Overcoming Frames and Ethical Dilemmas:

The Effects of Large Numbers on Decision-Making

Alexander Johnson, Bryce Jones, Andrew Kent, Alex Polta

University of Missouri

In partial fulfillment of the requirements for Psych 4003

Dr. Ines Segert

May 11, 2007

Abstract

Bernoulli's Law of Large Numbers renders single-event probabilities nearly meaningless. This study took two previous studies on decision-making, both of which found that participants violated rational-decision-making norms, and both of which neglected the Law of Large Numbers, and applied the Law of Large Numbers to the questions used in those studies to see if participants would respond more rationally if questions were presented in a more mathematically accurate format. This study found that participants are almost perfectly rational when the Law is applied to word problems that do not require much mathematical calculation, but did not do consistently better on problems which involved more than one step of mathematical calculation. Though the split within the results makes clear conclusions difficult, the complete elimination of irrationality in participants under certain conditions where irrationality had previously been shown should be significant enough to merit further research in the area.

Overcoming Frames and Ethical Dilemmas:

The Effects of Large Numbers on Decision-Making

In many studies on decision theory, the conclusions have been that humans make decisions based on heuristics or emotions, rather than reasoning, and thus often come to false conclusions or make inconsistent decisions. (Kahneman & Tversky, 1983; Greene et al., 2001) Kahneman and Tversky's 1983 study did show that when faced with risky decisions, people, in general, are risk-averse in cases of possible gain, risk-seeking in cases of possible loss (risk-intuitionism). This is puzzling, and seemingly irrational, behavior.

Citing Daniel Bernoulli's argument for the *subjective value*[i] rather than *expected value*[ii] of outcomes[iii], Kahneman & Tversky (1983) expanded on this theory to create a descriptive heuristic for risky decisions: individual's choices in cases of uncertainty do not employ the objective pragmatism of the Neumann and Morgenstern (1947) decision heuristic, specifically their *invariance* axiom, which states that "the preference order between prospects should not depend on the manner in which they are described.[iv]" Rather, Kahneman & Tversky argue, choices are determined by frames, i.e., preference order between prospects *does* depend on the manner in which they are described. Kahneman & Tversky hypothesized that subjects will continue to choose the non-risk option in cases of possible gain, the riskier one in cases of loss, even when these options are equivalent or worse, simply because of the framing of the state of affairs[v]. They gave participants questions involving expected value and utility calculations. Kahneman & Tversky's findings supported their theory; subjects were far more affected by framing than by actual expected value, that is, framing led subjects to make different choices on problems with equivalent expected value. Further, in questions which were framed so that one option dominated[vi], but violated the risk-seeking or adverse condition, subjects consistently chose the option that was clearly dominated by another (that is, the irrational but risk-intuition satisfying option). According to Kahneman, "We conclude that frame invariance cannot be expected to hold and that a sense of confidence in a particular choice does not ensure that the same choice would be made in another frame." Kahneman & Tversky showed that the normative standard of invariance set out by Neumann and Morgenstern is not descriptive of the decisions that humans actually make.

Another Study by Greene, Sommerville, Nystrom, Darley, and Cohen (2001) used an fMRI to show that when given moral dilemmas, parts of the brain responsible for emotion and parts of the brain responsible for reasoning are used, and the proportion of emotional response to rational response within the brain is determinate of the choice made by the participant. They argued that this reasoning-emotion tension explains why so many people make irrational[vii] choices in these dilemmas.

