinction would certainly bolster the relevance of this work. On this point, I was surprised that this paper does not incorporate or mention relevant research from the communication literature, which seem equally if not more closely linked to this topic. For example, politeness theory (Brown & Levinson, 1987), work on ‘bad news’ messages (Jablin & Krone, 1984), and/or attribution models (e.g., Folkes, 1982) seem potentially relevant since several of the authors’ findings appear consistent with aspects of prior theory.
5. The authors’ findings are confined to the context of cancellations in friendships based on the prompt that was used (i.e,. the purpose of this study is to examine how and why people cancel plans with a friend), and have limited generalizability in speaking to cancellation behavior across other types of relationship contexts. To this end, I think the authors would need to be more precise when introducing and discussing their findings on what the best/worst ways to cancel plans, to ensure this is confined to the context of friendships (unless additional justification can be provided).
6. Although the authors mention individual difference factors potentially influencing the results in the Discussion, I was also expecting some commentary on more context-dependent moderators, such as the canceller’s typicality (i.e., evaluations of the appropriateness/unexpectedness of cancellation based on past behavior) or if the degree of closeness with the friend would be expected to shift what the optimal cancellation strategies are.

More minor comments:

* The reference to comparative rejection should be incorporated earlier when introducing the relevance of rejection
* “The study was envisioned as a qualitative project and completely exploratory—thus, it was not pre-registered”. Conducting exploratory qualitative work does not inherently preclude pre-registration, since various aspects can be considered beforehand (e.g., the target sample, the qualitative analysis to be used).

Reviewer 2

**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.) |  |  |  | ✔ |  |

**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 authors investigate what people consider as appropriate or inappropriate reasons for ‘cancelling plans’. I found this topic to be novel and interesting and I appreciate the author’s descriptive approach. The paper is mostly well written, but I find that the contribution of the research to be quite limited in its current form. I have made a number of suggestions for improvements below.

*Introduction*

I appreciate the brevity of this introduction. Having said that, it could be a little bit more comprehensive in how it introduces various concepts (e.g. what is meant by “plans”?, the investment model) . In particular, the authors mention the investment model, without providing a full explanation of what it is and what it entails. The components of the model should be clearly explained.

The authors state that they will investigate “how they (participants) would hypothetically respond to a few different conditions of being cancelled on”. I did not see any investigation of this kind in the results.

*Methods*

There was overall a high degree of transparency in the methods, which I appreciate. However, it was not clear to me how participants were paid/compensated, and I think this is important to mention.

With respect to sample size justification, the authors mention “We occasionally ran one-sample t-tests against a scale midpoint (i.e., 3 on a 5-point scale).” – What does this refer to?

The authors should include more information on how and why they generated the questions that were included in the survey.

*Results*

It was not clear to me what the “% mentioned” in Table 2 refers to. I assume it is the percentage of participants who included that feature (i.e. code name) in their response, but the authors should clarify. It also might help to use a term for the code derived from the analysis (e.g. features).

Is it implied that the frequency with which something is mentioned, is relevant to how more, or less, reasonable, or inappropriate it is? This may not always be the case. For example, only 1.3% of people mentioned “no excuse provided”, but it is likely that most people would consider that highly inappropriate, the same goes for “change of mind”. A follow up study could present a new sample of participants with these “codes” as hypothetical reasons for cancelling, and have them rate each one according to their appropriateness/reasonableness.

Overall, I appreciate that the research is descriptive and the scope of the question is relatively narrow, but I think more work is needed to give us a more complete picture of how people judge the appropriateness of reasons for cancelling plans. In addition, an alternative method of sampling would be beneficial. Right now, it appears that racial/ethnic diversity comes exclusively from M-Turk, so it may that minorities in the sample are also the only ‘opt-in’ participants. I think this confounds the representativeness of the sample, and such factors are important for descriptive work.

**Author Response**
Nov 29, 2022

**Response to Editor and Reviewers**

We would like to thank the editor and reviewers for their thoughtful comments on the manuscript. We very much appreciate the constructive feedback and believe that the manuscript has improved significantly as a result of their suggestions. Below, we report how each of the reviewer issues was addressed and the corresponding changes to the manuscript. The reviewer comments are non-bolded, and our responses are bolded.

