



Contents lists available at ScienceDirect

## Journal of Experimental Social Psychology

journal homepage: [www.elsevier.com/locate/jesp](http://www.elsevier.com/locate/jesp)

## FlashReports

## In pursuit of the proper null: Reply to Chen and Risen (2009)

Brad J. Sagarin \*, John J. Skowronski

Northern Illinois University, Department of Psychology, DeKalb, IL 60115, USA

## ARTICLE INFO

## Article history:

Received 28 August 2008

Available online 30 September 2008

## Keywords:

Cognitive dissonance

Measurement

Choice

## ABSTRACT

Chen argued that the proper null hypothesis for free-choice studies examining shifts in choice was 66.7%. Sagarin and Skowronski (2009) questioned the appropriateness of this value, noting that it was based on an unwarranted assumption that subjects always choose preferred options over less preferred options. In this paper, we respond to the points raised by Chen and Risen (2009), noting that: (a) violations of an additional unwarranted assumption (perfect transitivity) also move the proper null hypothesis towards 50%; (b) the validation of pretest measures would enable researchers to estimate an upper bound on the proper null; (c) the “blind” choice methodology proposed by Sagarin and Skowronski places the null unambiguously at 50%; and (d) Sagarin and Skowronski correctly call for null-hypothesis tests where needed. In the end, we again endorse the idea that this debate is best resolved empirically, but we believe the empirical avenues available are wider than those endorsed by Chen and Risen.

© 2008 Elsevier Inc. All rights reserved.

Chen and Risen (2009) properly note that the debate about free-choice studies that examine post-decision shifts in choice centers on the appropriate value for the null hypothesis. The analysis presented in Chen (2008) suggested that the proper value for the null is 66.7%. The analysis presented in Sagarin and Skowronski (2009) questioned the appropriateness of this value.

It did so by noting that Chen's (2008) reasoning was based on an unsubstantiated assumption: that choice-making maps onto pre-choice preferences in such a way that the preferred option is *always* chosen over the non-preferred option. We noted that there is a long history of research in choice-making suggesting that subjects *do not always* choose their preferred option. Accordingly, we argued that subjects would select a preferred option over a less preferred option only a proportion of the time, and that this proportion is likely related to the magnitude of the preference difference between the two options. In this regard, we noted that pretesting prior to choice experiments usually indicates that the preference difference among options is small and that this provides additional reason to believe that people would only be somewhat more likely to choose the preferred option over a less preferred option.

The extent to which our analysis is valid suggests that the actual expected value for choice-making behavior is lower than Chen's (2008) proposed 66.7% figure. In fact, we suggested that researchers' pretesting makes it likely that the proper null is much closer to

50% than to 66.7%—a likelihood that can be increased through procedures such as those used by Egan, Santos, and Bloom (2007). These researchers “conservatively used each subject's least preferred color of the three (i.e., the one the monkey took longest to obtain during preference testing) as option C [the new item pitted against the item rejected at time 1]” (p. 980).

Other unsubstantiated assumptions might similarly lead one to believe that the true expected value is less than 66.7%. For example, the value that is output from Chen's (2008) analysis requires perfect transitivity in choice-making. Perfect transitivity is illustrated as follows: If a red M&M is preferred over a blue M&M, and a blue M&M is preferred over a green M&M, then the red M&M will always be preferred over the green M&M. To illustrate the implications of imperfect transitivity, we return to an example used in our original paper. The example involved choices among three M&Ms of differing colors. Given the assumption of perfect transitivity, there are six possible permutations of preference order:

- |                        |                        |
|------------------------|------------------------|
| (1) Red > Blue > Green | (4) Blue > Green > Red |
| (2) Red > Green > Blue | (5) Green > Red > Blue |
| (3) Blue > Red > Green | (6) Green > Blue > Red |

Chen (2008) argues that the first choice provides information regarding subjects' ordinal preferences; this constrains subsequent choices. For example, if at time 1 a subject chooses the red M&M over the blue M&M, Chen argues that such a choice eliminates three permutations (3, 4, and 6). It does so because those permutations specify that blue is preferred to red, which is inconsistent with

\* DOI of original article: [10.1016/j.jesp.2008.08.026](https://doi.org/10.1016/j.jesp.2008.08.026)

\* Corresponding author. Fax: +1 815 753 8088.

