Testing Rationality

MICHAEL NEUMANN *Trent University*

ABSTRACT: Amos Tversky, Daniel Kahneman, Dan Ariely and others detect irrationality when decision makers get led astray by how a decision problem is framed. They find that test subjects respond inconsistently when the same decision problem is described differently. But when are two decisions the same? The participants in their experiments are not decision theorists and cannot be counted on to read or approach the problems 'properly.' They may find sources of utility where researchers least suspect, and change payoffs that 'ought' to remain constant.

RÉSUMÉ : Amos Tversky, Daniel Kahneman, Dan Ariely et quelques autres prétendent trouver de l'irrationalité quand des décideurs sont trompés par la formulation du problème qu'ils abordent. Les chercheurs notent que les sujets offrent des réponses incohérentes lorsqu'un même problème est formulé de plusieurs façons différentes. Mais quand doit-on estimer que deux problèmes sont «les mêmes»? Les sujets ne sont pas des chercheurs ni des experts en théorie de la décision. On ne peut s'attendre à ce qu'ils lisent les problèmes «comme il faut». Peut-être découvrent-ils des sources d'utilité là où les chercheurs n'en soupçonnaient guère; peut-être changent-ils des valeurs qui «devraient» rester constantes.

Keywords: rational choice, behavioural economics, decision theory, utility, rationality

The work of Amos Tversky and Daniel Kahneman, followed by Dan Ariely, caused great stir among economists by challenging the notion that we act according to conventional ideas of behavioural rationality. The response has sometimes been sceptical, but only within certain limits. Very rarely is it claimed that the experiments fail to attain what might be considered their first objective, namely to show that the narrow conception of behavioural rationality as the maximization of expected utility cannot be guiding our decisions. It is rather

***Dialogue* 57 (2018), 615–631.**

doi:10.1017/S0012217318000094

the implications of this result that have been called into question: should our 'non-standard' decision making count as irrational?

Some have maintained that the experimenters have too narrow an idea even of conventional economic rationality. These critics say that maximization is not the whole story, that such tendencies as risk-aversion are entirely compatible with rational economic behaviour.[1] These replies seem to underestimate the sophistication of the experimenters. The subjects are not judged irrational because they use some criterion of rationality other than pure expected utility maximization, e.g., one that incorporates risk aversion. The irrationality judgement comes instead from divergent responses to what are plausibly claimed to be equivalent descriptions of the same problem. The experimenters' criteria of equivalence are strong, and go beyond merely equivalent payoffs: they often incorporate, for instance, equivalent risks. So to demand consideration of expanded criteria of rationality such as risk aversion is to demand something that has already been granted. Thus in the famous lost ticket case, the complaint is not that subjects prefer a sure thing to maximal expected utility. It's that, even with the payoffs and risks constant, they go for the sure thing in one case but not in the other. Tversky and Kahneman are, for instance, quite explicit about this in their 'Asian disease' problem, discussed below.

Other responses do not challenge the experimenters' results so much as defend the test subjects. Some have asserted that the subjects' 'performance' on the experiments can, with practice, be improved. Some have asserted that such improvement is a part of rationality itself. Some maintain that alternatives were described so as to *induce* risk aversion or some other bias.[2] They suggest that irrationality was 'unfairly' induced: had, for example, the problem been formulated differently, risk-aversion would not have kicked in under one description but lain low under another. Still others advocate more flexible, expansive criteria of rationality.[3] These criticisms, however, accept

---

[1] An extended and wide-ranging exposition of this critique may be found in Richard McKenzie's *Predictably Rational?*

[2] Two examples among many: Richard Samuels, Steven Stich, and P.D. Tremoulet hold that "the experimental results are robust and can be readily replicated," but that people will respond more rationally if the tests take account of deep biases found in "the pre-human and early human environment." The revised tests promise more rational responses: "when information is presented in the right way, performance on reasoning tasks should improve dramatically." See Richard Samuels, Stephen Stich, and P.D. Tremoulet's "Rethinking Rationality." Others suggest, more generally, that questions that invite the decision maker to rely on 'familiar modes of reasoning' will avoid the cognitive disruption that 'degrades performance.' See Marvin S. Cohen's "The Bottom Line."

[3] See Nathan Berg's "The consistency and ecological rationality schools of normative economics."

the basic contention that test subjects deviate from the standard models of rational choice.

What follows argues that the tests are unreliable on other grounds. Perhaps they have not been set up 'unfairly.' Perhaps the test subjects aren't pursuing non-standard approaches to rationality. Perhaps there is no systematic explanation for why the subjects appear to deviate from standard models of rationality: perhaps these subjects have no principles or attitudes or biases with which they approach all decision problems, or at least a wide range of them. They may decide to ignore some features of the test or apply various decision-making tactics (such as satisficing) episodically and capriciously.