**Editor**

Major revisions:

1. Reviewer 1 noted several instances where question wording limits the conclusions we can draw from the data. I also had this reaction to some of the questions (e.g., participants are asked “How much notice do you think your friend should give when cancelling plans?” but the answer options go from “none at all” to “a great deal” without any specifics). I would also have been curious to know whether participants would want a friend to hang out with them when that friend would prefer to be doing something else (e.g., they got a better offer). And, whether participants would prefer that a friend who cancels for a better offer is honest, or vague (not necessarily lying) about the reason. And, whether they think friends/acquaintances are obligated to provide an excuse when canceling. These reflections aren’t particularly helpful post-data-collection, but they do lead me to wonder if the authors can do more to explain their wording and design decisions. And, perhaps they could be helpful if the authors decide to further pursue this line of research.

**We thank you for highlighting some of the particularly important issues brought up by reviewers. We have now added some extended sections to the manuscript where we better articulate how these decisions were made and the limitations of these items for drawing firm conclusions from the data. We have also removed general statements/conclusions that rely on interpreting the more problematic items in the sample (e.g., the most people feel this or that way). A fuller description of these changes and acknowledgments can be found below in our response to R1 #1.**

**Your particular recommendations for diving further into this phenomenon are well taken, too. We have now sprinkled these future directions throughout the Discussion on pages 24-25.**

1. I appreciated Reviewer 2’s point question about how frequencies are to be interpreted. If something is cited infrequently as an “inappropriate excuse,” does that suggest it’s not that inappropriate. Some additional guidance for readers here could be helpful.

**We appreciated this point as well. The current study merely sought to characterize what various excuses were and their relative frequency. We ultimately cannot make any subjective judgment about their quality (or how unreasonable they are) beyond participants labeling them as reasonable and inappropriate. In fact, we have now added such a caveat to the paper, lest people think we are making a judgment about whether these excuses can be evaluated based on how spontaneously they are mentioned. The reviewer provided a suggestion for future research—in which we could ask an independent sample to rate each of the features/codes generated in this study (this could even lead to further measure developments of this phenomenon). We view this suggestion as incredibly important for future research and have added this consideration to the manuscript on pages 24-25.**

Minor revisions:

1. As both reviewers note, it would be informative to include more information on the two sub-samples.
2. The first heading in the results section is: “the affective and relational correlates of canceled plans” but this does not seem to be what is reported (nor does this seem to be analyzed anywhere in the MS).

**We have now provided descriptive statistics for each sample in Footnote #1X (on pg 8). In terms of notable differences between the two samples, there were significant differences in the age and gender distributions of the samples (i.e., MTurk participants were older and had a more equal gender balance). Surprisingly, the racial/ethnic breakdown of the sample was comparable between the two. Sample comparisons via *t*-tests, chi-squared tests, and effect sizes are now provided. We also clarified that the samples were collected in parallel (on pg 8). We also added a section to the Discussion where we acknowledge that the results from the current study would likely not generalize to other samples (see pg 25).**

**We have now changed this heading to be more descriptive of what was actually measured: “Evaluating Cancellations: Preferences for Advanced Notice, Desire to Reschedule, and Comparing Cancellations by Level of Closeness” (see pg 12).**

**Reviewer #1**

Thank you for the opportunity to review the manuscript, “How to Cancel Plans: A Mixed Methods Study of Strategy and Experience”. This paper investigates in an exploratory manner the ways that people cancel plans with friends as well as their perceived impact. Its strengths include its large cross-sectional sample and examination of qualitative and quantitative data. I also found this paper to cover an interesting topic. However, I had a number of methodological and conceptual questions/concerns, which I outline.

