
ELSEVIER

Journal of Economic Psychology 22 (2001) 91–101

JOURNAL OF Economic Psychology

www.elsevier.com/locate/joep

Evaluating framing effects

James N. Druckman *

Department of Political Science, University of Minnesota, 1414 Social Sciences Building, 267-19th Avenue South, Minneapolis, MN 55455-0410, USA

Received 25 July 2000; accepted 10 October 2000

Abstract

This paper examines Tversky and Kahneman's well-known Asian disease framing problem (A. Tversky, D. Kahneman, Science 211 (1981) 453–458). I describe an experiment where respondents received a version of the disease problem using a survival format, a mortality format, or both formats. The results from the survival and mortality formats replicate Tversky and Kahneman's original experiment both in terms of statistical significance and, in contrast to some other studies, in terms of magnitude. I then argue that the "both format" condition constitutes an important and previously unused baseline for evaluating the strength of framing effects. This standard of comparison provides a way to evaluate the impact of a frame on unadulterated preferences – that is, preferences unaffected by a particular frame. The implications for future framing effect experiments are discussed.

PsycINFO classification: 2260; 2340

JEL classification: C91; D81

Keywords: Framing effect

* Corresponding author. Tel.: +1-612-626-1672; fax: +1-612-626-7599.
E-mail address: druckman@polisci.umn.edu (J.N. Druckman).

0167-4870/01/$ - see front matter
PII: S0167-4870(00)00032-5

1. Introduction

In their famous experiment, Tversky and Kahneman (1981) provided participants with the following scenario:

> Imagine that the US is preparing for the outbreak of an unusual Asian disease, which 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 programs are as follows...

Some respondents ($N = 152$) then received a description of the programs using a survival format:

> If Program *A* is adopted, 200 people will be saved.
>
> If Program *B* is adopted, there is a 1/3 probability that 600 people will be saved, and a 2/3 probability that no people will be saved.

Of these respondents, 72% expressed a preference for Program *A* – the risk-averse alternative. Another group of respondents ($N = 155$) received a description of the same exact programs except it used a mortality format:

> If Program *C* is adopted, 400 people will die.
>
> If Program *D* is adopted, there is a 1/3 probability that nobody will die, and a 2/3 probability that 600 people will die.

Only 22% of these respondents opted for Program *C*, even though it is the same risk-averse alternative as Program *A*, and Program *D* is the same risk-seeking alternative as Program *B*.

This result reflects a general tendency for people to be risk-averse when exposed to a gains (survival) format and risk-seeking when exposed to a losses (mortality) format. Traditionally, such a shift in risk preferences due to alternative descriptions of the problem is called a "framing effect" (see Levin, Schneider, & Gaeth, 1998 for distinctions among types of framing effects). Tversky and Kahneman's work on framing effects – and in particular their Asian disease problem – has spawned numerous literatures throughout

the social sciences (e.g., Zaller, 1992; Camerer, 1995; Levin et al., 1998; Li, 1998).

In this paper, I build on these literatures by addressing two issues. First, I present an experiment designed to replicate Tversky and Kahneman's original Asian disease experiment (just described). This is interesting because several other replication attempts have produced a framing effect smaller than the one first documented by Tversky and Kahneman. I discuss a possible explanation for these variations in the magnitude of framing effects. Second, I offer a new baseline for evaluating the existence and/or strength of framing effects. Specifically, I compare respondent's preferences from the survival and mortality formats (i.e., conditions mimicking the original experiment) with preferences from a (newly included) condition where respondents received both formats simultaneously. The advantage of this standard of comparison, relative to other standards that I will discuss, is that it serves as an *empirically* derived baseline that may approximate preferences in the *absence* of an effect from either the survival or mortality format. The experimental results reveal that, contrary to common assertions (e.g., Kühberger, 1998, 235), the mortality format may not have a stronger effect on preferences than the survival format. Before turning to the experiment, I offer a more detailed discussion of why a replication is important and why a new standard of evaluation is useful.