The most radical manifestation of such attitudes might be called '*editing*': in one way or another, the subjects feel entitled to change the assigned test and arrive at different probabilities or utilities.

'Editing' may be understood by comparing it to some other practices attributable to decision makers. Some test subjects may exhibit an *endowment effect*, as when more utility is attached to a lost ticket than to the purchase of a replacement.[4] Still others may be *risk-averse*, assigning more disutility than others to the 25% chance of losing a bet. But these deciders work with the problem's parameters as they are given.

Some deciders work with the given parameters but refine them. John Broome[5] gives examples where decision makers vary the ranking of alternatives according to what other choices are available. *Staying at home* or *going to Rome* becomes *staying at home when the only other choice is mountaineering*, or *staying at home when the only other choice is visiting Rome*. But the 'editors' discussed here do not suggest more fine-grained descriptions of their preference rankings and subdivide the alternatives. They do not work with what they are given. They change the content of the original alternative set, as if Broome's Maurice would simply eliminate visiting Rome and substitute visiting Padua.

The prospect of this sort of radical surgery or capricious fiddling with the tests suggests difficulties that seem to elude decision theorists who are used to working with ideal cases and pliable test subjects: in non-ideas cases, you can't trust people to accept a problem as you've written it, sometimes even if they see the problem as they're supposed to see it. This need for caution is detectable in three experiments that have become classics in the field. A review of the experimental programs these classics have inspired—beyond the scope of this essay—might raise questions about the often-heralded demise of conventional economic rationality.

To see how 'editors' (and other uncooperative subjects) might weaken the implications of the famous tests discussed below, consider their conceptual basis. Tversky, Kahneman, Ariely, and others detect irrationality when decision

---

[4] See Daniel Kahneman, Jack L. Knetsch, and Richard H. Thaler's "Anomalies."

[5] John Broome, *Ethics out of Economics*, 70.

makers get led astray by how a decision problem is *framed*. Tversky's notion of framing is perhaps best summarized by his editor and collaborator, Eldar Shafir:

> … decision situations can arise in which alternative descriptions of the same decision problem give rise to different choices. This is known as a framing effect and is in violation of the principle of invariance, which requires that logically equivalent representations of a decision problem, as well as logically equivalent methods of elicitation, yield the same preferences.[6]

Framing is the smoking gun in hunts for irrationality: you are responding *differently*, that is, inconsistently, to the *same* decision problem. You do so for various 'reasons' having to do with the fact that the same problem is described in 'alternative,' that is, different ways.

But what is the 'same' decision problem? It is agreed in decision theory that different descriptions of the 'same' alternatives should not matter. Suppose in problem 1, *A* has a payoff of 2 and *B* a payoff of 3. Suppose in problem 2, *buying a coke* has a payoff of 2 and *buying a Pepsi* has a payoff of 3. That's not two problems; it's one and the same. And decision-theoretically 'equivalent' descriptions have the same non-effect: if the alternatives in problem 2 are described as *buying the red can* and *buying the blue can*, it's still problem 2.

However, it's no longer the same problem if the payoffs have changed: alternative descriptions are inessential to the decision problem, but payoffs are essential. If we change the payoffs of *A/coke/red-can* from 2 to 3, we have a new problem. A further complication, scaling, does not concern us here. It may or may not be the same or at least equivalent problem if we multiply all payoffs by 100.

Tversky, Kahneman, and Ariely seem confident that the problem of sameness does not affect their work because 'only' the descriptions are changed. In some cases the payoffs are even better than equivalent; they're literally identical. Granted, the payoffs are not actual utilities but sources of utility, but that can't matter, because, after all, hello—they're the very same sources.

This confidence may be misplaced. To assert the sources are the 'same' makes a subtle but unwarranted assumption: that the sources of utility are fully specified in the payoff boxes. A simple example will illustrate how that assumption may founder. Suppose, for example, I have two choices: buy a bag of

---

[6] Eldar Shafir, "Editor's Introduction," in *Preference, Belief, and Similarity*, 406f. Tversky himself, as far as I can see, contents himself with a loose, contextual definition of 'framing,' sometimes equating it to 'characterization.' See, e.g., Barbara J. McMeil, Stephen G. Pauker, Harold C. Sox, and Amos Tversky's "On the Elicitation of Preferences for Alternative Therapies."

potatoes from vendor A, or buy a bag of potatoes from vendor B. The potatoes are of the same quality. A charges $10 and B charges $11, and my decision matrix is

store	price of the potatoes	utility of the potatoes
Webb's Foods	10	10
Arthur's Foods	11	10

so I'd seem rationally compelled to buy from Webb's. The same would be true if the payoffs involved sources of utility—bags of the very same brand and weight. But perhaps the choice is poorly described: Webb's is a big supermarket, and Arthur's is a small stand run by an impoverished family. In that case, assuming some compassion on my part, I will get far less utility buying from Webb's, and it would be irrational for me to do so. Yes, the 'social' utility of the purchase should have been in the boxes. But the point is that it may look like the boxes adequately specify the sources of utility when in fact they don't, and an experimenter, posing the problem, might not realize this.