1. A central issue is that it is difficult to interpret some of the results based on the survey questions and design, which affects both the internal and external validity of the findings. For example:
* The dependent variable changes across the main research questions outlined (e.g., ‘helpful’, ‘unforgiveable’, ‘good/bad’) as well as the key questions administered (e.g., upset, offended, and annoying). For the latter, these are arguably distinct emotional responses, and both the rationale and justification for including these variables is unclear. Was the intention to assess different indicators of ‘distress’? At a minimum, justification should be provided for incorporating these particular emotional states.
* I would have a difficult time answering these questions if I were a participant because the reference points for a lot of the questions are vague and unclear. Question 4 asks how much cancellation would ‘affect’ the friendship - affect how? Findings from Question 1 state that people preferred a ‘moderate amount of notice’, but limited clarification is provided on what exactly constitutes a moderate amount.
* Pg. 9: “It is worth noting that most participants were not particularly upset when any of these social actors had to cancel plans (i.e., the absolute ratings on the scales were low)”. Statements such as these are somewhat dubious. Likewise, subjective comparisons between the mean and median as described is not sufficient and more compelling metrics should be used to support conclusions such as “being cancelled on might be upsetting for some people, but it is not particularly upsetting for most people.”

**We have now provided some initial context regarding how the close-ended items were generated. Specifically, they were generated from a group discussion between the corresponding author and two students about the possibly negative emotional responses and considerations someone might have to be cancelled on. We focused on what we considered to be the main negative emotions someone would likely have based on previous literature on related experiences (social rejection/ostracism; Hartgerink et al., 2015); however, we acknowledge that the responses queried in these questions are not an exhaustive treatment of the emotions and reactions of people who experience cancellations. Of course, these three questions only captured a sliver of the variety of responses people could have. For example, they might feel relatively indifferent to having a friend cancel on them. Likewise, there are plenty of situations that make people feel relieved when plans are cancelled. Because we went in with the assumption that being cancelled on was a negative experience—and our group discussion to generate these items reflects that—we likely missed a more accurate characterization of this experience. We hope that future research will use more established scales to capture the many different ways that people experience cancellations. We discuss this point further in the Discussion and caution readers from over-interpreting our materials as being a comprehensive assessment of people’s reactions to being cancelled on. This addition can be seen on pages 9-10 and 24-25.**

**We acknowledge that some of the questions might have been vague, unclear, and/or difficult to answer. This is particularly the case given that the questions we created only focused on a few distinct negative emotions. Regarding advance notice of cancellations, we purposely asked a more definitive question (providing participants with exact, time-discrete bins for when they would be most annoyed) and a more subjective assessment (ranging from none at all to a great deal for how much notice they would prefer). Worth noting, these two questions had a very small correlation with one another, *r* = -.093, *p* < .001, suggesting that people’s preferences for notice and perceived annoyance do not overlap highly, which was a surprise to us. We have now acknowledged this limitation in Footnote #2 on page 10. In that same footnote, we have also acknowledged the limitation that there are many ways a cancellation could “affect” a friendship and that our item was unable to capture these nuances.**

**In response to the reviewer’s comment, we have removed such qualitative statements based on the scale responses. We were trying to make a statement based on a percentage of people (last column of Table 1) who said they would be at least a moderate amount of upset, but we realize that strong statements like these neglect the variation in these responses and do not quite characterize people well (especially given the limitations of the item wordings that the reviewer pointed out).**

1. More information needs to be provided on the final sample being comprised of separate subsamples of undergraduate and Mturk participants. Were there notable differences between these samples? For example, Mturk paritcipants were recruited “to increase the age and racial/ethnic diversity of the overall sample”. Given this, the sample demographics should also be reported separately somewhere. Also, please clarify whether the recruitment of the complementary Mturk sample was planned before any exploratory analyses had been done.

**We have now provided descriptive statistics for each sample in Footnote #1 (on pg 8). In terms of notable differences between the two samples, there were significant differences in the age and gender distributions of the samples (i.e., MTurk participants were older and had a more equal gender balance). Surprisingly, the racial/ethnic breakdown of the sample was comparable between the two. Sample comparisons via *t*-tests, chi-squared tests, and effect sizes are now provided. We also clarified that the samples were collected in parallel (on pg 8). We also added a section to the Discussion where we acknowledge that the results from the current study would likely not generalize to other samples (see pg 25).**

1. How were the open-ended responses distributed before initial coding of the 20% began? Were they randomly distributed? I see on the data file provided that responses are organized by data source, so I presume coding occurred almost exclusively for the Mturk subsample, which may have affected results.