1.1. Replication

Scholars attempting to replicate Tversky and Kahneman's (1981) Asian disease experiment have had mixed success. Some replication attempts have failed to find a significant framing effect (e.g., Miller & Fagley, 1991), while others have successfully mimicked the original experiment (e.g., Bless, Betsch, & Franzen, 1998). More generally, it seems that while the effect is highly reliable (Kühberger, 1998), the magnitude of the effect tends to be smaller than that found in the original experiment (e.g., Fagley & Miller, 1990; Bohm & Lind, 1992; Kühberger, 1995, 235; see Levin et al., 1998 for an overview). Fagley and Miller (1997, 357) explain that "Framing results have been mixed, particularly if one looks at the magnitude of the effect..." In what follows, I present an experiment that attempts to replicate the original results; as I will discuss, my replication attempt differs from many others in that I use the original problem without alteration and ask subjects to reply only to the original problem (and no other problems).

1.2. Evaluating framing effects

Two approaches have been taken to evaluate the existence and/or strength of framing effects (Kühberger, 1995, p. 235; Kühberger, 1998, pp. 30–31; Wang, 1996; Levin et al., 1998, p. 153). The first approach is sometimes called a unidirectional effect or a choice shift. This approach entails comparing the percentage of participants who opted for the risk-averse alternative (or the risk-seeking alternative) in the gains (survival) format with the percentage of participants who opted for the risk-averse alternative (or the risk-seeking alternative) in the losses (mortality) format. For example, using Tversky and Kahneman's original data, 72% opted for the risk-averse alternative in the gains format and 22% opted for it in the losses format. These percentages significantly differ from one another, so one would conclude that there is a "framing effect" using this standard of comparison. This evaluative approach reveals just how different preferences can be when one frame is used instead of another (objectively identical) frame. Moreover, by using frames that push preferences in opposite directions (e.g., towards more risk-averse behavior versus more risk-seeking behavior), this approach provides insight into the maximal power of framing.

The other approach – sometimes called a bidirectional effect or a choice reversal – entails the investigation of if (1) risk-averse choices predominant for the gains format and (2) risk-seeking choices predominant for the losses format (Wang, 1996). This requires that significantly greater than (a risk-neutral) 50% of respondents opt for the risk-averse alternative under the gains format and significantly fewer than 50% of respondents opt for the risk-averse alternative under the losses format. Using Tversky and Kahneman's original data again, 72% is significantly greater than 50% and 22% is significantly below it, and thus, one would conclude that a framing effect occurred using this standard of comparison as well. This bidirectional approach has the advantage of revealing if one frame causes a majority of subjects to opt for one alternative (e.g., the risk-averse alternative) while the other frame causes the majority to opt for the other alternative (e.g., the risk-seeking alternative). Identifying such a majority reversal in preference can be important for those interested in if different frames can generate majority support for one alternative instead of another.

Wang (1996, pp. 154–155) offers another justification for the bidirectional approach, arguing that only when the formats push preferences significantly away from 50% (in the predicted directions) does the effect

violate the description invariance assumption underlying expected utility theory. That is, the assumption that "different representations of the same choice problem should yield the same preference" (Tversky & Kahneman, 1987, p. 69). The status of this justification is unclear, however, because even when a unidirectional effect occurs, it still suggests that some percentage of the respondents violate the preference invariance assumption (even if it does not suggest that a majority violate the assumption). Put in another way, evidence of violations of preference invariance among some respondents need not require a bidirectional effect. Indeed, Tversky and Kahneman (1987, pp. 70–71) themselves cite a framing experiment that results in significant difference between formats but not significant differences from 50% (in the predicted directions) as a prime example of an invariance violation.