E-mail address: [bsagarin@niu.edu](mailto:bsagarin@niu.edu) (B.J. Sagarin).

the initial choice of red. Importantly, green is preferred to blue in two out of three of the remaining permutations (2 and 5, not 1). Hence, Chen argues that the baseline probability of choosing a green M&M in the second choice is 66.7%.

However, this logic assumes perfect transitivity. Imperfect transitivity adds two additional permutations:

- |   |
|---|
| (7) Red > Blue; Blue > Green; Green > Red |
| (8) Blue > Red; Red > Green; Green > Blue |

Under these conditions, a time 1 choice of red over blue eliminates four permutations (3, 4, 6, and 8). Of the remaining permutations, green is preferred to blue in only two out of four (2 and 5, not 1 or 7). Thus, violations of the transitivity assumption will also move the proper null hypothesis towards 50%, with the null reaching 50% if intransitive permutations (7 and 8) are as common as transitive permutations (1 through 6).

How common are intransitive permutations? Psychological researchers have produced them in the laboratory, often through the careful manipulation of the characteristics of choice options (e.g., see Gonzalez-Vallejo, 2002; Mellers & Biagini, 1994). This suggests that such permutations exist, although they are likely less frequent than transitive permutations. Nonetheless, this research demonstrates that violations of transitivity in choice can occur, and to the extent that they do, such violations work against the argument proposed by Chen.

The general thrust of our argument is that the more often the assumptions made by Chen (2008) overstate subjects' real behavioral tendencies in choice-making, the more the expected value for the second choice in a typical two-choice study falls below 66.7% and approaches 50%. As the analysis presented in Chen and Risen (2009) demonstrates, a lowering of the expected value to a level somewhat above 50% would not threaten inferential conclusions to be derived from all experiments in which the actual choice probability is in the mid-60% range. Another way of saying this is that we think that Chen's analysis leads to the possibility that the current literature might overestimate the effect size that is present in the two-choice paradigm, but that the inferential conclusion that an initial choice affects subsequent choices may still be valid.

However, we again emphasize that this is ultimately an empirical question (which is an explicit admission that, despite our arguments and intuitions, Chen, 2008, might be correct). We see two broad approaches to answer this question:

The first approach is to try to bag the elusive null "in the wild." This is no easy task. Researchers have used a variety of ancillary measures of preference such as selection latency (e.g., rapid selection of especially desirable alternatives; see Egan et al., 2007). However, as Chen and Risen (2009) point out, single ancillary measures of preference might lack validity. Perhaps subjects quickly snatch offending M&Ms to remove them from their sight while leisurely contemplating desirable M&Ms before consuming them. However, proper methodology could eliminate such problems. For example, researchers could validate a pretest by examining choices between items pretested to be more or less separate in magnitude of preference, thus establishing a relation between the magnitude of any pretest difference and choice probabilities. Alternatively, researchers could instill preferences of varying magnitudes through reinforcement and estimate the relationship between choice probabilities and pretest differences. Moreover, multiple measures could be used to more accurately triangulate in on latent preferences. Then, when a sufficiently accurate pretest (or suite of pretests) has been validated, it could be applied to our first proposed experimental methodology: estimating an upper

bound of *P*. Results of past studies could then be reinterpreted in the context of this new null.

In this regard we wish to point out that a careful read of Sagarin and Skowronski (2009) will reveal that Chen and Risen's (2009) claim that we misunderstand the function of the null-hypothesis is incorrect. Where appropriate, we included direct calls for significance testing (e.g., "Statistical tests would be needed, of course, to determine whether these data-based choice rates differ significantly from 54.2%, but they certainly suggest rejection at time 2 of the unchosen time 1 option."). Elsewhere, in conditions in which significance tests are not needed, we do not call for them (e.g., "It turns out that fully accounting for the choice data observed in the study of capuchins would require a value of *P* of 88.7%—a value that, in the context of the pretest data showing no measurable preferences, seems highly implausible.").

