Friedman's Risky Methodology

by

Chris Starmer

*School of Economics

University of Nottingham

NG7 2RD, UK.

Tel ++44 115 846 6067

chris.starmer@nottingham.ac.uk

Word count: 7364

Introduction

In this paper my aim is to offer some reflection upon Friedman's methodology focussing in particular upon his work relating to the theory of risky choice. With this objective in mind, I revisit two key papers in which Friedman and Savage (1948,52) discuss the empirical support for expected utility theory (EUT)[1]. In these papers there are clear traces of the methodological position set out in Friedman's 1953 essay (henceforth F53). For instance, a recurrent theme is that EUT should be judged in terms of its predictive accuracy relative to its intended realm of application. At the same time, both papers are motivated by the scarcity of 'direct' empirical evidence and in light of this each paper suggests some remedy. My interest will focus especially on the second of these contributions which argues that the normative appeal of EUT axioms can be read as a source of 'indirect' evidence for the hypothesis.

I will argue that this claim is methodologically problematic partly because the argument set out in the joint 1952 paper fails to successfully forge an explicit link from normative appeal of EUT axioms to predictive success of the hypothesis. A key question addressed in this paper is how, if at all, such a link could be established? I suggest the issue is of more than passing historical interest because there has been an apparent tendency in modern economics for the selection of descriptively oriented theories to be guided by normative criteria. Put differently, at times economists behave as if they believe that normative appeal is a form of empirical evidence and I illustrate this with a particular example in the contemporary theoretical literature on risk.

Development of the Expected Utility Hypothesis

EUT is an extremely widely used piece of economic theory. In one respect it is easy to see why. Risk is a pervasive aspect of the economic world and for some time now, EUT has been the standard approach to modelling decisions involving risks. But, although the theory dates as far back as Bernoulli (1738), it did not find much favour with modern economists until the 1950s. This may be partly

explained by the fact that, in the form presented by Bernoulli, the theory presupposes the existence of a cardinal utility scale; an assumption which did not sit well with the drive towards ordinalization during the first half of the twentieth century. FS48 suggest that, prior to that, a firm (but mistaken) belief in the principle of diminishing marginal utility may have been a significant impediment to its acceptance[2]. Whatever the precise historical reasons for disinterest, it is clear that economists became much more interested in the theory following the demonstration by von Neumann and Morgenstern (1947) that the expected utility hypothesis could be derived from a set of 'appealing' preference axioms.

It will be useful for what follows to provide a thumbnail sketch of key ingredients in a typical axiomatisation. A common approach has been to model risky alternatives as 'prospects' or lotteries. I will use bold case letters to represent prospects and any prospect $\mathbf{p} = (p_1, x_1; \dots \dots; p_n, x_n)$ is a probability distribution over a fixed set of consequences where p_1 represents the probability of consequence x_1 and so on[3]. The expected utility hypothesis can then be derived from three axioms on preferences over prospects: ordering, continuity and independence. Ordering requires there to be a complete and transitive pairwise ranking of prospects[4]; the addition of a continuity principle[5] is then sufficient to guarantee that preferences can be represented by *some* function V(.) which attaches a real-value to every prospect[6]. The independence axiom of EUT[7] adds a restriction of the following form. Consider two compound prospects *q'= (q, p; s, 1-p) and r' = (r, p; s, 1-p)* where *q'* is itself a probability mixture of two prospects; it results

---

[1] For compactness I will refer to these two papers as FS48 and FS52 respectively.

[2] "The rejection of utility maximisation as an explanation of choices among different degrees of risk was a direct consequence of the belief in diminishing marginal utility. If the marginal utility of money diminishes, an individual seeking to maximise utility will never participate in a "fair" game of chance…… But this implication is clearly contradicted by actual behavior. (Para break) Marshal resolved this contradiction by rejecting utility maximisation as an explanation of choices involving risk. He need not have done so, since he did not need diminishing marginal utility – or, indeed, any quantitative concept of utility- for the analysis of riskless choices". (FS48, 280).

[3] Take it throughout that, for all prospects, $p_i \geq 0$ for all i, and $\Sigma_i\, p_i = 1$.