While Tversky and Kahneman's original result is significant using either of the two evaluative approaches, it is easy to find examples where the two approaches lead to very different conclusions (Kühberger, 1998, pp. 30–31). For example, imagine that 35% opted for the risk-averse alternative in the gains format and 5% opted for it in the losses format. Then, the 50% baseline approach (or the bidirectional approach) implies either no framing effect (because 35% does not exceed 50%) or a reverse framing effect (because 35% may be significantly below 50%). In contrast, assuming a sufficiently large N, 35% and 5% might well be significantly different from one another in the expected (relative) direction; thus, one would conclude that there is a framing effect using the unidirectional approach.

Apart from evaluating if framing effects occurred, these approaches also have implications for the strength of different effects. Most notably, those who use the 50% baseline often conclude that the losses format is stronger than the gains format because it results in a larger deviation from 50% – for example, 22% is further from 50% than is 72% (see, e.g., Kühberger, 1995, p. 235).

In any given study, both approaches for evaluating framing effects could be used, as each provides distinct insights – the unidirectional approach reveals the relative impact of alternative frames on preferences while the bidirectional approach reveals if a majority preference reversal occurred. However, neither approach provides insight into another question: what is the impact of a given frame on unadulterated preferences – that is, preferences unaffected by a particular frame? In many real-world circumstances, individuals receive a single frame that may or may not move their prior (unadulterated) preference (e.g., they receive a single frame from a strategic

politician or salesperson); it is thus important to somehow evaluate the impact of frames on unadulterated preferences.

The unidirectional approach fails to capture this process because it compares preferences shaped by one frame with preferences shaped by another frame. One could argue that bidirectional approach applies in that it reveals how far a single frame can move preferences from a risk-neutral 50%. (Since the programs have the same expected values, if decision-makers are risk-neutral, they will be indifferent and randomly choose each program with equal 50% probability.) However, there is no reason to *assume* that respondents have risk-neutral preferences before being exposed to the problem. Indeed, it may be the case that in the *absence of influence from the gains or losses format*, people would exhibit risk-seeking or risk-averse behavior (and not risk-neutral behavior). For example, imagine that 15% (instead of 50%) opted for the risk-averse alternative in the absence of influence from a losses (or gains) format. In this case, if 45% opted for the risk-averse alternative when given the gains format, then, assuming a large N, we would conclude that the frame did in fact significantly affect unadulterated preferences. We would not reach this conclusion using the 50% baseline.

Evaluating the effect of a frame on unadulterated preferences thus requires the measurement of prior risk preferences. A general risk-taking propensity measure may be inappropriate, however, since risk attitudes tend to vary across specific problems (e.g., Wang, 1996). In the experiment that follows, I attempt to gauge prior risk attitudes for this problem by incorporating a new condition that includes *both* the gains (survival) and losses (mortality) information. The idea is that by exposing respondents to both formats, the impact of each of the formats on risk attitudes will cancel out. This will serve as an appropriate *empirically derived* standard to evaluate the impact of a frame on unadulterated preferences. Note that using both formats in this way differs from Kühberger (1995) in that it does not separate out domain effects from framing effects because the goal is to have the domains cancel each other out.

It is of course possible that when presented together, one of the formats will be more salient and push preferences in a certain direction; however, previous work on a different problem by Levin, Johnson, and Davis (1987, p. 52) suggests that a format with both types of data (e.g., gains and losses data) is "neutral in evaluation". In some sense, this new survival–mortality format serves as the control group that has not been included in prior Asian disease experiments.

2. Method

I randomly assigned 320 student participants (from a large US public university) to one of three conditions. The conditions asked participants to respond to a version of the Asian disease problem that used either (1) the survival format ($N = 69$), (2) the mortality format ($N = 79$), or (3) a format that included both survival and mortality data ($N = 172$). (These three variations of the problem constitute the three conditions.) Using students in the experiment should not limit its generalizability. Indeed, based on a meta-analysis of 136 framing effect studies with nearly 30,000 participants, Kühberger (1998, p. 36) finds that the behavior of student participants does not significantly differ from the behavior of non-student participants.