When examining specific findings of irrationality, these factors loom large, because invariably the experiments find that a *minority*—often a small minority—of the respondents are irrational. This nurtures the suspicion that not all respondents adopted the most obvious reading of the problem: for them, the descriptions may have contained sources of disutility or utility unsuspected by the experimenters. So it takes only a vaguely plausible reading, not a 'best reading,' to undermine the experiment. The alternative readings gain still more stature because simpler and/or more elegant explanations are preferable. An explanation that posits irrationality requires a much more elaborate theory than one that does not.

So it may be in the researchers' experiments. A lost-ticket problem 'reframed' as a lost-money problem may, for some people in some circumstances, alter the utilities of the alternatives. Lost lives may have different significance for different agents, depending on exactly how those lives are lost. The test subjects may in effect rebel against the experimenters' reading of the problem. What follows offers suggestions for how this may happen, and how it may change payoffs that appear set in stone.

Rebellion of this sort, however annoying to researchers, is not irrational. It needs considering when evaluating the experiments' results. What follows suggests that, because of the possibility of rebellion, the experiments offer no compelling evidence of irrational choice. The apparently irrational responses can be explained, not by the effects of the formulations, but by the difficulty of getting test subjects to read problems like decision theory professionals. The test subjects often don't even think they are supposed to respect the given parameters of the problem; they feel free to elaborate a bit. The results are

utility assignments that defy the experimenters' intentions and indeed explicit instructions. It may well be these defiant assignments, and not any failure of rationality, that explain the experimenters' results.

The Ticket Problem

Here is the ticket problem as described by Tversky and Kahneman:

> Problems 8 and 9 illustrate another class of situations in which an existing account affects a decision:
>
> Problem 8 [$N = 183$]: Imagine that you have decided to see a play where admission is \$10 per ticket. As you enter the theater you discover that you have lost a \$10 bill.
>
> Would you still pay \$10 for a ticket for the play?
>
> Yes [88 percent] No [12 percent]
>
> Problem 9 [$N = 200$]: Imagine that you have decided to see a play and paid the admission price of \$10 per ticket. As you enter the theater you discover that you have lost the ticket. The seat was not marked and the ticket cannot be recovered.
>
> Would you pay \$10 for another ticket?
>
> Yes [46 percent] No [54 percent]

and the authors comment:

> The marked difference between the responses to problems 8 and 9 is an effect of psychological accounting. We propose that the purchase of a new ticket in problem 9 is entered in the account that was set up by the purchase of the original ticket. In terms of this account, the expense required to see the show is \$20, a cost which many of our respondents apparently found excessive. In problem 8, on the other hand, the loss of \$10 is not linked specifically to the ticket purchase and its effect on the decision is accordingly slight.[7]

Describing the Alternatives

One of the complications attending empirical studies of rationality is that your test subjects cannot be held to exactly the reading of the problem you want them to accept. Anyone who has taught the material knows this: students will, for example, regularly suggest that, given a stark choice, there is actually 'something else you could do, namely' … This of course has nothing to do with irrationality. In analyzing the responses to the ticket problem, this complication has to be kept in mind, particularly since the problem, as presented, positively

[7] Amos Tversky and Daniel Kahneman, "The Framing of Decisions and the Psychology of Choice," 457.

https://doi.org/10.1017/S0012217318000094 Published online by Cambridge University Press

invites further interpretation. Indeed, it appears as if the experimenters themselves have re-framed the decision problem, even loosely re-framed it. Their descriptions invite interpretations that alter the payoffs in a fundamental way, so that we have two problems rather than one in different guises.

The first situation is curiously described: it is said that, upon entering the theatre, I discover I have lost $10. To 'discover' that I have lost the money implies that I don't know how it came to be missing.

What might have happened that makes me sure the $10 is lost rather than stolen? Answers to questions like that usually require some sort of investigation that, since I am about to enter the theatre, I did not undertake. For all I know, some carelessness was involved on my part. It is not, after all, as if the wind has blown it away because I didn't hold the bill properly, or I let it slip and it fell into a river. This seems to be a case where I have suffered a loss, where the $10 goes missing, and I don't know how that happened. It is not a clear case of *my* having lost it, in the sense that the responsibility must be mine. Who knows? Maybe the money is only lost *temporarily*; it may turn up somewhere. That's certainly happened to me.