[4] More precisely, this axiom entails that: (1), for any pair of prospects $\mathbf{q}$, $\mathbf{r}$: either $\mathbf{q} >\sim \mathbf{r}$ or $\mathbf{r} >\sim \mathbf{q}$ or both where $>\sim$ represents the relation "is (weakly) preferred to"; and (2) for any three prospect $\mathbf{q}$, $\mathbf{r}$, $\mathbf{s}$, if $\mathbf{q} >\sim \mathbf{r}$ and $\mathbf{r} >\sim \mathbf{s}$, then $\mathbf{q} >\sim \mathbf{s}$.

[5] Continuity requires that for all prospects $\mathbf{q}$, $\mathbf{r}$, $\mathbf{s}$ where $\mathbf{q} >\sim \mathbf{r}$ and $\mathbf{r} >\sim \mathbf{s}$, there exists some probability p such that there is indifference between the middle ranked consequence *r* and the prospect $(\mathbf{q}, \text{p}; \mathbf{s}, 1\text{-p})$.

[6] V(.) represents preferences in the sense that $V(\mathbf{p}) \geq V(\mathbf{q})$ if, and only if, prospect $\mathbf{q}$ is not preferred to $\mathbf{p}$.

in **q** with probability p otherwise **s**. Similarly, **r'** is a mix of **r** and **s** and notice that the probability of **s** is the same (i.e. 1-p) for both compound prospects. Independence implies that the common component of the two compound prospects (in this case the 1-p chance of **s**) is irrelevant to their relative ranking which should depend purely on the ordering of the simple prospects *q and* **r**[8]. Given this additional assumption, preferences can be represented by the familiar expected utility function $V(\mathbf{p})=\Sigma_t p_i.u(x_i)$ where **p** is any prospect, and u(.) is a 'utility' function defined over consequences.

Where lies the alleged appeal of this set of axioms or, equivalently, the appeal of EUT so axiomatised? One, I think enlightening, way to reconstruct it is as follows. The standard method of modern economics has been to understand behaviour through the lens of optimisation, that is, to assume that agents are rational actors moved along by coherent and stable preferences. Hence, to assume the existence of some preference function over prospects $V(\mathbf{p})$, is essentially to apply the standard method of economics to decisions over risks. But what justifies, from a normative point of view, the added restriction of independence?

A classic answer is provided by Samuelson (1952) who argues firmly that the type of independence condition entailed by EUT is itself a compelling normative principle of rationality. The nub of the argument is essentially this. Suppose you have to choose between *q'* and *r'* as defined above and suppose, for the sake of illustration, that p=0.5 with the compound lotteries resolved by a coin flip which determines the outcome to be **s** if 'tails' comes up. Samuelson argues that, if tails does come up, you won't care which option you chose (because you get **s** either way) so you can safely ignore this common element of the prospects when choosing between **q'** and **r'**, just as independence requires. Once it is conceded that the outcomes associated with tails can be ignored, it then seems plain that an agent should choose between the compound prospects on the basis of their ordering over the simple prospects **q** and **r**.

---

[7] Whenever I refer to 'independence' principles later in this paper, note that I have in mind Expected Utility Independence conditions and not other types of independence condition that feature elsewhere in economic theory.

[8] More formally, the independence axiom of EUT entails that for all prospects **q**, **r**, **s**: if **q** >~ **r** then (**q**, p; **s**, 1-p) >~ (**r**, p; **s**, 1-p), for all p.

The argument for ignoring the tails outcome, and for (EUT) independence more generally, turns on the proposition that there can be no (rationally justifiable) *complimentarity* between the outcomes within a prospect because they are *mutually exclusive*. If this is conceded, and independence is accepted as an implication of rationality, then EUT has much more significance than simply being just one amongst many possible models of risk preference; EUT then has a claim to be interpreted as the logical extension of rational economic analysis to the realm of risk. Whether or not individual axioms of EUT can be defended as requirements of rationality has, of course, been a matter of much debate[9]. Fortunately, we need not enter these tricky debates because my primary concern will be to examine what follows granting, for the purpose of the argument, that the axioms of EUT *can* be taken as appealing principles of rationality.

