

# The Great Between-Subjects Assumption

Charles Lambdin\* & Victoria Shaffer

*Department of Psychology, Fairmount College of Liberal Arts and Sciences*

There seems to be an assumption that some judgment and decision-making (JDM) findings will only be supported in between-subjects designs, even though between-subjects results can be misleading. This paper tested this assumption by taking two famous between-subjects JDM findings (Shafir, 1993; Tversky & Kahneman, 1986) and replicating them within subjects. Replicating within subjects allowed us to conduct a manipulation check, and the qualitative data gathered surprisingly suggest the original experiments lack construct validity

## Introduction

Results sometimes disappear or even reverse themselves when switching from a between- to a within-subjects design and vice versa [1]. Because of this it is sometimes important to treat design type as an explicit independent variable [2]. For instance, Jones and Aronson's famous "Just World Hypothesis" finding — that respectable women are judged to be at greater fault at causing their own rape than less-respectable women — can be reversed by switching from a within- to a between-subjects design [1]. Birnbaum argued this reversal occurs because in the between-subjects design the stimulus is confounded with the context, adding that "It may be that many of the counterintuitive findings so well-liked in social psychology are merely results of the *confounding of contexts* that can occur in between-subjects designs" [1].

In the field of judgment and decision making (JDM) some researchers seem to assume that within-subjects designs should be avoided due to increased task transparency, even though between-subjects designs take longer, require more subjects, decrease power and contain unwanted context effects [1, 3]. The present experiment took two standard JDM between-subjects experiments [4, 5] and replicated them using a within-subjects design.

## Experiment

As of yet, 77 undergraduate Wichita State University students have volunteered. Students received extra credit for their participation in our study. Participants completed a computerized questionnaire which was administered using Medialab. Participants answered the questions from each experiment [4, 5], back-to-back, within subjects. After each replication participants answered questions about whether anything struck them as odd or unusual about their answers, thoughts or impulses.

## Results

Shafir [5] had participants compare two parents (*A* and *B*) in a hypothetical custody case. He hypothesized that since one of the parents in the hypothetical custody case, Parent *B*, had both more extreme positive and negative attributes, that according to the theory of reason-based choice most participants would both award and deny custody to Parent *B*. Shafir found (between subjects): Parent *B*:  $P_c + P_r = 119$ ; Parent *A*:  $P_c + P_r = 81$  ( $P_c$  and  $P_r$  represent the percentage of participants who chose and denied an option). Shafir thus assumed responses should be complementary, an assumption that only makes sense in a within-subjects design. We found: ( $N = 77$ ) Parent *A*:  $P_c + P_r = 104$ ; Parent *B*:  $P_c + P_r = 96$ .

Tversky and Kahneman [4] had participants read the following hypothetical scenario: Imagine that the United States is preparing for the outbreak of an unusual Asian disease that is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimates of the consequences of the program are as follows: If **Program A** is adopted, 200 people will be saved. If **Program B** is adopted, there is a one-third probability that 600 people will be saved and a two-thirds

probability that no people will be saved. Participants in one group chose from among these options and participants in another from these: If **Program C** is adopted, 400 people will certainly die. If **Program D** is adopted, there is a one-third probability that no one will die and a two-thirds probability that 600 people will die. We presented these options within subjects. Results for both are shown in Figure 1.

**Discussion**

In our replication of Shafir [5], 25% of participants awarded and denied custody to the same parent. Participants were asked to type anything that struck them as odd about their responses or impulses. Most responses were rationalizations: e.g., “I like to go for the underdog.” Only 9% of comments betrayed an understanding of the manipulation. Further, participants did not seem to agree with the experimenter assumption that Parent *B* had both more negative *and positive* attributes. Without being asked to comment on it, 19% of participants explicitly stated Parent *B* is an unfit parent simply because s/he has an “extremely active social life.” This perhaps suggests the original experiment may lack construct validity, and participants do not seem to interpret the stimulus materials in the manner hypothesized (Parent *B* having both more negative AND positive attributes).

Tversky & Kahneman (1986): Even though they were not explicitly asked for this information, about 10% of participants indicated they did not see the “Programs” as complementary. This is actually a valid interpretation! Given this interpretation, “400 people will certainly die” does *not* imply that “200 will certainly live,” but rather, “one can be certain at least 400 out of 600 will die, maybe 450, maybe 575 will die.” “Two hundred will live” was taken by some to mean “with Program *A* at least 200 will live.” This does not actually tell one how many will die! Thus with Program *D* one does not know if anyone will die. Given this interpretation, it is not only rational to choose *A* and *D* but it also does not constitute a framing effect in any way, shape or form! One hundred percent of participants who “switched” answers chose programs *A* and *D*; *nobody* chose *B* and *C*! This again suggests the original experiment may lack construct validity.

**Conclusion**

These results also suggest the “between-subjects assumption” was in these cases unwarranted; e.g., one would think that when presenting both Parents *A* and *B* to the same participants at the same time that no one would both award and deny custody to the same person! And yet, within subjects, 25% of participants did just that! This indicates that even with such a clear-cut, either-or, accept/reject decision presented within subjects, many participants do not realize their choices should be complementary! Thus, within subjects replication clearly did not make the design transparent.

[1] M. Birnbaum, Controversies in Psychological Measurement, Social Attitudes and Psychological Measurement (Ed. B. Wegener), pp. 401-485. 1982.  
 [2] G. Keren, Between or Within-Subjects Designs: A Methodological Dilemma, A Handbook for Data Analysis in the Behavioral Sciences (Eds. G. Keren & C. Lewis), pp. 257-272. 1993.  
 [3] M. Birnbaum & B. Mellers, Bayesian Inference: Combining Base Rates with Opinions of Sources Who Vary in Credibility, Journal of Personality and Social Psychology, 45, pp. 792-804. 1983.  
 [4] A. Tversky & D. Kahneman, Rational Choices and the Framing of Decisions, Journal of Business, 59, pp. 251-278. 1986.  
 [5] E. Shafir, Choosing Versus Rejecting: Why Some Options Are Both Better and Worse than Others, Memory and Cognition, 21, pp. 546-556. 1993.

Figure 1

Our Results		Tversky & Kahneman’s Results	
Program	% Chose	Program	% Chose
A	77%	A	72%
B	23%	B	28%
C	40%	C	22%
D	60%	D	78%