The case of the lost ticket is a different matter altogether, for more than one reason. In the first place, money gets stolen all the time. Ordinary theatre tickets do not. In the second place, I might well understand the scenario to mean that the ticket has only recently been in my possession: I bought it in the ticket line (the example predates ticket machines and online purchase), and entered the theatre soon after. At the very least, if I discover on entering the theatre that I have lost $10, that very much allows for the possibility that the loss, even if it was my fault, occurred quite some time ago. Finally, theatre tickets, unlike money, are normally valid only for a certain time. Even if they turn up again, they're worthless.

Now suppose that the respondents divide, not into accounts of different sorts, but simply into optimists and pessimists. The optimists think it likely that the money will turn up somewhere at home; the pessimists do not. Of course, among the optimists and pessimists, some are also well-off, others not. Put these factors together, and what do you get? Among the worse off, there are:

- the optimists, who discount the possibility that the money is gone for good;
- the pessimists, who think it likely that the money is gone for good;
- the ticket-losers. Since the ticket clearly *is* either gone for good and/or worthless, optimists and pessimists alike consider themselves permanently out of pocket. Their calculations shouldn't differ.

Of course, there are optimists and pessimists among the better off. But since neither group lets their views about the $10 bill or the ticket affect their choice, we don't need to look at them closely. They'll probably all buy a second ticket because it makes no appreciable difference to them whether they pay $10 or $20 to see the play. Their actions should count as consistent.

Those accused of inconsistency are those who changed their preference from one situation to the other. In the money-loss situation, they bought a second ticket. In the ticket-loss situation, they did not.

Now take Joan, a worse-off optimist. She discounted the possibility that the money was gone for good, and bought a second ticket. Being worse off, she would not have done so had she thought she was permanently out the $10.

When Joan confronts the ticket-loss situation, she is sure she is permanently out the $10 she spent. She believes herself too poor to buy a second ticket, so she doesn't.

Perhaps Joan is like the preference-switchers who are accused of inconsistency. They are not inconsistent; they are optimists. They pretty much assumed they were not permanently out $10 in the money-loss case. But, like everyone, optimist or pessimist, they knew they were permanently out $10 in the ticket-loss case. (Inconveniently for some experimenters in behavioural rationality, most people realize they are rarely in a position to assign probabilities to uncertain outcomes.) They now feel they cannot afford a second ticket. Optimism—belief that the lost bill will turn up—not some form of behavioural irrationality, accounts for their choices.

Could Joan's optimism be irrational? We don't know: maybe she often 'loses' money only to rediscover it. But it doesn't matter. Even if Joan is irrational in the sense that she weighs negative evidence too lightly, that only determines how she assesses the facts of the choice situation, not the rule she applies to that situation. She applies her rules consistently, to maximize net expected return. She is therefore not irrational in the sense explored in the experiment.

There is even a certain amount of wiggle room in this reading of the results. Perhaps not all of the worse-off people who didn't change their preference were pessimists. Perhaps some were just people who read the situation as the experimenters intended. They took the money, like the ticket, to be definitively lost, not because they tended to expect the worst, but simply because they understood that the experimenters intended them to consider the loss permanent.

One could also question what is 'irrational' about accountings similar to the ones posited by Tverksy and Kahneman. Suppose some ticket-losers say the following:

> If I had not bought the ticket, I would not be out $10. But, when I lost the $10, I would have been out $10 whether or not I had bought the ticket: the loss is in no way causally related to the ticket-buying. Therefore the loss of the ticket should come out of my entertainment budget. The loss of the $10, however, should not: this is simply a lesson to be more careful with my cash. So, in the case of the money-loss, I won't count it against my entertainment budget. In the case of the ticket-loss, I will, and my budget does not permit the expense of an additional ticket. Therefore I will buy an additional ticket in the case of the money-loss, but not in the case of the ticket-loss.

This doesn't seem to be 'psychological' accounting; it is just accounting. It is not clear why it violates any canons of utility calculation or behavioural rationality. These ticket-losers seem to have grounds for holding that they are paying twice for the ticket in the ticket-loss case—where in fact they buy a ticket twice—but not in the money-loss case—where in fact they do not buy a ticket twice. If so, the ticket-loss case and the money-loss case have, for these people, different payoffs. Then we cannot have a case of two ways of framing the *same* (or essentially the same) decision.

