Journal of Experimental Social Psychology 45 (2009) 425–427

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

journal homepage: www.elsevier.com/locate/jesp

FlashReports

Is choice a reliable predictor of choice? A comment on Sagarin and Skowronski ☆

M. Keith Chen [a,*], Jane L. Risen [b]

[a] School of Management and the Department of Economics, Yale University, 135 Prospect Street, New Haven, CT 06511, USA
[b] Booth School of Business, University of Chicago, 5807 South Woodlawn Ave, Chicago IL, 60637, USA

ARTICLE INFO

Article history:
Received 30 August 2008
Available online 25 September 2008

Keywords:
Cognitive dissonance
Free-choice paradigm
Mere-choice effects
Preference construction
Revealed preferences

ABSTRACT

In a recent working paper, Chen (2008) argues that a methodology central to the cognitive dissonance literature (the free-choice paradigm) has suffered from an inability to separately measure how much choices affect people's preferences, and how much they simply reflect those preferences, by failing to fully control for the fact that subjects tend to choose goods they prefer. Although Sagarin and Skowronski concede this, they discount Chen's argument, claiming that for revealed preferences to completely account for observed choice-effects the relationship between choice and preference would have to be unrealistically high. In this comment, we argue that their critique both misses the crux of Chen's analysis, and is incorrect. Specifically, to properly test whether choices affect preferences, it is essential that researchers experimentally control for revealed preferences rather than speculate how much of a role they may play. Moreover, Sagarin and Skowronski's critique rests on two fundamental errors—a misunderstanding of the function of the null-hypothesis and a misunderstanding of preference-measurement psychometrics. These errors leads Sagarin and Skowronski to suggest alternative experimental designs which while a good first step, do not address the problems identified in Chen (2008).

Since Brehm (1956) initial free-choice experiment, psychologists have argued that choosing between alternatives creates dissonance, causing people to shift their preferences to rationalize their choices. In his working paper, Chen (2008) argues that the free-choice paradigm fails to include a proper control group which would account for the information that is revealed by participants' choices. In other words, although psychologists recognize that an individual's choice conveys information about their preferences, the free-choice paradigm does not account for this information. As a result, Chen argues, we cannot know whether the free-choice paradigm measures preferences shifts resulting from making choices (i.e., because of dissonance reduction) or whether these shifts are simply predicted by participants' existing preferences (which have been revealed by their choices).

In their critique of Chen (2008), Sagarin and Skowronski's (hereafter S&S) acknowledge that the free-choice paradigm does not control for choice information. However, they argue that for revealed preferences to account for existing results, subjects would have to display a relationship between choice and preference that they believe is unrealistically high. In this comment we argue, first and foremost, that rather than speculate about the information revealed in choice, researchers should use an experimental design that explicitly controls for that information (see Chen & Risen, in preparation; Risen & Chen, submitted for publication). Second, we argue that S&S's argument rests on two errors. We believe that they argue that the relationship between choice and preference needs to be unrealistically high because they both misunderstand the function of a null hypothesis and misapply core psychometric principles. This leads S&S to suggest experimental designs that do not correct for the biases Chen (2008) identifies. To make our arguments clear, we include brief discussions of the free-choice paradigm, Chen's (2008) criticism of the paradigm, and S&S's critique of Chen.

The free-choice paradigm

The typical free-choice paradigm (based on Brehm, 1956) has participants rank several items, choose between two of them, and re-rank the items. In a host of studies, researchers have found that participants report liking the chosen item more, and the rejected item less, after making a choice. Recently, Egan, Santos, and Bloom (2007) (hereafter ES&B) tested the effect of choice on subsequent choice (rather than on subsequent preference) by modifying Brehm's seminal experiment. In this version, participants rate several items, choose between two items that they rated similarly, and then choose between a third item that was rated similarly and the item that was initially rejected. ES&B (2007) report that when choosing between a new item and one that has just been

DOI of original article: 10.1016/j.jesp.2008.08.027

☆ A response to Sagarin and Skowronski's critique of Chen (2008).

* Corresponding author.
E-mail addresses: keith.chen@yale.edu (M.K. Chen), jane.risen@chicagogsb.edu (J.L. Risen).