To the extent that its axioms can be justified as sound principles of rational choice to which any reasonable person would subscribe, the axioms provide grounds for interpreting EUT *normatively;* that is as a model of how people ought to choose. Some writings have placed emphasis on this normative interpretation of EUT. For example, Savage (1954) presents what has become one of the most celebrated derivations of EUT explicitly as an attempt to extend logical reasoning to situations of uncertainty. His primary aim is not to provide an empirical theory for predicting human behaviour, but instead to develop logical tools for deciding between alternative courses of action:

> "Decisions made in the face of uncertainty pervade the life of every individual and organisation. Even animals might be said continually to make such decisions, and the psychological mechanisms by which men decide may have much in common with those by which animals do so. But formal reasoning presumably plays no role in the decisions of animals, little in those of children, and less than might be wished in those of men. It may be said to be the purpose of this book, and indeed of statistics generally, to discuss the implications of reasoning for the making of decisions." p6

That said, Savage does accept that EUT may also have some potential as a simple – if 'crude' - empirical theory for predicting human behaviour, albeit in 'suitably limited domain'[10]. It is to the interpretation of EUT as an empirical theory that I now turn.

EUT as an empirical hypothesis

In two now classic papers co-authored by Friedman and Savage (FS48, FS52) – and published shortly before Foundations and F53 - there is much more focus on the empirical interpretation of EUT. Both papers interpret EUT explicitly as an empirical hypothesis and present these papers as 'crude' tests of it. The tests are crude, partly because of the limited supply of direct evidence suitable for testing the theory. Nevertheless, the two papers propose complementary strategies for crude testing.

So how does EUT stand up to empirical scrutiny? In FS48 they argue that, given a suitable specification of the utility function over wealth, the theory can explain a set of stylised facts about how people respond to risk, particularly in the context of gambling and insurance. They present a set of five stylised facts to be explained and they propose a functional form for the utility of wealth that is shown to be consistent with the presented facts. A crucial feature of their proposal is that the utility function contains both concave and convex segments.   As such it allows the same individual to express both risk averse (insurance) and risk seeking (gambling) behaviour – one of the key stylised facts they set out to explain.

It is the argument of the second paper (FS52), however, that will be my primary focus. This paper also presents EUT as an empirical hypothesis to be evaluated relative to the 'evidence', but it brings a new and interesting twist by introducing an explicit distinction between *direct* and *indirect* evidence. Direct evidence comes from comparing the implications of theory with observations generated within the

---

[9] For the reader wishing to explore these debates a good place to start is Sugden (1991).

intended domain of application, while indirect evidence relates to phenomena that are not in the 'primary' domain of interest. So, to the extent that EUT is intended to explain market phenomena such as gambling and insurance decisions, FS48 is in good part an evaluation of EUT on the basis of direct evidence.

In contrast, the main purpose of FS52 is to articulate indirect support for EUT. An interesting feature of the argument is that it seeks to licence reading the normative appeal of EU axioms as a source of empirical support. I will refer to this manoeuvre as the *FS-twist.* The argument goes like this. There is, FS52 argue, "indirect evidence rendering it plausible that the hypothesis will continue to fail to be contradicted, in at least some important domains" (FS52, 466). The primary source of this evidence[11] is "the plausibility of a set of postulates that are sufficient for derivation of the hypothesis" p466. The postulates are a set of axioms essentially equivalent to those discussed above. The question I wish to pursue here is how 'the plausibility of postulates' translates into (indirect) empirical support for EUT? At first blush this seems slightly mysterious, but some insight is provided by the following passage from FS52:

> "In saying that these postulates are more plausible than the hypothesis to which they are logically equivalent, we mean that the postulates immediately call to mind a host of implications (or predictions from the hypothesis) susceptible to casual observation. With respect to the class of phenomena to which these implications relate, the hypothesis has had many opportunities for contradiction and has repeatedly failed to be contradicted. The evidence is indirect because this is not the class of phenomenon we are primarily interested in using the hypothesis to predict" (p466)