The Candies

Another experiment is reported by Ariely, and allegedly involves an irrational bias in favour of free items. Ariely describes the situation as follows:

> In one experiment, … [w]e set up a table at a large public building and offered two kinds of chocolates—Lindt truffles and Hershey's Kisses. There was a large sign above our table that read, "One chocolate per customer." Once the potential customers stepped closer, they could see the two types of chocolate and their prices.
> … When we set the price of a Lindt truffle at 15 cents and a Kiss at one cent, we were not surprised to find that our customers acted with a good deal of rationality: they compared the price and quality of the Kiss with the price and quality of the truffle, and then made their choice. About 73 percent of them chose the truffle and 27 percent chose a Kiss. Now we decided to see how free! might change the situation. So we offered the Lindt truffle for 14 cents and the Kisses free. Would there be a difference? Should there be? After all, we had merely lowered the price of both kinds of chocolate by one cent. But what a difference free! The humble Hershey's Kiss became a big favorite. Some 69 percent of our customers (up from 27 percent before) chose the free! Kiss, giving up the opportunity to get the Lindt truffle for a very good price. Meanwhile, the Lindt truffle took a tumble; customers choosing it decreased from 73 to 31 percent.[8]

Must these results indicate an irrational lust for free items? Not if the test subjects don't approach the problem with the desired seriousness. Though the experimenters try to cover the ways in which subjects might take the problem other than as intended, they can't cover all the bases.

Conventionally, rational utility-assessments could induce you to prefer taking a Lindt in the first situation and a Hershey's in the second. Several factors undermine this assessment.

Suppose that the Lindt is just barely preferred in the first decision. It has much more utility than the Hershey's, but the money spent eats up a lot of that value, so that the *net* utility of *choosing* the Lindt is just 0.1 'utiles' greater than the net utility of choosing the Hershey's. The disutility of each decision is, of course, generated by the money spent.

---

[8] Dan Ariely, *Predictably Irrational*, 51f.

https://doi.org/10.1017/S0012217318000094 Published online by Cambridge University Press

Here is a table indicating the value of the 1st and 2nd alternatives for some of those who choose the free Hershey's in the second situation. If money is linear with utility, there is no ground for preference reversal.

	1st prices	1st net utilities	2nd prices	2nd net utilities
Lindt	15¢	(25 - 15) = 10	14¢	(25 - 14) = 11
Hershey's	1¢	(10.9 - 1) = 9.9	0¢	(10.9 - 0) = 10.9

But there is another not-implausible interpretation of the results. For quite a few people I know, the (money-equals-utility) analysis of the situation would be stilted, especially at such low price points. These people really don't care at all whether they pay a penny more or less for something. Money, for them, is of course not linear with utility. In this case, we might suppose that 15 cents (*and* 14 cents) is worth a bit less than that in utility units, say 13.5. A single cent gets 0 units of utility.[9] But what bugs these people is having to dig up change for a trivial transaction. They might resent this more for a one-cent purchase than for a 15-cent or 14-cent purchase, but let's suppose not. In each case, they assign a utility of -2 for the transaction. Then we might get the following:

item	item utility	1st prices	price disutility	transaction disutility	1st net utilities	2nd prices	price disutility	transaction disutility	2nd net utilities
Lindt	25	15¢	-13.5	-2	9.5	14¢	-13.5	-2	9.5
Hershey's	10.9	1¢	-0	-2	8.9	0¢	-0	-0	10.9

Here preference-reversal happens not simply because the Hershey's Kiss no longer requires the outlay of a penny, but because it no longer requires a trivial and therefore irksome transaction.

Later, the researchers got worried about transaction costs and devised another experiment. This time the test subjects went through a cafeteria line where they could pay for a Lindt or a Hershey's in the first run, or just pay for a Lindt or take a free Hershey's in the second run. But this experiment does not exclude the possibility that some people who took a free Hershey's may have ignored the Lindt offer rather than assessed its utility. For those individuals, there *were* no relative utility assessments biased towards free.

[9] There are two ways a penny could have zero utility: because its monetary utility is zero, or because its monetary utility is 1, but there is a disutility of 1 because you have to carry the thing around with you. Similarly, spending 14 cents might have 1 unit less monetary utility than 15 cents, but there will be the disutility of getting a penny back and having to carry the damn thing around.

https://doi.org/10.1017/S0012217318000094 Published online by Cambridge University Press

Crucially, the respondents were not required to choose between the chocolates, so they were not required to weigh the utilities of both alternatives. The experimenters themselves say: "it was difficult to separate customers who decided not to participate from those who did not notice the offer; therefore, all customers who passed by the cashier and did not select any of our chocolates were coded as 'nothing.'"[10] But what about the non-participants who consistently ignore all 'conveniently situated' promotional sale offers—as some consumer advocates recommend—but scoop up freebies? This is not necessarily irrational, especially in a checkout lineup. With freebies, unlike for-pay promotions, I am absolutely certain not to miscalculate utilities, nor to waste time calculating them, nor indeed to be seduced into irrational impulse purchases. Finally, it is noteworthy that, in the cafeteria experiment, those taking the free Hershey's declined significantly, from 42% to 31% in the most nearly comparable cases.