Our study proposes that Kahneman & Tversky's conclusion was incorrect. Many of the questions given in Kahneman & Tversky's study reguired calculations involving probabilities (as is necessary in any expected value calculation), but ignored a fundamental concept in statistics, discovered by James Bernoulli, called The Law of Large Numbers:

> The Law of Large Numbers says that in repeated, independent trials with the same probability *p* of success in each trial, the chance that the percentage of successes differs from the probability *p* by more than a fixed positive amount, $e > 0$, converges to zero as the number of trials *n* goes to infinity, for every positive *e*. [viii]

The implication of the Law of Large Numbers for Kahneman & Tversky's study is that for a very small number of trials, e.g. 1, as in Kahneman's study, "the chance that the percentage of successes differs from the probability *p*" is very high, rendering probability statements almost meaningless. To put this another way, a coin weighted to fall heads 60 percent of the time would likely fall heads about 600 out of 1000 throws, give or take a few tails (~60% heads), but it would not be surprising if in one toss or a few tosses it only landed tails (0% heads). Given this, expected values will not or may not be consistently correct on single trials, and the Kahneman study is based around single trial expected value calculations. Thus, Kahneman may have misattributed the findings of his study, because the subjects in Kahneman's study may not have been irrational, but rather, had an understanding of or intuitive sense about the Law of Large Numbers. Also, though the Greene, et al (2001) study does not involve explicit expected value calculations, it could easily be conceived of as an implicit task in the test; for example, the probability of someone from one group of possible victims in the dilemma being a mother or child or Gandhi, or the probability of being punished for your actions. These sorts of "random" probabilistic occurrences with small numbers undermine the capacity for rational decision-making, while large numbers normalize this: if each group is exceedingly large, if there's a mother or child or Gandhi in one group of possible victims, there is likely to be a mother or child or similar peacemaker in the other. Though the expected value calculations in Greene et al's study are not explicit and can be calculated many different ways, this does not undermine the fact that with small enough numbers, knowing how to correctly understand or calculate the expected value of each action and thus act consistently is impossible.

We reasoned that if given a sample size large enough to properly reflect the expected value of the options provided, participants would exhibit invariance in their answers to the questions. If so, humans are more "rational" in decision making than current research would suggest.

Method

Participants

149 participants volunteered from a university psychology class. Participants were presumably between the ages of 18-24, as daytime classes at this university generally do not have students outside of this age range. Demographic information was not obtained for the participants, but the university student body as a whole is 84% Caucasian, 2% Hispanic, 3% Asian, 6% African American, 1% Native American, and 1% international students,[1] and participants likely reflected this. All participants were given consent forms that were signed and collected. Participants were offered five bonus points in the class for their completion of the surveys, as well as a false lure of three extra bonus points if they answered a majority of questionnaire questions correctly.

Apparatus

Participants were given paper questionnaires that had 11 questions.

Design

To see if the Law of Large Numbers had an effect on decision-making in our participants, seven different questionnaires were used. The control group, "A", was a set of questions from the Kahneman & Tversky (1983) and Greene, Sommerville, Nystrom, Darley, & Cohen (2003) studies. Groups B,C,D,E,F,G were composed of revised versions of the questions in A, in which the number of trials (questions 1-9, based on Kahneman & Tversky,) or the number of victims (questions 10-11, based on Greene et al) were manipulated, while the formal structure of the question remained unchanged. B,C and D were mathematically significant numbers: x10, x100, x1,000,000, respectively. E,F, and G, were supposed anthropologically significant numbers: x30, x300, x300,000, respectively. Participants were assigned to groups A-G randomly.

[1]http://www.princetonreview.com/college/research/profiles/studentbody.asp?listing=1022672&ltid=1&intbucketid=

For question sets 1&2, 5&6, 8&9, and 10&11[2], there was not a correct answer for each individual question. Instead, these questions tested for invariance: though there is no “correct” answer to each question, rational agents should choose consistently within the question sets (see discussion of invariance in introduction.) For an example of the survey see the appendix.

Procedure

Upon entering the class, the researchers were introduced. Consent forms were handed out, signed by the participants, and collected. Participants were informed of the bonus point structure mentioned in the *Subjects* section and told that they must get six of the 11 questions correct to receive the extra points. All participants were given all of the bonus points; the lure of extra bonus points merely served as motivation for the participants. Questionnaires were handed out randomly to participants. When finished, participants placed questionnaires facedown on a table in the front of the class and were offered debriefing forms. For a few participants, there were not enough debriefing forms, and these participants were given oral debriefings.