The suggestion that the "postulates immediately bring to mind a host of implications" seems very reasonable at least partly because the axioms make explicit particular formal properties of EUT which may be less than transparently obvious from a statement of the hypothesis. For instance, once I see and

---

[10] " ….[EUT] can be interpreted as a crude and shallow empirical theory predicting the behavior of people making decisions. This theory is practical in suitably limited domains, and everyone in fact makes use of at least some aspects of it in predicting the behaviour of others". p20.

comprehend a statement of the independence principle, it then becomes clear that it is implicit in EUT that common components of prospects under consideration should be irrelevant to my choice among those prospects. The argument, however, involves two further claims that strike me as less obviously justified. One is the assertion that implications which the postulates bring to mind are 'susceptible to casual observation', the second is the assertion that such casual observation has broadly supported the hypothesis.

A difficulty arises in relation to both of these latter assertions because it is not obvious what to understand by 'casual observation' in this context. Perhaps the most natural interpretation would be to think that casual observations refer to instancess of behaviour thrown up by choices that people happen to make in the world. But if we put aside, for the moment, observations from specifically designed experimental tests of axioms, it is doubtful that the world throws up much data bearing on the validity of EUT axioms. An important reason for this is that the axioms are (for the most part) propositions about *consistency* between choices and as such they do not typically tell us how individuals will choose in specific decision problems. Because of this, it is far from obvious that casual observations of decisions in the world generate rich opportunities for testing EUT axioms. For example, if we know that an agent prefers some prospect **a** over **b**, and **b** over **c**, we can test transitivity if we can also observe their preference between **a** and **c**. But it seems to me that casual opportunities for collecting such observations are rather infrequent.

Another possibility is that 'casual observation' is to be understood as arising from introspective thought experiments, that is from personal reflections of roughly the form: would I violate the transitivity principle in this setting; or would I violate the independence principle in that setting; etc, etc? Such introspections, however, could hardly be regarded as providing objective tests of what the agent's *behaviour* would be. If an agent is, or tends to become convinced of the proposition that 'I would not deliberately violate choice principle x', it seems that further introspections of the form 'would I violate x

---

[11] They also suggest as an aside that "Coherence with rest of economic theory" (p466) may count as

in this specific case' can no longer be considered independent observations. To the extent that the agent is concerned to reason coherently, it seems the cognition that 'I would not deliberately violate choice principle x' will move them to conclusions consistent with this, in thought experiments.

There is, however, another potential argument articulated in FS52 to motivate the FS-twist. This works in two steps. The first is to argue that the axioms are "introspectively very appealing" (p468) and seeks to persuade the reader that each principle "is not one he would deliberately violate"[12] (FS52, p4690. It seems reasonable to suppose that many people would so endorse the principles. The second step is to assert that the introspective appeal of the axioms is "some reason for supposing that people do actually tend to avoid flagrant violation of the principle"[13] (FS52, p469). This second step appears to involve a questionable leap from propositions about the intuitive appeal of choice principles to propositions about choice *behaviour*. In order to make this leap explicit, consider propositions P1 and P2:

(P1) the agent accepts axiom x as a principle they would not deliberately violate

(P2) the agent probably behaves in accordance with axiom x

As I read it, this part of the FS52 paper essentially asserts that because P1 holds, P2 is likely. But while P1 and P2, are not inconsistent, in the absence of further premises, P2 is not implied by P1. P1 is a proposition about normative judgements while P2 is a proposition about behaviour. In order to generate the implication suggested by FS52 it is necessary to introduce some premise linking normative beliefs with behaviour. For example:

---

indirect evidence for EUT

[12] While FS52 discuss each of the EU axioms, a key part of their argument is concerned with justifying the independence axiom. This seems natural given that this was and continues to be the most controversial element of the EUT system: it is also the assumption that gives EUT most of its empirical content.

[13] Although the claim here seems a moderate one given the qualification that normative appeal is just 'some' reason, I take it there is meant to be a substantive claim here given that the main argument of the paper seeks precisely to use normative appeal as a source of evidence.