Even if the experiment were rock-solid, it would not deliver quite as advertised. Ariely is convinced that the dramatic change in preference-order indicates a *strong* bias in favour of free.[11] This doesn't follow. That many people reversed their preferences can only indicate a *widespread* bias. Any bias driving these reversals may be quite weak, even it surfaces in a large proportion of the population: indeed, it may surface only when people, confronted with a trivial choice, don't bother with calculating utilities.[12] Were Ariely to seek evidence of *strong* bias in favour of free, he would instead present the test subjects with a far more important choice. Suppose, instead of chocolates, the experimenters were offering two cancer treatments called 'Good' and 'Fair,' one of which was far more effective than the other. It seems unlikely that many of us would prefer Good at $1,000 to Fair at $2, but Fair at $0 to Good at $998. Our love of

---

[10] This remark raises uncertainties about the procedure. There is no mention of anyone who passed by the cashier and *did* select a chocolate but *did not* make a purchase. This seems odd. Generally speaking, in a cafeteria, you can't exit without passing by the cashier. The sign just said: "one chocolate per person"; it did not say anything about having to purchase something in order to take a free chocolate. So, weren't there people who didn't buy anything but took a free chocolate? How were they coded? Since these people likely existed, but were not mentioned, it sounds as if the experimenters assumed anyone taking a free chocolate had also purchased something. This would invalidate the experiment, because these people had no transaction costs. See Kristina Shampanier, Nina Mazar, and Dan Ariely's "How Small is Zero Price?" 748.

[11] After describing his results, Ariely exclaims: "What is it about free! that's so enticing?" See *Predictably Irrational*, 54.

[12] Failing to calculate is itself a response to a quite different decision problem—whether or not to calculate. There is no *a priori* way to determine whether a particular individual's failure to calculate in the candy-choice situations is irrational.

freebies is properly measured by how *much* we will give up for something free, not by how *many* of us are willing to sacrifice a small quantity of utility for a free item.

It does not end there. If the bias towards the free is very weak—the most that can have been established—it is probably not irrational, even if it leads to non-maximizing choices. Why? Because it is also rational to be a *satisficer*. With these trivial utilities at stake, both outcomes are, for many good *enough*. If a bias tilts towards 'free' outcome in cases like these, when trivial differences obtain, it still tilts to a satisficing outcome. So, even at best, the experiment does not show irrationality.

It is worth noting the unsystematic aspect of the test subjects' behaviour. Perhaps these individuals do not always discount free offers, or always exhibit indifference to small quantities of utility, or always get annoyed by fiddling with change. The problems here are not a matter of recognizing possible *strategies* for decision problems. Perhaps all we have here are intermittent whims; hence the unpredictability.

The Asian Disease

More subtle problems emerge from another reputed example of framing. It is quite elaborate, and described as follows:

> It is often possible to frame a given decision problem in more than one way. … rational choice requires that the preference between options should not reverse with changes of frame. Because of imperfections of human perception and decision, however, changes of perspective often reverse the relative apparent size of objects and the relative desirability of options. …
>
> The effect of variations in framing is illustrated in problems 1 and 2.
>
> Problem 1 [$N = 152$]: Imagine that the U.S. 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 estimate of the consequences of the programs are as follows:
>
> If Program A is adopted, 200 people will be saved. [72 percent]
>
> If Program B is adopted, there is 1/3 probability that 600 people will be saved, and 2/3 probability that no people will be saved. [28 percent]
>
> Which of the two programs would you favor?
>
> The majority choice in this problem is risk averse: the prospect of certainly saving 200 lives is more attractive than a risky prospect of equal expected value, that is, a one-in-three chance of saving 600 lives.