0022-1031/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jesp.2008.08.026

rejected, both 4-year-old children and capuchin monkeys choose the novel item over the previously rejected item at a rate greater than 50% (63% and 60%, respectively).

Chen's critique

In his working paper, Chen (2008) argues that both the traditional and modified paradigms fail to control for the information that is revealed by participants' choices. The criticism neither denies the existence of choice-based dissonance nor confirms it. Instead, the argument is simply that neither the traditional nor ES&B's modified versions can properly test for dissonance because both fail to account for revealed preferences. To make his argument, Chen begins with a simple re-analysis of the results found by ES&B. The re-analysis treats participants' choices as informative rather than random. Chen (2008) points out that if participants tend to choose the good they prefer, then the correct null hypothesis for ES&B's experiment is strictly higher than 50%. That is, as long as participants choose the good they prefer at a rate greater than chance, then we learn something about their preferences, and the null hypothesis that ES&B use is no longer appropriate. Chen then makes the assumption that participants nearly always choose the item they prefer and calculates a null of 2/3rds, a number remarkably close to what ES&B find.

Because this simple re-analysis does not address every other prominent papers in the FCP literature, however, Chen (2008) also develops a full Bayesian analysis of the more traditional free-choice experimental design used in such papers as Brehm (1956), Gerard and White (1983), Steele, Spencer, and Lynch (1993), Lieberman, Oschner, Gilbert, and Schacter (2001), Kitayama, Snibbe, Markus, and Suzuki (2004).

Sagarin and Skowronski's critique

The critique that S&S offer specifically addresses Chen's simplified analysis of a single paper by Egan et al. (2007). They do not (as they suggest), challenge Chen's broader analysis of the traditional free-choice paradigm. Specifically, S&S argue that Chen's (2008) analysis of the ES&B paper fails to account for the possibility that agents may not always choose the good they prefer. In his analysis of ES&B, Chen makes the simplifying assumption that if a participant prefers good A to B then, when asked to choose between them, he will choose A with near certainty. S&S suggest that a more realistic analysis might assume that the participant's chooses A only 75% percent of the time. Expressed mathematically, S&S ask the reader to consider Eq. (1):

$$\mathrm{P}(\text{participant chooses good A over B} \mid \mathrm{X}) = 75\% \quad (\mathrm{S\&S}) \tag{1}$$

where X is that the participant prefers A to B.

Note, however, that a virtually identical equation (though this is not acknowledged by S&S) appears in Chen's (2008) general analysis of the traditional free-choice paradigm; there Chen considers Eq. (2):

$$\mathrm{P}(\text{participant chooses good A over B} \mid \mathrm{X}') = \frac{3}{4} \quad (\mathrm{Chen}) \tag{2}$$

where X′ is that a participant has already twice stated they prefer A to B (see cases two and three of the main analysis in Chen (2008)).

In fact, Chen's general formulation (from which Eq. (2) is drawn) encompasses S&S's concern that participants may not always choose the item they prefer (call this "choice error") as well as the possibility that, when asked, participants' stated preferences may not always reflect their true preferences (call this "reporting error").[1]

In other words, a complete reading of Chen (2008) makes it clear that the argument for controlling for revealed preferences holds even if choice error (and reporting error) are assumed. Indeed, S&S's critique is a simplification of the more general framework Chen develops, applied to one of his initial analyses.

S&S draw out the implications of choice errors for Chen's analysis of ES&B, pointing out that the null hypothesis would only be 2/3rds (as Chen computes) if participants always choose the item they prefer (to use their notation, call this a P of 100%). Further, they correctly calculate that for the null to be 60% (one amount ES&B observe), participants would have to choose the good they prefer 89% of the time. S&S argue that from psychometric first-principles, subjects should not be expected to choose goods they prefer as often as 89%, therefore they claim that Chen's objection to ES&B's paper "cannot fully account" for the data reported in ES&B (2007).