**Upon completion of data collection, open-ended responses were randomly assigned an identification number (for later re-merging following coding) and migrated to a separate file for coding purposes. Their order was scrambled prior to the provision for the initial coding and subsequent coders. This separation and scrambling were purposely done for the very reason the reviewer suggested, in addition to another—so (a) the coding was not unduly influenced by having only data from one source and (b) to keep the coders blind to other characteristics that the participants provided (and their demographics). After codes were generated, the data were merged back into the main data file and replaced the text responses (so we could publicly share the data and reduce the chance of identification from open-ended responses). Of course, because most of the data came from undergraduate participants (82.6%), it is possible that the codes might disproportionally reflect undergraduate student responses, but this was not intended or done in a systematic way to bias the coding in that regard. We realize none of this was articulated in the manuscript and added it to Footnote #3 on page 13.**

1. Greater specific on the content coding approach should be provided (e.g., in a supplemental file). The authors describe what they did and reference (Morgan, 1993 Woike, 2007), but it is not entirely clear how they went about the various selection decisions that govern the general recommended steps described. Related to this, I was confused by some of the coding categories that did not seem like distinct categories. For example, what is the difference between Health versus Physically/emotionally not well? As for the example Important category (“Something important came up”), would these not fall under the Unexpected category (and similarly for Romance within Better Offer, Being Mean within Feelings)

**We thank the reviewer for encouraging us to provide more details on the content coding approach. We have now prepared a separate supplementary file and uploaded it with the manuscript (and placed on OSF) that describes our process of developing the content analysis in line with the steps described by Woike (2007). The supplement is referenced on page 13.**

**In this file, we provide some greater clarity within this file regarding some of the conceptual overlap of categories. The criterion for creating a coding category was the occurrence of that feature at least once within a particular domain and was done inductively (reasons that were health- or illness-related were categorized as “health”; a different health-adjacent theme of not feeling well was distinct enough [and didn’t include a reference to a health/illness problem] to warrant its own category). Specifically, the coding procedure tried to stay faithful to participants’ original entries—if they said “better offer” that was categorized as “better offer.” If they referenced a romantic possibility, that was categorized as romantic possibility rather than a “better offer.”**

**Although some specific excuses (i.e., a family-, health-, or work-relevant excuse) may be reasonably considered to be something unexpected or important, we still stuck to participants’ specific mentions. In other words, an entry was categorized as “Health” if it had to do with an individual health concern (not generalized feeling); an entry was categorized as “unexpected” if that or another word (e.g., emergency”) was used; and an entry was categorized as “important” if that specific qualifier was used by the participant. Assuredly, family-, health-, and work-relevant excuses could reasonably be considered important, too. However, they were only coded that way if participants spontaneously labeled them as important. In this way, multiple categories could be nominated/coded for a particular unit of analysis, and we avoided making the assumption that something like a work-relevant excuse is considered important or unexpected unless otherwise stated by the participant. We acknowledge that there is likely some conceptual overlap in these categories, but in the cases where a code was not explicitly mentioned, it was not coded.**

1. What is the conceptual significance of this behavior compared to other related social psychological behaviors such as rejection communication? Incorporating aspects of this distinction would certainly bolster the relevance of this work. On this point, I was surprised that this paper does not incorporate or mention relevant research from the communication literature, which seem equally if not more closely linked to this topic. For example, politeness theory (Brown & Levinson, 1987), work on ‘bad news’ messages (Jablin & Krone, 1984), and/or attribution models (e.g., Folkes, 1982) seem potentially relevant since several of the authors’ findings appear consistent with aspects of prior theory.