Now the authors discover irrationality and inconsistency when the test subjects are presented with what is said to be the same choice, differently 'framed':

> A second group of respondents was given the cover story of problem 1 with a different formulation of the alternative programs, as follows:
>
> Problem 2 [$N = 155$]:
>
> If Program C is adopted 400 people will die. [22 percent]
>
> If Program D is adopted there is 1/3 probability that nobody will die, and 2/3 probability that 600 people will die. [78 percent]
>
> Which of the two programs would you favor?
>
> The majority choice in problem 2 is risk taking: the certain death of 400 people is less acceptable than the two-in-three chance that 600 will die. The preferences in problems 1 and 2 illustrate a common pattern: choices involving gains are often risk averse and choices involving losses are often risk taking. However, it is easy to see that the two problems are effectively identical. The only difference between them is that the outcomes are described in problem 1 by the number of lives saved and in problem 2 by the number of lives lost. The change is accompanied by a pronounced shift from risk aversion to risk taking. … Inconsistent responses to problems 1 and 2 arise from the conjunction of a framing effect with contradictory attitudes toward risks involving gains and losses.[13]

This experiment invites three interpretations that undermine the results. The interpretations could apply simultaneously to different respondents. These respondents *may* be reasoning deviantly, but it seems quite likely that they are reading or reacting deviantly instead. The possibilities are surveyed below.

> *They use expected utility calculation on all alternatives, and, realizing it makes no difference, choose on a whim.*

Suppose all respondents see the problems *exactly* as they are apparently supposed to. They do a calculation of expected deaths and lives saved, assigning a probability of 1 to the outcomes A and C. It is of course assumed that those not saved will die, and those who die will not be saved. Then all four outcomes have exactly the same value: 400 will die, 200 will be saved. But then it's unclear how indulging biases can count as irrational, because doing so is utility-neutral. Technically all strategies are weakly maximizing and therefore equally rational choices. A finding of irrationality would require that the respondents *sacrificed* at least some expected value in favour of risk, or against it.

> *They mistrust probability*.

Perhaps the respondents take seriously only the outcomes stated in non-probabilistic terms. They embrace A because it states, non-probabilistically,

[13] Tversky and Kahneman, *op. cit*., 453.

https://doi.org/10.1017/S0012217318000094 Published online by Cambridge University Press

that 200 will be saved, but not that anyone will die. They avoid C because it states, non-probabilistically, that 400 will die, but not that anyone will be saved. In this case, the respondents are not inconsistently risk-averse; they are consistently probability-sceptical.

Here is another case where real-life test subjects misbehave, refusing to take the problem as stated. That they are told the consequences represent an 'exact scientific estimate' would only put them further off trusting the payoff matrix. 'Exact … estimate' is almost comical, and, if this is an 'unusual Asian disease,' where does the alleged exactitude come from? The experimenters' description conjures up visions of blinkered scientists entirely too confident of their estimates in areas where, real-life experience shows, exactitude is unattainable. When have the (probability-weighted!) consequences of a disease outbreak been so precisely foreseen, let alone with perfect accuracy? And what are we to make of the 'definite' losses which, though they occur in the future, appear to be assigned a probability of 1, the same as would be assigned a logical truth? Obedient graduate students know that such assumptions are mere conveniences, never to be taken seriously. Ordinary respondents, even if they did know this, would be unlikely to accept it, at least if my own teaching experience is any indication.

Suppose, then, that the respondents take seriously only the non-probabilistic certainties, and do so outside a probabilistic framework. They believe that the experts have very good grounds for predicting 200 lives saved in scenario A, and 400 dead in scenario C. However, they do not feel compelled to infer that, under A, 400 will die, and under C, that 200 will be saved. The disease, absent any special countermeasures A through D, is merely 'expected' to kill 600 people. Maybe, under A, those who aren't strongly expected to be saved won't die after all. So A might quite *reasonably* seem a better bet than B, which is described in terms of probabilities that the scientists very likely do *not* have strong evidence to assert. Who knows, in short, what will happen under B?

When we get to the second choice, the same consideration work in reverse. A non-probabilistic assertion that 400 will die might be taken quite seriously. The probability spread under D might be taken as what we might *reasonably* suppose it is, a guess. And, under C, as under A, we don't really know what is going to happen to those *not* asserted to die. Will there really be 200 saved? We don't know. So, someone might think, we might as well opt for the uncertainties of D, which at least don't seem to rest on any firm evidence about the outcome. This is *not* to accept a probabilistic calculation of 'expected deaths.' It is to embrace an uncertainty that might not be so bad over what seems a near-certainty of a terrible outcome, in C. Whatever one might think of such reasoning, it is certainly not a case of inconsistent risk-aversion, because the probabilities that determine the risks have been rejected in both scenarios.

They see agency in A/B, but not in C/D.

The problems may seem 'effectively identical' to consequentialists like me and, apparently, the authors, but perhaps not to non-consequentialist respondents. 'To save' is a transitive verb; it implies that something is done to something. 'To die' is intransitive, it merely implies that something happens to someone. So 'saved' strongly suggests agency while 'die' does not. That people are saved may well be taken to imply that someone saves them. In this formulation, the risky alternative seems like 'gambling with people's lives.' This may well seem wrong, a gross violation of their autonomy, so it may be thought to involve greater moral liability. So A may well seem morally preferable. However, in the second pair of choices, it seems like the deaths will just happen, so neither choice involves a moral liability. Though this sort of analysis is highly questionable, it reflects common moral biases.[14] It is not behaviourally irrational and hardly, as the authors assert, 'contradictory.'