Results

Questions generally fell into three types: explicit probability problems involving multi-step calculations (questions 1-6), subjective utility problems (8-9), and ethical dilemmas (10-11). The latter two types required no math beyond single-step addition or subtraction, while the first six questions required calculating expected value, which requires not only a fair amount of “mental math”, but also a background knowledge of how to calculate expected value at all. As to how participants ought respond, as mentioned in the *Design* section, some questions tested for variance (making consistent choices among changing frames) between pairs of questions (1&2, 5&6, 8&9, and 10&11), while other questions were not set up as pairs and instead required

[2] Questions 10 and 11 arguably have a correct answer, and pheasibly could have different correct answers, but Greene et al assumed a single “rational” response for each, so that same paradigm was used for this study. See endnote vii

expected or subjective value calculations (3, 4, 7). Among the questions in which participants had to make consistent choices against changing frames (fig. 1), and calculate expected value using probabilities (Fig. 5), as a whole the test groups (B-G) did not show consistent improvement over the control group (A) (Fig. 2). However, on question set 8-9, which tests for framing without the probability calculations necessary in 1-6, the improvement of the test groups over the control group was stark (fig. 3). Also, the test group showed significant improvement on consistency between their answers in questions 10 and 11, and made more “rational” choices on these two questions individually. This distinction, between questions 1-6, the “former” questions, which involve multi-step mathematical calculations; and 8-11, the “latter” questions, which involve little or no math, is essential for understanding the results. When the latter questions are separated from the former, test groups show no real improvement on the former questions, sometimes doing much worse but with a slight trend to do better as numbers get larger (Figs. 3, 5). On the latter questions, however, the test groups did not just improve from the control group; in all but a few groups they performed perfectly, totally overcoming variance.

Question 7 was most likely useless because in its test format it became implausible (e.g., on Form D, it provided the scenario of a “dilemma” in which the participant had to decide whether to drive 20 minutes to save $5,000,000). Its inclusion or exclusion, does no effect the meaning or significance of the data.

Fig. 1: Percent variance between equivalent questions
(x-axis= test form, y-axis=percent variance)

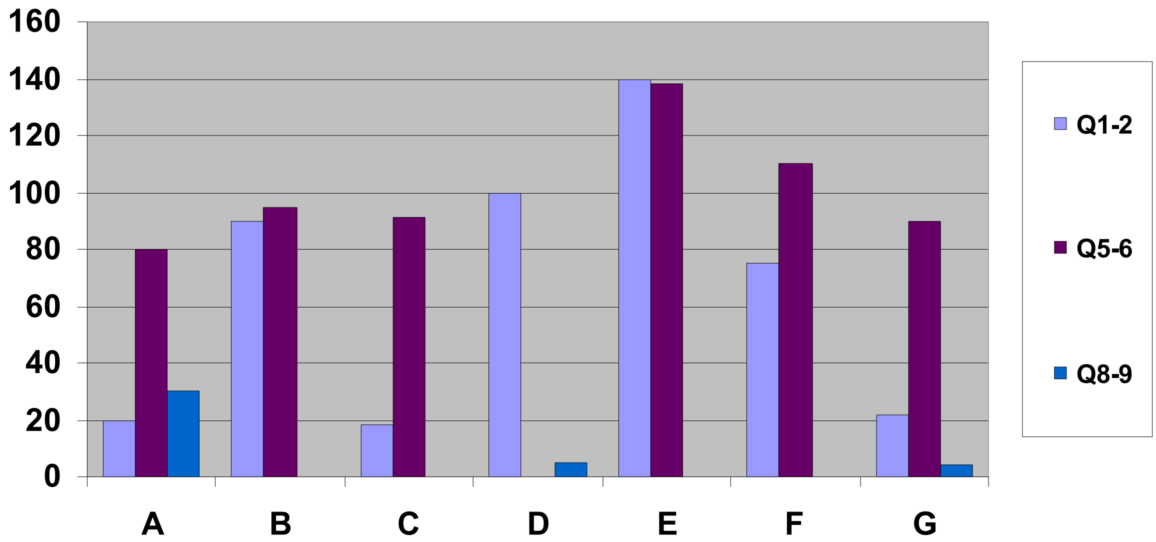


Fig. 2 Average Percent variance by Test Form
(x-axis= test form, y-axis=percent variance)

