

Philosophy 2²3³: Intuitions and Philosophy
Fall 2009
Tuesdays and Thursdays, 1pm - 2:15pm
Library 209

Hamilton College
Russell Marcus
Office: 210 College Hill Road, Room 201
email: rmarcus1@hamilton.edu

Class 10 - Practical Reasoning

I. Irrationality and intuition

Stich and Nisbett discussed three kinds of errors in inductive reasoning.

Shafir explores several further problems.

We can take these examples as further support for Stich and Nisbett's claim that we can be in reflective equilibrium while holding illegitimate principles of inference.

Shafir also alleges that they constitute a violation of consequentialism, which, as the cornerstone of Expected Utility Theory, serves as an underlying principle of economic theory.

Shafir, who is a psychologist, defines 'consequentialism' slightly differently than philosophers do.

For psychologists and economists, a person is a consequentialist if she chooses her actions according to the anticipated consequences.

As social scientists use the term, it is not a moral theory, but a descriptive one.

It is not a way of justifying one's moral claims, but a prudential guide.

Cases in which people act in ways that do not maximize their interests are refutations of the social scientist's claim of consequentialism.

Such evidence does not undermine the, say, utilitarian philosopher's claim that the good is to be identified with that which brings about the best results.

The moral philosopher takes consequentialism to be a normative theory, a way of evaluating the rightness or wrongness of actions.

The examples that Shafir discusses are more properly interpreted, for the philosopher, as claims about people's irrationality.

Our interest in these claims lies in their application to the framework of seeking reflective equilibrium.

If people find themselves in reflective equilibrium while holding irrational beliefs, we have good reason to believe that being in reflective equilibrium is not sufficient for justifying our beliefs.

Later, we will have to ask the question whether such worries undermine philosophical methodology.

That is, our real concern is whether worries about the social scientist's consequentialism, like those of Shafir, translate into worries about the intuitions we use in philosophy.

But, that's a big question, for throughout the term.

Let's look now at the data.

II. An outline of Shafir's data

Shafir reviews recent research in several areas of irrational behavior.

1. Violations of procedure invariance (preference reversal) due to
 - a. compatibility
 - b. prominence
 - c. affect
2. Violations of the sure-thing principle (STP) due to the disjunction effect

III. Compatibility

Preference reversal is the name for the phenomenon in which people do not choose the most advantageous option.

To determine whether a bet is advantageous or not, you calculate its expected value (EV).

To calculate the expected value of a bet, you multiply the probability of winning the bet by its payoff; if applicable, you subtract the probability of losing the bet by its loss.

Two distinct bets were compared.

H: an 8-in-9 chance of winning \$4 (EV=\$3.56)

L: a 1-in-9 chance of winning \$40 (EV=\$4.44)

In both H and L, there was no cost for the bet, so no further money had to be subtracted.

Given that the expected value of L is higher than that of H, it would be purely rational (in the sense of maximizing one's expected outcome) to choose L over H.

Still, most subjects chose H over L.

This is an example of preference reversal.

Preference reversal is not necessarily a problem for a theory of rationality, since it could be argued that there are other factors which subjects are weighing.

That is, one might argue that subjects were not merely looking to maximize their expected monetary value.

There might be extraneous elements of the bet that are being weighed: the thrill of the bet, say, or the aversion to risk.

Those claims are less plausible, when a further case is examined.

If the source of preference reversal were extraneous factors, those factors would presumably be procedure invariant.

Logically equivalent elicitation procedures [must] give rise to the same preference order. Thus, if one option is chosen over another, it is also expected to be priced higher. Procedure invariance is essential for the interpretation of both psychological and physical measurement (60).

That is, the method of comparing the two bets should not affect one's preference.

Consider the same pair of bets, but with a different method of comparing the bets.

Instead of asking subjects which bet they prefer, the researchers asked subjects to price the two bets. (You can price something by asking subjects how much they would pay for it, or by asking them how much they would accept in order to sell it; in this case, the latter method was used.)

Subjects, when pricing the bets, preferred L to H.

So their preferences were reversed from the choice study.

This reversal is a violation of procedure invariance.

Shafir's explanation of preference reversal in this case relies on the claim that subjects are more likely to choose properly when the form of the survey is compatible with the example question.

Thus, since the example used a bet, involving money, the pricing option elicited rational responses.

Compatibility will explain why the pricing option elicits the maximization of expected value, but it does not explain the preference reversal of the choice option.

Shafir notes that another difference between the two cases is that one can price each bet in isolation

whereas choice must be done by considering the bets together, by comparison. These comparisons, he claims, lead to confusion.

IV. Prominence and affect

The confusions that arise from comparison are influenced by both prominence and affect. Subjects priced a 10,000 entry music dictionary with a new cover higher than a 20,000 entry dictionary with a torn cover, when priced in isolation. But, when compared, the relevant factor (number of entries) became more prominent, and subjects preferred the ratty, more comprehensive one.

Similarly, subjects' affective evaluation of the robbery scenarios led them to award greater compensation to the victim in an abnormal event than to the victim in his local convenience store.

Abnormal events generally trigger greater sympathy.

(Shafir's explanation of the greater sympathy has to do with which counterfactual scenarios we consider; the example from the Olympics, that silver-medal winners are less satisfied than bronze-medal winners, is particularly compelling.)

But, when the two cases were presented together, subjects reasoned that the victims should receive the same compensation.