**We thank the reviewer for recommending this literature to integrate into the paper. We have now done so in the Introduction (pgs 5-7) and Discussion (pg 21). However, we stopped short of using these theories to make formal predictions about what to expect in the data (an unethical consideration given that we already know the results; Kerr, 1998).**

**Regarding the conceptual significance of cancelling plans, we now point out in the manuscript that cancelling plans is quite a bit different than outright rejecting someone (as featured in the recommended work on job applicants and romantic/sexual advances). Cancelling plans involves an initial commitment to spend time with someone and then later reneging on those plans (which is different than what the interpersonal rejection literature has focused on to date). Despite some theory and intuition suggesting a right and a wrong way to go about cancelling plans, several unanswered questions remain about cancelling plans, specifically. This addition is now articulated on page 7.**

1. The authors’ findings are confined to the context of cancellations in friendships based on the prompt that was used (i.e,. the purpose of this study is to examine how and why people cancel plans with a friend), and have limited generalizability in speaking to cancellation behavior across other types of relationship contexts. To this end, I think the authors would need to be more precise when introducing and discussing their findings on what the best/worst ways to cancel plans, to ensure this is confined to the context of friendships (unless additional justification can be provided).

**We couldn’t agree more. We have now changed the title to include “friends” to make this crystal clear. We have also further highlighted throughout the paper that the study at hand was about friendships, specifically. Finally, we have added the observation that our study was specific to friendships (and not other contexts) to the Limitations section on page 25.**

1. Although the authors mention individual difference factors potentially influencing the results in the Discussion, I was also expecting some commentary on more context-dependent moderators, such as the canceller’s typicality (i.e., evaluations of the appropriateness/unexpectedness of cancellation based on past behavior) or if the degree of closeness with the friend would be expected to shift what the optimal cancellation strategies are.

**We found these recommendations fascinating and have incorporated them into the manuscript on page 23-24.**

More minor comments:

* The reference to comparative rejection should be incorporated earlier when introducing the relevance of rejection
* “The study was envisioned as a qualitative project and completely exploratory—thus, it was not pre-registered”. Conducting exploratory qualitative work does not inherently preclude pre-registration, since various aspects can be considered beforehand (e.g., the target sample, the qualitative analysis to be used).

**Thanks for these as well. We have now mentioned the comparative rejection in the second paragraph when we discuss politeness theory (i.e., that it would be particularly important to maintain a positive face with very close friends; see pg 5). Also, we have now removed the qualifier regarding not pre-registering because of the exploratory and qualitative elements of the study.**

**Reviewer #2**

The authors investigate what people consider as appropriate or inappropriate reasons for ‘cancelling plans’. I found this topic to be novel and interesting and I appreciate the author’s descriptive approach. The paper is mostly well written, but I find that the contribution of the research to be quite limited in its current form. I have made a number of suggestions for improvements below.

1.) Introduction

I appreciate the brevity of this introduction. Having said that, it could be a little bit more comprehensive in how it introduces various concepts (e.g. what is meant by “plans”?, the investment model) . In particular, the authors mention the investment model, without providing a full explanation of what it is and what it entails. The components of the model should be clearly explained.

The authors state that they will investigate “how they (participants) would hypothetically respond to a few different conditions of being cancelled on”. I did not see any investigation of this kind in the results.

**We thank the reviewer for their kind words about the manuscript. Following the recommendation of R1, we have significantly expanded the Introduction to provide a broader context (i.e., theories) for interpersonal communication and how they might be related to cancelling plans. Regarding defining the various concepts, we have now clarified that “plans” are primarily comprised of shared leisure activities (Larson et al., 1986) on the edge of page 3 onto page 4. We have also further defined the elements of the Investment Model (on page 4). In line with the comments provided by the Editor and R1, we have also edited the manuscript to clarify our contributions by sticking to what was actually done and assessed (i.e., not hypothetical responses to cancellations, rather their thoughts and opinions on the best ways to cancel plans with friends).**

2.) Methods

There was overall a high degree of transparency in the methods, which I appreciate. However, it was not clear to me how participants were paid/compensated, and I think this is important to mention.

With respect to sample size justification, the authors mention “We occasionally ran one-sample t-tests against a scale midpoint (i.e., 3 on a 5-point scale).” – What does this refer to?

The authors should include more information on how and why they generated the questions that were included in the survey.