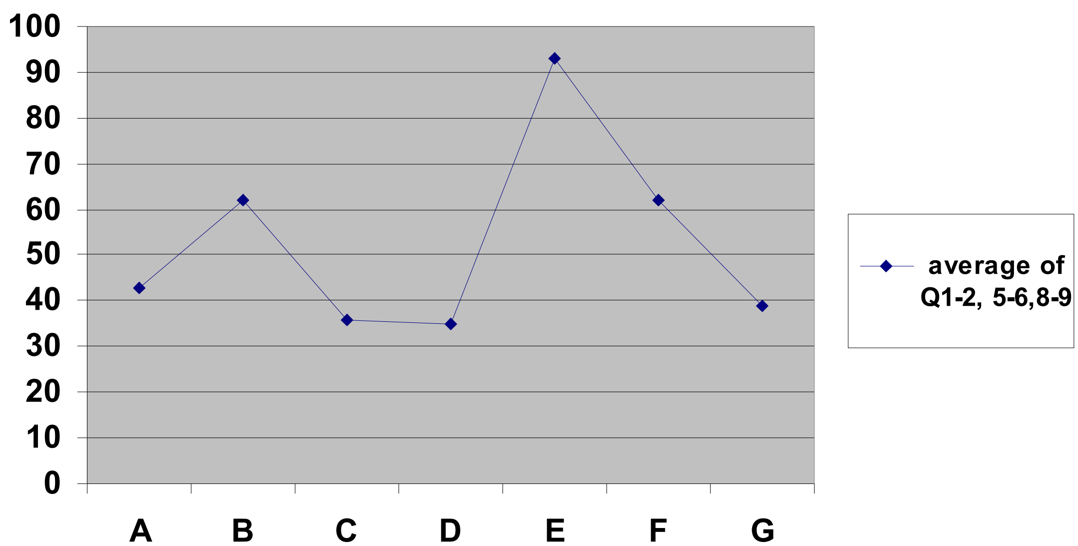


Fig. 3: Question sets 1-2 and 5-6
(x-axis= test form, y-axis=percent variance)

