A significant portion of the data analysis for this paper involved content coding open-ended responses to three questions. To code the responses, we migrated through the recommended steps proposed by qualitative researchers. These steps of qualitative content coding are articulated in Woike (2007). After each step, we discuss the particular decisions and considerations that went into the coding of the cancelling plans data.

Step 1: What is the research question? What is it to be identified, described, or measured?

The purpose of the study was to characterize the phenomenon of being cancelled on. For this particular analysis, the research questions were: how would participants prefer their friends go about cancelling plans on them? What do participants consider a reasonable excuse to cancel plans? What do participants consider an inappropriate excuse to cancel plans?

This decision is in line with the practice of using open-ended responses (and content coding) to assess people's understandings and interpretations of events (Pennebaker et al., 1997).

Step 2: Decide whether content analysis will provide the needed information, either by itself or in conjunction with another method.

Content analysis provides the best way to assess, via participant-generated data, the occurrence and frequency of preferences for cancellations and good/bad excuses for doing so. These directed questions were selected because of this targeted goal, rather than a more expansive prompt (e.g., "discuss a time that you were cancelled on in depth") which might have been subjected to a different qualitative analysis. The content analysis is examined in isolation of the other closed-ended data collected, although further analysis is possible.

Step 3: Decide what type of qualitative material will best provide the information needed and how to obtain it.

These relatively narrow open-ended items provide the type of responses that lent themselves to the research questions (see above). Participants' short-answers were the best way to assess this phenomenon. This was also done for the ease of online data collection as face-to-face interviews were not possible and expanded essay-type questions would have burdened participants (both of which were beyond the scope of the current work). These were people's spontaneous answers and the questions were not meant to pull for any particular context or reasoning for cancelling.

Step 4: Determine the unit of analysis to be coded.

The unit of analysis was the participants' whole responses to each of the prompts/questions. They were coded in their entireties and not combined across questions.

Step 5: Select or develop a content coding system.

The coding system of "frequencies" was selected (instead of categories or ratings). Specifically, this approach was chosen to develop meaningful units of the response and recording the number of instances of these units in the data. This also allowed us to have multiple codes per entry, which is not always the case (e.g., rating systems that provide one evaluative judgment per unit of analysis). Before the corresponding author developed the coding categories, they elected to create codes for recurring instances within each question across units of analysis. This involved reading 20% of the open-ended responses and identifying general categories that the separate coders could then use. Thus, the category generation was not done a priori and was entirely a descriptive assessment of the actual data.

The criteria for creating a coding category was the occurrence of that feature at least once within a particular domain. Coding was done inductively (reasons about health-related was categorized as health; a different health-adjacent theme of not feeling well was distinct enough to warrant its own category). In general, we assigned descriptive codes to summarize participants' responses. The coding procedure tried to stay faithful to participants' original entries—if they said "better offer" that was categorized as "better offer" (i.e., In Vivo coding; Miles et al., 2018). However, when responses specifically referenced a romantic possibility/offer, that was categorized as romantic possibility rather than a "better offer". Similarly, the list of reasonable excuses included general excuses (e.g., something unexpected or important) and specific excuses (e.g., family member is in the hospital). We categorized responses with the more specific code when possible and evaluations (of whether something is important) were solely based on participant's own words. Although some specific excuses (i.e., a family-, health-, or work-relevant excuse) may be reasonably considered to be something unexpected or important, we 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 relevant 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 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.

This resulted in 8 categories for cancelling preferences, 6 (later 7) categories for reasonable excuses, and 7 (later 10) categories for the unreasonable excuses. In the event that coders encountered additional categories when reading through all the responses, the study team agreed to revisit the codes to see if adding another category made sense. Indeed, this happened for some categories that were not in the original 20% of responses read by the corresponding author (i.e., the additions of the money categories for the reasonable/inappropriate question coding; being mean, and change of mind were added to the inappropriate question coding). These codes are presented in Table 2 of the main manuscript.

Step 6: Obtain pilot data to test and to refine the coding system.

The pilot data phase consisted of 20% of respondents read over by the corresponding author. As stated above, the coding system could be refined for additional categories not defined in the initial set. The coding scheme/definitions/examples were relatively straightforward and are presented in Table 2 of the main manuscript. Two research assistants unfamiliar with the project were recruited as coders (i.e., the empirical/independent approach).

Step 7: Train coders and ensure that inter-coder agreement is satisfactory.

The corresponding author then discussed the coding scheme (see Table 2) with two research assistants. They reviewed some exemplar responses (also in Table 2). They then coded 200 randomly chosen responses (non-overlapping with these aforementioned 20%). The two coders were reliable (kappa (k) = .95). Disagreements were discussed among the two coders and the corresponding author (who developed the coding scheme). The two coders then proceeded to code the remaining entries (including the original 20% for due diligence). Coders noted whether a particular response was either gibberish or blank. These responses were treated as missing such that the percentages presented below reflect the frequency a code is mentioned out of eligible responses. This approach yielded a large amount of usable data for how people should go about cancelling (N = 1134; 58 exclusions), what a good excuse is (N = 1065; 127 exclusions), and what a bad excuse is (N = 1192; 0 exclusions).

Step 8: Obtain final material to be analyzed.

After coding 200 random responses, the two research assistants coded the remaining responses (split evenly between them), as the training data set had come from the full data set.

Step 9: Code the material with identifying characteristics removed, and determine interrater reliability; or perform computer-assisted content analysis.

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 so the coding was not unduly influenced by having only data from one source and 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 they could be publicly shared). We reported on the reliability from the training sample; research assistants coded the remaining entries in a non-overlapping fashion. No computer-assisted content analysis was conducted.

Step 10: Analyze the data; carry out cross-validation if appropriate.

Data were analyzed by reporting the frequencies of each category from the entire sample of entries for each question. The final frequencies are reported in Table 2. The “% mentioned” column is calculated by summing the number of occurrences of a feature, dividing by the total number of eligible entries that feature could appear in, and multiplying by 100.

References

Miles, M. B., Huberman, A. M., & Saldaña, J. (2018). *Qualitative data analysis: A methods sourcebook*. Sage Publications.

Pennebaker, J. W., Mayne, T. J., & Francis, M. E. (1997). Linguistic predictors of adaptive bereavement. *Journal of Personality and Social Psychology*, *72*, 863.

Woike, B. A. (2007). Content coding of open-ended responses. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 292-307). The Guilford Press.