It should be noted, of course, that non-significance in such reinterpretations is not synonymous with equivalence. A non-significant effect can still contribute to a confident conclusion that an effect is actually present when synthesized meta-analytically with other effects. This is especially likely when the non-significant study may exhibit a large effect size, yet yield a result that is non-significant because of an insufficient number of observations. Hence, the cautious way to proceed would be to withhold conclusions pending re-analysis and meta-analytic synthesis of the entire corpus of existing data.

A second approach is to eliminate the ambiguous null. To this end, in Sagarin and Skowronski (2009) we suggested an experimental methodology in which the first choice is made "blind." Because subjects would not know which item they were choosing, their choice could not reflect their preferences, but it still could affect their preferences. Importantly, this methodology requires no assumptions regarding the validity of ancillary measures. Even better, the proper null stands unambiguously at 50%, thus addressing the problems identified in Chen (2008).

Either of these two types of experiments, properly executed, would yield fruitful data and would belie Chen and Risen's (2009) assertion that such experimental methods cannot address the problem. We make no claim, of course, that these are the only methodologies capable of addressing Chen's (2008) concerns. We look forward to reading the approaches referred to in the Chen and Risen response paper.

Lastly, we note that Sagarin and Skowronski's (2009) comments addressed Chen's (2008) criticism of free-choice paradigms that examine shifts in choice; Chen and Risen (2009) are correct in asserting that Sagarin and Skowronski did not address Chen's criticism of free-choice paradigms that examine shifts in ranking or ratings. In this regard, we note that one element of our critique was based on Chen's failure to account for the imperfect relationship between preference and choice in studies that examine shifts in choice. We concur with Chen and Risen that Chen's criticism of studies that examine shifts in ranking or ratings models this imperfect relationship. However, we were puzzled as to why Chen incorporated this imperfect relationship in one section of his paper (the section on shifts in rankings or ratings) but ignored it in another (the section on shifts in choice). Nonetheless, we reiterate that our analysis is limited to Chen's critique of shifts in choice; we leave theoretical and/or empirical critiques of Chen's analysis of shifts in rating or ranking to others.

In the end, we return to the common ground that we share with Chen and Risen (2009): This debate is best resolved empirically. We are excited by the myriad approaches available to researchers, and we look forward to seeing how they turn out—and we still wonder if anyone has a few spare M&M's that can be put to good use in research.

**References**

- Chen, M. K. (2008). *Rationalization and cognitive dissonance: Do choices affect or reflect preferences?* (Cowles Foundation Discussion Paper No. 1669). New Haven, CT: Yale University, Cowles Foundation for Research in Economics. Available from <http://cowles.econ.yale.edu/P/cd/d16b/d1669.pdf> (Retrieved July 14, 2008).
- Chen, M. K., & Risen, J. L. (2009). Is choice a reliable predictor of choice? A comment on Sagarin and Skowronski. *Journal of Experimental Social Psychology*, 45, 427–429.
- Egan, L. C., Santos, L. R., & Bloom, P. (2007). The origins of cognitive dissonance: Evidence from children and monkeys. *Psychological Science*, 18, 978–983.
- Gonzalez-Vallejo, C. (2002). Making trade-offs: A probabilistic and context-sensitive model of choice behavior. *Psychological Review*, 109, 137–155.
- Mellers, B. A., & Biagini, K. (1994). Similarity and choice. *Psychological Review*, 101, 505–518.
- Sagarin, B. J., & Skowronski, J. J. (2009). The implications of imperfect measurement for free-choice carry-over effects: Reply to M. Keith Chen's (2008) "Rationalization and cognitive dissonance: Do choices affect or reflect preferences?" *Journal of Experimental Social Psychology*, 45, 424–426.