Here we have three types of unsystematic deviation from the intended problem—not three cases of a systematic but deviant strategy. Completely ignoring probabilities is equivalent to editing the problem. Choosing on a whim is, well, whimsical, not the expression of some notion about how to decide when the difference between alternatives is negligible. Excluding an alternative on moral grounds is a different sort of editing: it adds a moral dimension unseen by the experimenters, but does not change the decision problem. Instead, the test subjects choose on moral rather than decision-theoretic criteria. Here the test subjects either deal with a different decision problem, or approach the problem from a standpoint distant from rules of behavioural rationality.

Conclusion

None of this shows that researchers' principal findings are wrong. There may indeed be such a thing as negative framing and it may indeed undermine the congruence between economic rationality and real-world decision making. The foregoing merely suggests that the evidence supporting these contentions is flawed: not because some alternate decision strategies are overlooked, but because the test subjects 'edit' or rebel against the problems in a manner in which well-behaved test subjects do not. Since findings of inconsistency always depend on choosing different strategies in the 'same' situations, one must always ask whether the test subjects are working with the same situations as the experimenters—and if not, *why* not. Until we know how much 'editing' and rebellion are going on, we don't know the prevalence of framing in a given group of test subjects.

---

[14] More complex moral reasons for the purportedly irrational choice patterns are offered by Frances M. Kamm, "Moral Intuitions, Cognitive Psychology, and the Harming-Versus-Not-Aiding Distinction," and by Tamara Horowitz in a paper extensively quoted by Kamm: "Philosophical Intuitions and Psychological Theory" (paper presented at the Rethinking Intuition Conference, Notre Dame University, South Bend, Ind., April 1996).

References

Ariely, Dan
2008 *Predictably Irrational: The Hidden Forces That Shape Our Decisions*. New York: Harper-Collins.

Berg, Nathan
2014 "The *consistency* and *ecological rationality* schools of normative economics: Singular versus plural metrics for assessing bounded rationality." *Journal of Economic Methodology* 21 (4): 375–395.

Broome, John
2004 *Ethics out of Economics*. Cambridge: Cambridge University Press.

Cohen, Marvin S.
1993 "The Bottom Line: Naturalistic Decision Aiding," in *Decision making in action: Models and methods, edited by* G.A. Klein, J. Orasanu, R. Calderwood, and C.E. Zsambok. Norwood, NJ: Ablex Publishing Corporation, pp. 51–102.

Horowitz, Tamara
1998 "Philosophical Intuitions and Psychological Theory." *Ethics* 108 (2): 367–385.

Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler
1991 "Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias." *Journal of Economic Perspectives* 5 (1): 193–206.

Kamm, Frances M.
1998 "Moral Intuitions, Cognitive Psychology, and the Harming-Versus-Not-Aiding Distinction." *Ethics* 108 (3): 463–488.

McKenzie, Richard
2010 *Predictably Rational?: In Search of Defenses for Rational Behavior in Economics*. Berlin: Springer.

McMeil, Barbara J., Stephen G. Pauker, Harold C. Sox, and Amos Tversky
2004 "On the Elicitation of Preferences for Alternative Therapies," in *Preference, Belief, and Similarity: Selected Writings of Amos Tversky*, edited by Eldar Shafir. Cambridge, Massachusetts: MIT Press, pp. 583–591.

Samuels, Richard, Steven Stich, and P.D. Tremoulet
1999 "Rethinking Rationality: From Bleak Implications to Darwinian Modules," in *Cognition, Agency and Rationality*, edited by K. Korta, E. Sosa, and X. Arrazola X. Philosophical Studies Series 79. Springer, Dordrecht, pp. 21–62.

Shafit, Eldar (Ed.)
2004 *Preference, Belief, and Similarity: Selected Writings of Amos Tversky*. Cambridge, Massachusetts: MIT Press.

Shampanier, Kristina, Nina Mazar, and Dan Ariely
2007 "How Small Is Zero Price? The True Value of Free Products." *Marketing Science* 26 (6): 742–757.

Tversky, Amos, and Daniel Kahneman
1974 "Judgment under Uncertainty: Heuristics and Biases." *Science* 185 (4157): 1124–1131.
Tversky, Amos, and Daniel Kahneman
1979 "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* 47 (2): 263–292.
Tversky, Amos, and Daniel Kahneman
1981 "The Framing of Decisions and the Psychology of Choice. *Science* 211 (4481): 453–445.