The prominence of a principle of equity prevails.

I would be interested in your responses to the case of the two job offers.

Take the MBA case:

- A. \$75,000, the same as other grads
- B. \$85,000, but others will be earning \$95,000

Subjects tended to choose A when evaluated in isolation, but B when A and B were compared.

What is the role of affect, here?

V. The sure thing principle and the disjunction effect

The cases in Shafir's article which have the most direct relevance to our interests are the violations of the sure-thing principle (STP).

STP: If x is preferred to y knowing that event A obtained, and x is preferred to y knowing that event A did not obtain, then x should be preferred to y even when it is not known whether A obtained (66).

STP is an application of deductive logic to ordinary reasoning.

Since event A either does or does not obtain, then x should just be preferred to y .

But Shafir presents cases in which STP does not hold.

[These cases remind me of two interesting paradoxes: [Simpson's paradox](#) and [Parrando's paradox](#).]

STP is a foundational assumption of theories of rational choice, just as procedure invariance is essential to measurement.

But, the disjunction effect leads to violations of STP, and thus to consequentialism, in Shafir's sense.

In the case that Shafir discusses, people are asked to consider a gamble, a fifty percent chance of winning \$200 and a fifty percent chance of losing \$100.

Whether they win or lose that gamble, they are offered a second instance of the same bet.

Note that the expected value of the gamble is $.5(200) - .5(100) = +\$50$.

So, it's a good bet whether you've won or lost.

When subjects were presented with the second bet, knowing whether they had won or lost the first, they tended to take the second bet.

But, when subjects were presented with the second bet, *not* knowing whether they had won or lost the first, they tended to decline the second bet.

This result is called the disjunction effect: we respond differently to a disjunction than we do to the individual disjuncts.

The disjunction effect held even when subjects' decisions were calibrated to account for generally accepted principles governing the subjective value of modest gains and losses.

It turns out, according to a well-supported body of research known as prospect theory, that people's decisions in such cases tend to be loss-averse.

Because people tend to value avoiding small losses greater than achieving small gains, people will be disinclined to make the bet described above.

But, the same value function Shafir presents (pp 68-9) predicts that people will tend to accept the second gamble if they know what the result is on the first bet; the prediction holds.

Prospect theory further provides an explanation for why people reject the second bet if they don't know what the result of the first bet is.

Instead of reasoning as if the first bet were already completed, they reason from their present epistemic state.

And, prospect theory predicts that subjects should not take the gamble if they have not already done so.

When confronted with the disjunctive scenario above, people do not evaluate the attractiveness of the second gamble from two alternative positions, one assuming a gain and one assuming a loss, as implied by STP. Instead, not knowing whether they have won or lost the first, people segregate the two gambles and evaluate the second from their current position, as if for the moment no change had occurred. Uncertain about the previous outcome, people evaluate the situation as if no outcome had obtained (70).

Having won the first gamble assures the person of a no-loss situation, and having lost compels her to try to recover her losses. Uncertainty, on the other hand, brings about a state that is not a disjunction of the former two, but an independent tendency to be cautious and avoid further losses (ibid).

(We will return to prospect theory, which Shafir uses to do the calibration, later in the term.)

VI. Inductive inference and philosophical intuition

The conclusion we should make, according to Shafir, is that these results should be troubling to philosophers.

[There is] a discrepancy between the nature of people's everyday experiences and the conditions that yield philosophical intuitions (72).

Considered in isolation, these results are somewhat damning to the claim that people behave rationally all the time.

But, no one really holds that claim.

So, that argument is a scarecrow.

[I should mention that I like 'scarecrow' as an alternative to the sexist 'straw man' and the closest PC alternative, 'straw person', which evokes nothing.]

One concern is whether Shafir's results point to a problem with reflective equilibrium.

Stich and Nisbett claimed that their results were cases in which people held invalid inductive inferential principles in reflective equilibrium.

Shafir's claims, though, seem clearly not to be in reflective equilibrium.

Consider the disjunction effect.

When reasoning in isolation, people are subject to the disjunction effect, and violate STP.

But, when they are presented with the cases together, they defer to STP.

Once the applicability of STP is detected, for example in a transparent within-subject design, people typically find it compelling to the point of being irresistible (70).

Similarly, in the robbery scenario, people awarded the same amounts when the cases were compared, though their affect led them to differential awards when the cases were considered in isolation.

Thus, while Shafir's claims may impugn our initial intuitions, they may not impugn our considered judgments.

A slightly different concern is whether Shafir's results show some deep fault in our intuitions.

He claims that a source of difficulty for philosophers is that they tend to consider how people behave in comparison, but that people are actually faced with their decisions, generally, in isolation.

When a philosopher introspects about how people will, or even ought to, evaluate different options when these are presumably encountered in isolation, the philosopher will be confined to a concurrent evaluation, with the various alternatives before his or her attention. to the extent that the two forms of evaluation - concurrent and in isolation - lead to differential weightings, there will be a systematic tendency for people to experience events in isolation that will remain beyond the scope of within-subject intuition. In everyday life, we tend to experience and evaluate scenarios one at a time, whereas intuitions about relative worth often arise from introspectively comparative evaluations. As a result, the weighting of dimensions that goes into the making of common intuition can differ systematically from the weights that are assigned in actual experience (62).

This seems like a reasonable claim to try to assess in class.

Next up: Gettier cases, and a more firm application of similar results to philosophical reasoning.