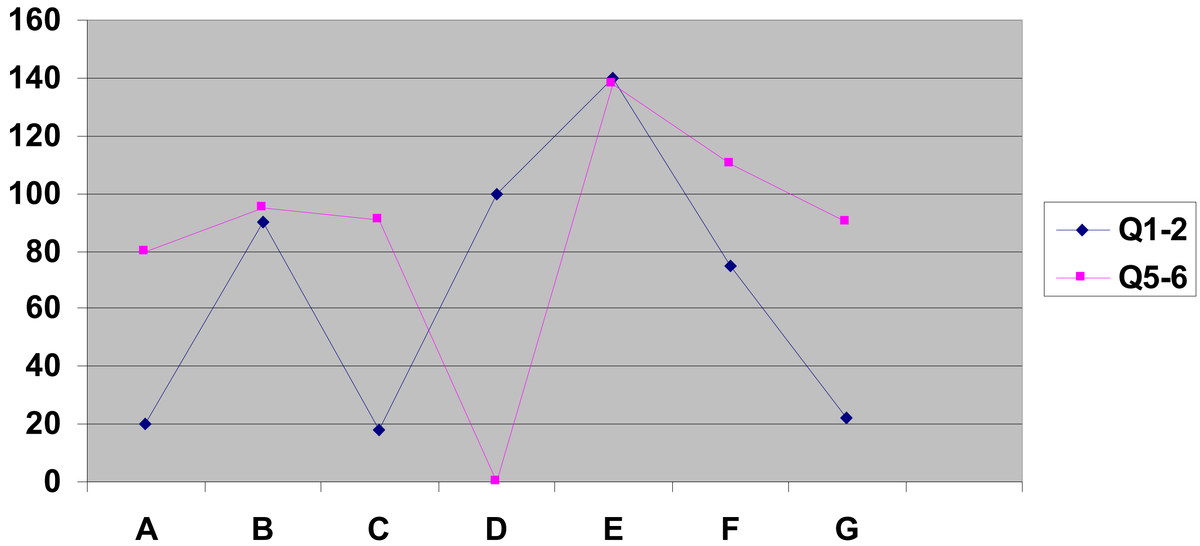


Fig. 4: Percent variance between Questions 8 and 9
(x-axis= test form, y-axis=percent variance)