Our response

While this computation is numerically correct, we would disagree with the conclusion that S&S draw, and argue their conclusion is unwarranted for two reasons. First, the number that S&S focus on—the P needed to produce a null-hypothesis which exactly matches the experimental results—is misleading and suggests a misunderstanding of the function of a null-hypothesis. S&S are asking the wrong question, calculating the choice–preference correlation that would make the null-hypothesis literally equal to observed behavior. But of course, the correct analysis would be to calculate what would render the observed results statistically indistinguishable from (rather than identical to) experimental results. Although S&S do not provide this analysis, it is possible to do the appropriate calculations using numbers reported in the original ES&B paper; the results are in Table 1.

What these shaded numbers show is the percent of time that participants would need to choose their preferred good (when presented with two options) in order for their expected behavior to be statistically indistinguishable from the behavior ES&B observe. If we take the conventional 5% significance level as our benchmark, the answer is that given any positive relationship between preferences and choice, the predicted result would be indistinguishable from ES&B's observed data for children. For their data on monkeys, a relationship of 78% would make the predicted result indistinguishable from the actual result. Notably, recent experimental evidence suggests that numbers higher than those in the table (even if adjusted to a 10% significance level) are commonly observed in similar experimental settings (see, for example, Savani, Markus, and Conner (2008).

The psychometrics of choice

This brings us to the second reason that we disagree with S&S, which is their interpretation of the empirical psychometric literature on discrete choice, which they claim suggests that a strong correspondence between choice and preference is unlikely (they cite Carroll & de Soete, 1991; Estes, 1984; Navarick & Chellsen, 1983; Williams, 1985). Briefly summarizing the main results of this literature: when two options have roughly equal predicted

[1] Both of these basic issues have long been recognized in the psychometric literature; sophisticated discussions of choice error dates as far back as the early literature on psychophysical discrimination (see Thurstone, 1927) and more formally in the psychometric literature on discrete-choice (see Luce, 1959; Marschak, 1960).

Table 1
Calculation accounting for experimental findings

	Egan et al. (2007), experiment on: 4-year-old children	capuchin monkeys
Observed third-stage choice	63.0%	60.0%
Reported *N*, *t*-value	29 (2.03)	6 (5.48)
Relationship between choice and preference (*P*-value) required to reach		
A 5% significance level	50.0%	77.9%
A 10% significance level	62.2%	80.5%

"values," participants tend to show only a relatively weak and probabilistic preference for the option predicted to be best. So for example, when two levers are equally reinforced (say, both levers produce a food reward 25% of the time), an animal will press each level roughly 50% of the time; if individuals rate two goods similarly on a numeric scale (say pretzels = 2.38 and potato chips = 2.40), then when offered a choice between the two, only slightly more than half will choose the higher-rated good.

S&S take this to suggest that preferences and choices are very loosely linked, and by extension that people's choices should be only minimally informative as to what goods they actually like in situations where external measures of value are roughly equal. This is a mistake. While this sounds compelling at first blush, it conflates two critically different psychometric concepts: the predictive ability of ancillary measures of preferences available to the experimenter, and the predictive ability of a participant's own choices. The fact that external measures may contain very little predictive power for an individual's choice does not imply that choice itself must be only very loosely tied to the preferences the individual has at that moment. For example, even if you know that an individual chooses pretzels and chips a roughly equal number of times across situations, you may still expect that the choice of pretzels or chips in a given situation will provide reliable information about that individual's preference at that moment. This conflation leads S&S to suggest alternative experimental designs which do not address the problems identified in Chen (2008).[2]

We reiterate, however, that even if one were to find these relationship levels of choice and preference implausible, the central point of Chen's critique still applies; when testing for choice-effects it is necessary to begin with an experimental design that explicitly controls for revealed preferences, rather than claiming that a revealed preference explanations are a-priori implausible.

Summary

We hope this comment will help the reader think through and evaluate these interesting questions. In summary, the calculations of S&S are correct; indeed their calculations are structurally identical to several done by Chen in his general analysis (and found in the appendix of Chen (2008)). We disagree that these calculations refute the argument of Chen. Moreover, we contend that S&S's argument both misunderstands the role of a null-hypothesis in scientific tests, and conflates two related but critically different psychometric concepts: the ability of ancillary measures of preferences to predict a participant's choice, and how informative a participant's own choice is, leading S&S to suggest several alternative experimental methods which do not address the fundamental critique of Chen (2008). S&S note at the end of their criticism that data is the only way to determine whether choice truly does reveal important information in the free-choice paradigm and whether it can account for the results that have been found. We are in complete agreement on this point. We have conducted several studies that control for revealed preferences, and we have found results that suggest that this control is in fact necessary (see Chen & Risen, in preparation; Risen & Chen, submitted for publication). We hope to continue the discussion with the inclusion of our empirical efforts.