The survival and mortality formats were as in the original experiment, described above. An example of the "both" survival-mortality format is:

> If Program *A* is adopted, 200 people will be saved and 400 people will die.
>
> If Program *B* is adopted, there is a 1/3 probability that 600 people will be saved and nobody will die, and a 2/3 probability that no people will be saved and 600 people will die.

The relatively large number of participants exposed to the survival–morality format is due to exposing them (randomly) to one of four versions of this format, where the order of the programs and outcome statistics varied. In presenting the results, I merge these four versions because no significant differences between them were found. Participants were debriefed at the conclusion of the experiment.

3. Results

Of the participants who received the survival format, 68.1% opted for Program *A* – the risk-averse alternative. In contrast, only 22.8% of the participants who received the mortality format opted for Program *C* – the identical risk-averse alternative. To test for a statistically significant difference between the two formats, I use a difference of proportions test that is

analogous to a difference-of-means test except that it applies to the case of two proportions from two samples or conditions (Blalock, 1979, pp. 232–233). The difference of proportions test indicates that the difference between the two formats is highly significant ($z = 5.54$, $P < 0.01$). Moreover, both the survival format and the mortality format results significantly differ from 50% in the expected directions ($\lambda_1^2 = 9.06$, $P < 0.01$; $\lambda_1^2 = 23.41$, $P < 0.01$, respectively).

This result is quite close to a replication of Tversky and Kahneman's original experiment, both in terms of significance and magnitude. Also, in contrast to Fagley and Miller (1990, 1997) (but consistent with Miller and Fagley, 1991, p. 520), women were no more susceptible to framing effects than men. Specifically, when given the survival format, 61% of women and 74% of men opted for the risk-averse alternative; when given the mortality format, 21% of women and 26% of men opted for the risk-averse alternative.

Of the participants who received a format with both survival and morality data, 43.6% preferred the risk-averse alternative. This is approximately midway between the survival format result (68.1%) and the mortality format result (22.8%). It also is marginally statistically significantly different from a risk-neutral 50% ($\lambda_1^2 = 2.81, P < 0.10$). I next discuss the implications of these results.

4. Discussion

The successful replication is consistent with several studies that also find a statistically significant framing effect. However, unlike some of these studies, the magnitude of the framing effect reported here approaches the magnitude found in the original experiment. A possible explanation for this finding is that, like Bless et al.'s (1998) similar replication, participants were exposed to only one problem and the original problem was used without alteration (also see Takemura, 1994). In contrast, several of the replication attempts that produce significant but smaller framing effects expose participants to several problems and/or slightly alter the original problem (Miller and Fagley, 1991; Bohm and Lind, 1992; Kühberger, 1995; Wang, 1996; Zickar and Highhouse, 1998). Indeed, Bless et al. (1998) demonstrate that framing effects can be quite sensitive to slight contextual changes.

The format with both mortality and survival data provides a new baseline for evaluating the impact of the frames. If the survival and mortality formats neutralize one another, making this format an unadulterated

measure of prior risk preferences, then the result of 43.6% suggests that participants had a slight (albeit marginally significant) tendency towards risk-seeking, in the absence of influence from a format. Thus, if the survival format alone had pushed participants significantly away from 43.6%, then it would have had an important effect (on initial preferences), even if it did not significantly differ from 50%. While 43.6% is only marginally significantly different from 50%, it still highlights the potential problem of assuming a risk-neutral 50% as an indication of unadulterated preferences. In short, there is no reason to assume that the respondents are risk-neutral; a better approach is to measure prior preferences so as to evaluate the impact of a frame on unadulterated preferences. In many contexts, individuals receive just one frame (e.g., from a politician or salesperson) and thus understanding the effect of a single frame on prior, unadulterated preferences is significant.

Notice also that if we treat 43.6% as an empirically derived baseline, then neither the survival nor the mortality format was more influential (since 43.6% approximates the mid-point between these singular formats). In contrast, using a 50% baseline figure suggests that, consistent with prior interpretations, the mortality format had a stronger influence than the survival format because it deviates more from the risk-neutral 50% baseline (i.e., 27.2% versus 18.1%).