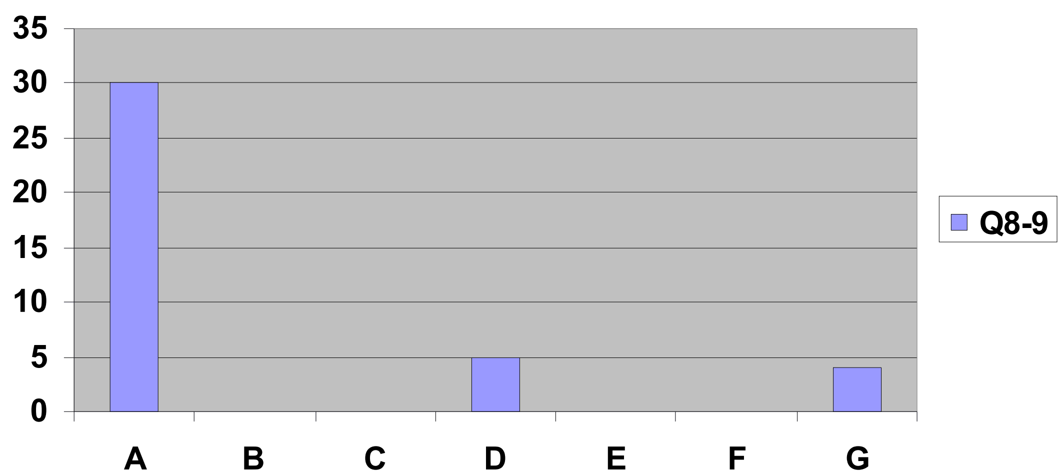


Fig. 5: Responses to single questions
(x-axis =test form, y axis =

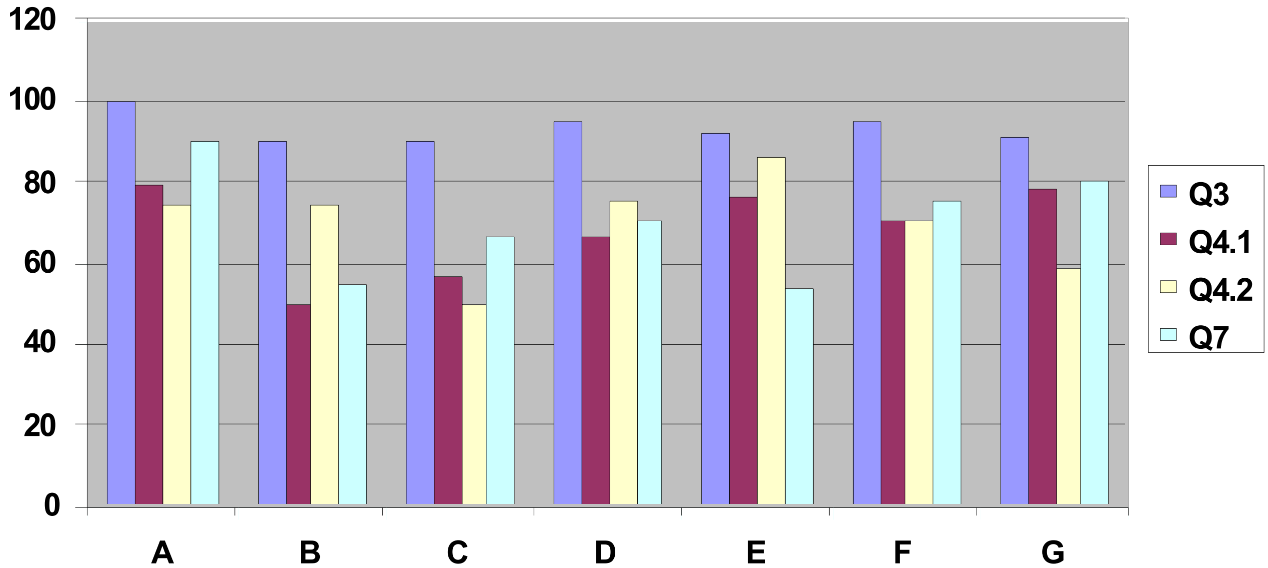
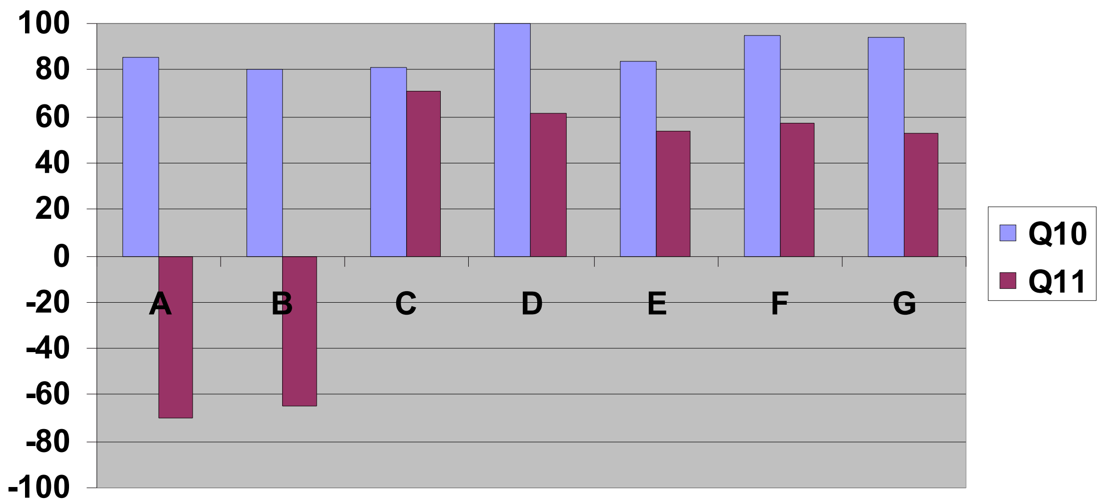


Fig 6: Questions 10-11
(x-axis= test form, y-axis= Positive numbers indicate yes response, negative no. The absolute value of the number is the percentage of of participants giving that answer)



Discussion

The results of this experiment partially support our hypothesis, but show a variable for which we had not planned: question format/difficulty. As the results show, participants in the test group did not perform consistently better on questions 1-6 than the control. However, on question sets 8-9 and 10-11, variance was eliminated in most control groups. For the latter questions, participants did significantly better than we had even hoped for: our test groups did not just perform better than the control, they performed almost *perfectly*. However, that our results only fit our hypothesis on the latter questions puts a different light on the results: it is difficult to say unequivocally that the Law of Large Numbers explains why people did poorly on previous studies and performed well on half of this study. On one hand, it may be the case that the Law of Large Numbers is behind the improved performance on the latter questions, and the former questions were too time-intensive to keep the interest of a university student who knows that he or she is allowed to go home when finished with the questionnaire, or who isn't familiar with the concept of calculating expected value. It is quite possible that this explains the results; the latter questions are just easier to do and take little time. Alternately, while the latter questions were real-life dilemmas, the tasks in problems 1-6 may have just been so unfamiliar as to alienate the participant: Walker, de Vries, and Trevethan (1987) have noted that most participants tend to perform poorly on tasked with which they do not personally identify. Walker et al's findings, however, are as likely a countering piece of evidence as a supporting one, because an ethical dilemma involving switching a trolley's path to save a million people (question 10, Form D of our study) is not likely to be a personally relevant criterion to many participants.