[2] For example, S&S suggest that to address the critique of Chen, ES&B could measure the latency time of a monkey to retrieve, say, red and blue M&M candies. If monkeys take roughly the same amount of time to retrieve either candy when measured repeatedly across time, S&S suggest that a number of monkeys then be asked to choose between the two, and the proportion choosing the color candy that showed the slightest speed advantage over time "will provide an upper bound of P. From this upper bound, researchers could then calculate the statistical "expected value" for the time 2 choice and compare actual time 2 choice data against this value." (S&S, 2008) But this mistakenly conflates the ability of latency time to accurately predict a subject's choices and the predictive content of a subject's own choices; two values that need not have any relationship. Over time, a monkey can take just as long to pick up red candies as blue, and this need not imply that in any particular trial their choice between red and blue can be treated as random chance.

References

Brehm, J. W. (1956). Postdecision changes in the desirability of alternatives. *Journal of Abnormal and Social Psychology, 52*, 384–389.

Carroll, J. D., & de Soete, G. (1991). Toward a new paradigm for the study of multiattribute choice behavior: Spatial and discrete modeling of pairwise preferences. *American Psychologist, 46*, 342–351.

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.

Chen, M. K., & Risen, J. L. (in preparation). *On the measurement of cognitive dissonance from choice*.

Egan, L. C., Santos, L. R., & Bloom, P. (2007). The origins of cognitive dissonance: Evidence from children and monkeys. *Psychological Science, 18*, 978–983.

Estes, W. K. (1984). Global and local control of choice behavior by cyclically varying outcome probabilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 258–270.

Gerard, H. B., & White, G. L. (1983). Post-decisional reevaluation of choice alternatives. *Personality and Social Psychology Bulletin, 9*, 365–369.

Kitayama, S., Snibbe, A. C., Markus, H. R., & Suzuki, T. (2004). Is there any free choice? Self and dissonance in two cultures. *Psychological Science, 15*, 527–533.

Lieberman, M., Oschner, K., Gilbert, D., & Schacter, D. (2001). Do amnesiacs exhibit cognitive dissonance reduction? The role of explicit memory and attention in attitude change. *Psychological Science, 12*, 135–140.

Luce, R. Duncan (1959). *Individual choice behavior*. New York: Wiley.

Marschak, Jacob (1960). Binary choice constraints on random utility indicators. In Kenneth Arro (Ed.), *Stanford symposium on mathematical methods in the social sciences*. Stanford, CA: Stanford University Press.

Navarick, D. J., & Chellsen, J. (1983). Matching versus undermatching in the choice behavior of humans. *Behaviour Analysis Letters, 3*, 325–335.

Risen, J. L., & Chen, M. K. (submitted for publication) *How choice affects and reflects preferences: Revisiting the free-choice paradigm*..

Sagarin, B. J., & Skowronski, J. J. (2008). The implications of imperfect measurement for free-choice carry-over effects: Reply to M. Keith Chen (2008) Rationalization and cognitive dissonance: Do choices affect or reflect preferences? *Journal of Experimental Social Psychology*. doi:10.1016/j.jesp.2008.08.024.

Savani, K., Markus, H. R., & Conner, A. L. (2008). Let your preference be your guide? Preference and choices are more tightly linked for North Americans than for Indians. *Journal of Personality and Social Psychology, 95*(4), 861–876.

Steele, C. M., Spencer, S. J., & Lynch, M. (1993). Self-image resilience and dissonance: The role of affirmational resources. *Journal of Personality and Social Psychology, 64*, 885–896.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review, 34*, 273–286.

Williams, B. A. (1985). Choice behavior in a discrete-trial concurrent VI-VR: A test of maximizing theories of matching. *Learning and Motivation, 16*, 423–443.