As mentioned, another explanation for the result from the format with both mortality and survival data is that the participants actually were risk-neutral (or even risk-averse) ex ante and, when given the morality–survival format, the mortality description was more salient than the survival description (and consequently it pushed participants in a risk-seeking direction, albeit to a lesser extent than the mortality format alone). Irrespective of whether this is true, the results presented here suggest an important amendment to the way in which we evaluate framing effects. As discussed, the unidirectional and bidirectional approaches have their advantages; however, neither provides an estimate of the impact of a single frame on unadulterated preferences. In contrast, the empirically derived baseline used here may provide such an estimate. In future framing experiments, all three approaches (or other analogous alternatives) could be used so that we can estimate the impact of the frames relative to one another, evaluate majority preference reversals, and judge the impact of each frame on unadulterated preferences. Framing effects clearly play a substantial role in the decision-making process, and thus, we should evaluate them from each of these varying perspectives.

Acknowledgements

I thank Gregory Bovitz, Daniel Druckman, Nicole Druckman, and the anonymous reviewers for helpful advice.

References

Blalock, H. M. (1979). *Social statistics*. New York: McGraw-Hill.

Bless, H., Betsch, T., & Franzen, A. (1998). Framing the framing effect: the impact of context cues on solutions to the 'Asian disease' problem. *European Journal of Social Psychology*, *28*, 287–291.

Bohm, P., & Lind, H. (1992). A note on the robustness of a classical framing result. *Journal of Economic Psychology*, *13*, 355–361.

Camerer, C. (1995). Individual decision making. In J. H. Kagel, & A. E. Roth (Eds.), *The handbook of experimental economics* (pp. 587–703). Princeton, NJ: Princeton University Press.

Fagley, N. S., & Miller, P. M. (1990). The effect of framing on choice: interactions with risk-taking propensity, cognitive style, and sex. *Personality and Social Psychology Bulletin*, *16*, 496–510.

Fagley, N. S., & Miller, P. M. (1997). Framing effects and arenas of choice: your money or your life? *Organizational Behavior and Human Decision Processes*, *71*, 355–373.

Kühberger, A. (1995). The framing of decisions: a new look at old problems. *Organizational Behavior and Human Decision Processes*, *62*, 230–240.

Kühberger, A. (1998). The influence of framing on risky decisions: a meta-analysis. *Organizational Behavior and Human Decision Processes*, *75*, 23–55.

Levin, I. P., Johnson, R. D., & Davis, M. L. (1987). How information frame influences risky decisions: between-subjects and within-subject comparisons. *Journal of Economic Psychology*, *8*, 43–54.

Levin, I. P., Schneider, S. L., & Gaeth, G. J. (1998). All frames are not created equal: a typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Making Processes*, *76*, 149–188.

Li, S. (1998). Can the conditions governing the framing effect be determined? *Journal of Economic Psychology*, *19*, 133–153.

Miller, P. M., & Fagley, N. S. (1991). The effects of framing, problem variations, and providing rationale on choice. *Personality and Social Psychology Bulletin*, *17*, 517–522.

Takemura, K. (1994). Influence of elaboration on the framing of decision. *The Journal of Psychology*, *128*, 33–39.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*, 453–458.

Tversky, A., & Kahneman, D. (1987). Rational choice and the framing of decisions. In R. M. Hogarth, & M. W. Reder (Eds.), *Rational choice: the contrast between economics and psychology*. Chicago: University of Chicago Press.

Wang, X. T. (1996). Framing effects: dynamics and task domains. *Organizational Behavior and Human Decision Processes*, *68*, 145–157.

Zickar, M. J., & Highhouse, S. (1998). Looking closer at the effects of framing on risky choice: an item response theory analysis. *Organizational Behavior and Human Decision Processes*, *75*, 75–91.

Zaller, J. (1992). *The nature and origins of mass opinion*. New York: Cambridge University Press.