It is also highly likely that our participant's performance had nothing to do with an intuitive or overt understanding of the Law of Large Numbers, but rather, the inclusion of large numbers in the question format allowed for a more frequency-based approach to interpreting the questions. Brase and Barbey (2006) have shown that "frequency presentations of information lead to better judgments and decision making."[ix]

Interestingly, this distinction between frequency representations and the use of the Law of Large Numbers may be a false one, as frequency-type probability requires data gathered from a sufficiently large number of trials. More research would need to be done to unwrap all of these questions. A study comparing probabilities represented as frequencies to those represented as belief-type probabilities (non-frequency type) over large numbers would be helpful to separate the effect of the Law of Large numbers on understanding of probability information from frequency-representation on understanding of probability representation. Also, a study on people's willingness or ability to do problems like those posed in questions 1-6 of our study would be needed to understand why our participants did so poorly on these questions.

The idea of the numbers in groups E-G being "Anthropologically Significant" was brought into question after the trial was run. Given that these groups did not exhibit results uniformly different from the mathematically significant numbers, it can be assumed *prima facie* that these numbers do not have a separate significance.

References

Bernoulli, D. (1954). Exposition of a New Theory on the Measurement of Risk. *Econometrica 22,* 23-36. (Original work published 1738)

Brase, G. (2002). Which Statistical Formats Facilitate What Decisions? Thee Perception and Influence of Different Statistical Information Formats. *Journal of Behavioral Decision Making*. 15, 381-401.

Brase, G., & Barbey, A. (2006). Mental Representations of Statistical Information. *Advances in Psychology Research.* 41, 91-113.

Cohen, J. (2005). The Vulcanization of the Human Brain: A Neural Perspective on Interactions Between Cognition and Emotion. *Journal of Economic Perspectives.* 19, 3-24.

Greene, J., Sommerville, R., Nystrom, L., Darley, J., & Cohen, J. (2001). An fMRI Investigation of Emotional Engagement in Moral Judgment. *Science*, 293, 2105-2108.

Hacking, I. *An Introduction to Probability and Inductive Logic*. Cambridge, UK: Cambridge University Press.

Kahneman, D., & Tversky, A. Choices Values Frames. (1983). *APA Award Address*.

Kauder, E. (2005) Genesis of the Marginal Utility Theory: From Aristotle to the End of the Eighteenth Century. *The Economics of Brains*. http://www-psych.stanford.edu/~span/Press/bk0505press.html

Neumann, J., & Morgenstern, O. (1947). *Theory of Games and Economic Behavior* (2nd ed.). Princeton: Princeton University Press.

Roeser, S. (2006). The Role of Emotions in Judging the Moral Acceptability of Risks. *Safety Science*. 44, 689-700.

Rosen, G. (2007). The Brain on the Stand. *The New York Times.* (nytimes.com)

Walker, L., de Vries, B., Trevethan, S. (1987) Moral Stages and Moral Orientations in Real-Life and Hypothetical Dilemmas. *Child Development.* 58, 842-858.

Appendix 1

1: Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 60,000 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:

If Program A is adopted, 20,000 people will be saved.
If Program B is adopted, there is a one-third probability that 60,000 people will be saved and a two-thirds probability that no people will be saved.

2: If Program C is adopted, 40,000 people will die.
If Program D is adopted, there is a one-third probability that nobody will die and a two-thirds probability that 60,000 people will die.