**We have now clarified that undergraduate students were awarded course credit; MTurk participants were compensated $.50 on page 8. We have now edited the statement referring to t-tests to be more specific in what we were referring to (i.e., comparing some of the closed-ended questions against each other (i.e., comparing cancellations from acquaintances, close friends, and friends) and one-sample t-tests to test whether the mean of a closed-ended question was significantly above or below the midpoint (for descriptive purposes).**

**In line with the comments from R1, we have provided additional context for how the questions were generated. Specifically, they were generated from a group discussion between the corresponding author and two students about the possibly negative emotional responses and considerations someone might have to be cancelled on. We focused on what we considered to be the main negative emotions someone would likely have based on previous literature on related experiences (social rejection/ostracism; Hartgerink et al., 2015); however, we acknowledge that the responses queried in these questions are not an exhaustive treatment of the emotions and reactions of people who experience cancellations. Of course, these three questions only captured a sliver of the variety of responses people could have. For example, they might feel relatively indifferent to having a friend cancel on them. Likewise, there are plenty of situations that make people feel relieved when plans are cancelled. Because we went in with the assumption that being cancelled on was a negative experience—and our group discussion to generate these items reflects that—we likely missed a more accurate characterization of this experience. We hope that future research will use more established scales to capture the many different ways that people experience cancellations. We discuss this point further in the Discussion and caution readers from over-interpreting our materials as being a comprehensive assessment of people’s reactions to being cancelled on. This addition can be seen on pages 9-10 and 24-25 (see R1 #1 for more details).**

3.) Results

It was not clear to me what the “% mentioned” in Table 2 refers to. I assume it is the percentage of participants who included that feature (i.e. code name) in their response, but the authors should clarify. It also might help to use a term for the code derived from the analysis (e.g. features).

**We have now clarified that the “% mentioned” column is calculated by summing the number of occurrences of a feature and then dividing by the total number of eligible entries that feature could appear in, and multiplying by 100 (this is now in the Note for Table 2 and listed in an expanded supplement requested by R1). We have now also relabeled them as “features” in Table 2 and in the main text as the reviewer recommended.**

4.) Is it implied that the frequency with which something is mentioned, is relevant to how more, or less, reasonable, or inappropriate it is? This may not always be the case. For example, only 1.3% of people mentioned “no excuse provided”, but it is likely that most people would consider that highly inappropriate, the same goes for “change of mind”. A follow up study could present a new sample of participants with these “codes” as hypothetical reasons for cancelling, and have them rate each one according to their appropriateness/reasonableness.

**We thank the reviewer for bringing this up. The current study merely sought to characterize what various excuses were and their relative frequency. We ultimately cannot make any subjective judgment about their quality (or how unreasonable they are) beyond participants labeling them as reasonable and inappropriate. We view this suggestion as incredibly important for future research and have added this consideration to the manuscript on pages 24-25.**

5.) Overall, I appreciate that the research is descriptive and the scope of the question is relatively narrow, but I think more work is needed to give us a more complete picture of how people judge the appropriateness of reasons for cancelling plans. In addition, an alternative method of sampling would be beneficial. Right now, it appears that racial/ethnic diversity comes exclusively from M-Turk, so it may that minorities in the sample are also the only ‘opt-in’ participants. I think this confounds the representativeness of the sample, and such factors are important for descriptive work.

**We thank the reviewer for their recommendations and comments on the manuscript. We agree with their comment that an alternative sampling method would be beneficial and have added a discussion about this to page 25.**

**Worth noting, and surprising to us (given the goal of recruiting MTurkers), when we formally examined whether racial/ethnic representation differed across the two samples, we found that the two samples were not significantly different from one another, *χ*2(8) = 13.15, *p* = .107 (e.g., 79.1% White among MTurk participants, 70.7% White among subject pool participants; this is now added as Footnote #1). We have now amended the statement in the Method to reflect that our coordinated collection enhanced diversity (i.e., age, gender) and maximized our sample size. Nevertheless, we accept the reviewer’s general observation and have now discussed it in the paper.**

**Editor Final Decision:** Accept

Dec 7, 2022

Dear William Chopik,

I have now had a chance to read over your manuscript “How to Cancel Plans with Friends: A Mixed Methods Study of Strategy and Experience”, along with the letter describing the changes you made. Sorry for the delay, and thank you for your patience and 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,
Alexa Tullett