3: You have a choice between two games, game A and Game B. You will play 100 rounds of whichever game you choose. All 100 rounds will only be played on one of the following games:
A: 25% chance to win $240 and
75% chance to lose $760
B: 25% chance to win $250 and
75% chance to lose $750

Which do you choose?

4: You are playing the same game as before: 100 rounds playing exclusively on game A or game B, but then another 100 rounds of either game C or game D.

Decision (i) Choose between:
A. a sure gain of $240
B. 25% chance to gain $1000 and
75% chance to gain nothing

Decision (ii) Choose between:
C. a sure loss of $750
D. 75% chance to lose $1000 and
25% chance to lose nothing

5: Consider the following two-stage game. As before, you will play 100 rounds exclusively of the game that you choose. In the first stage, there is a 75% chance to end the game without winning anything and a 25% chance to move into the second stage. If you reach the second stage you have a choice between:

A. a sure win of $30
B. 80% chance to win $45
Your choice must be made before the game starts, i.e.,
before the outcome of the first stage is known. Please indicate the option you prefer.

6 : Again, you will play 100 rounds exclusively of one of the following games that you choose. Which of the options do you prefer?

C. 25% chance to win $30
D. 20% chance to win $45

7: Imagine that you are about to purchase 100 jackets and 100 calculators for the math club. A jacket sells for $125 and a calculator for $15. The calculator salesman informs you that the calculator you wish to buy is on sale for $ 10 at the other branch of the store, located 20 minutes drive away. Would you make a trip to the other store? What if the jacket was $15, the calculators were $125, and the other
branch of the store had the calculators for $120?

8: Imagine that you have bought 100 tickets for a raffle at the price of $10 per ticket. As you enter the raffling place, you discover that you have lost all the tickets. The tickets cannot be recovered. Would you pay for 100 more tickets?

9: Imagine that you are on your way to buy 100 tickets for a raffle at the price of $10 per ticket. As you enter the raffling place,

you discover that you have lost $1000 in cash.
Would you pay for 100 more tickets?

10 A runaway trolley is headed for ten people who will be killed if it proceeds on its present course. The only way to save them is to hit a switch that will turn the trolley onto an alternate set of tracks where it will kill one person instead of ten. The track leading to the one person loops around to connect with the track leading to the ten people. We will suppose that without a body on the alternate track, the trolley would, if turned that way, make its way to the other track and kill the five people as well. Ought you to turn the trolley in order to save five people at the expense of one?

11. As before, a trolley threatens to kill ten people. You are standing next to a large stranger on a footbridge that spans the tracks, in between the oncoming trolley and the ten people. In this scenario, the only way to save the five people is to push this stranger off the bridge, onto the tracks below. He will die if you do this, but his body will stop the trolley from reaching the others. Ought you to save the ten others by pushing this stranger to his death?

[i] *Genesis of the Marginal Utility Theory: From Aristotle to the End of the Eighteenth Century* E Kauder, *The Economics of Brains*, http://www-psych.stanford.edu/~span/Press/bk0505press.html

[ii] (Expected value)=(probability)(utility) Ian Hacking's Introduction to Probability and Inductive Logic ch. 10

[iii] *Choices Values Frames* (1983) pg 342

[iv] Ibid. pg 343

[v] "State of Affairs" is a technical term. See Ian Hacking's Introduction to Probability and Inductive Logic, pgs. 117-119.

[vi] "Dominance" " is also a technical term. Kahneman & Tversky describe it in *Choices Values and Frames* as follows: "Dominance demands that if prospect A is at least as good as prospect B in every respect and better than B in at least one respect, then A should be preferred to B." For a further discussion see Ian Hacking's Introduction to Probability and Inductive Logic, pg. 115

[vii] "Rational", in the ethical dilemmas such as those in the Greene study, is arguably not synonymous with "correct", as Roeser (2006) and Cohen (2005) have noted.

[viii] (http://www.stat.berkeley.edu/~stark/Java/Html/lln.htm)

[ix]Within this experiment, however, they will be treated as synonymous.

Brase and Barbey 2006, pg 7