(P3) agents rarely behave in contradiction with principles they believe they would not

deliberately violate

Notice, however, that it is not enough just to assume P3: for it to do the job of converting normative appeal to evidence, it has to be empirically valid or at least plausible. How would one assess whether this is an empirically plausible claim? I can think of at least two possibilities. One would be to refer to a theory of the choosing mind. If one could point to some model of human mental processes which implied that normative beliefs govern choice behaviour, and show that there is satisfactory empirical support for it, that would be one way to motivate a principle like P3. This strategy, however, would involve a major departure from the 'as if' methodology famously set out in F53.

It is clear that Friedman and Savage do favour an 'as if' interpretation of EUT. For instance F48 provides this methodological aside with respect to evaluation of EUT:

"An objection to the hypothesis just presented that is likely to be raised by many, if not most, readers is that it conflicts with the way human beings actually behave and choose. Is it not patently unrealistic to suppose that individuals consult a wiggly utility curve before gambling or buying insurance, that they know the odds involved in the gambles or insurance plans open to them, that they can compute the expected utility of a gamble or insurance plan, and that they base their decision on the size of the expected utility?

While entirely natural and understandable, this objection is not strictly relevant. The hypothesis does not assert that individuals explicitly or consciously calculate and compare expected utilities. Indeed it is not at all clear what such an assertion would mean or how it could be tested. The hypotheses asserts rather that, in making a particular class of decisions, individuals behave as if they calculated and compared expected utilities and as if they knew the odds. The validity of this assertion does not depend on whether individuals know the precise odds, much less on whether they say that they calculate and compare expected utilities or think that they do, or whether psychologists can uncover any evidence that they do, but solely on whether it yields sufficiently accurate predictions about the class of decisions with which the hypothesis deals. Stated differently, the test by results is the only possible method of determining whether the as if

statement is or is not a sufficiently good approximation to reality for the purpose at hand." (FS48, p298)

This clearly has a great deal of resonance with the methodological position Friedman sets out in F53: we should not count it against EUT that its assumptions appear 'unrealistic'; EUT is not to be interpreted as a model of conscious human decision processes but as an 'as if' model with the purpose of predicting behaviour; as such, the only relevant test of the theory is taken to be its predictive performance, though note the important caveat that; predictive performance is to be judged relative to the theory's domain of application.

This 'as if' strategy entails that theories *not* be judged in terms of whether they are defensible models of mental processes. So to invoke a model of mental process as a defence of the theory would does not seem to provide an interpretation of the FS-twist which is consistent with F53 (or FS48).

Another possibility would be to interpret P3 as a purely empirical (but untheorised) principle. This, I suggest, provides an interpretation which is more consistent with F53 but it is not unproblematic. Part of the problem with this reading is that evidence from behavioural research runs against a general claim that peoples' behaviour satisfies principles that most would take to be normatively appealing. To support that claim, here are two illustrations.

1. Savage and the Allais paradox

A significant part of the debate about independence has centred on the well-known Allais paradox. Consider the following action/state payoff matrix where each row represents one of four risky actions (g1 to g4), and columns represent possible states of the world (with probabilities given at the top of each column). The numbers in the matrix are to be read as the state contingent payoffs to each (in, say, 000s $).

	0.1	0.01	0.89
g1:	500	500	500
g2:	2500	0	500
g3:	500	500	0
g4:	2500	0	0

Choice 1 { g1, g2 }
Choice 2 { g3, g4 }

Consider a choice between g1 and g2. The independence axiom implies that since these two actions give the same consequence in the third state of the world, that third state must be irrelevant to that choice. The same argument applies to the choice between g3 and g4. Notice that if the third column is blanked out, the choice between g1 and g2 is identical to the choice between g3 and g4 (i.e. 0.11 chance of 500 vs 01. chance of 2500). Hence, independence implies that if g1 is (not) preferred to g2, g3 is (not) preferred to g4. There is considerable evidence that many people faced with pairs of choices with this general structure choose g1 over g2 and g4 over g3 in violation of independence.

It is a matter of historical interest that Savage himself, in Chapter 5 of Foundations conceded that he also violated independence when first confronted with the Allais paradox. He reports that having once recognised the inconsistency, he reflected upon the situation and came to the conclusion that his belief in the normative appeal of independence was unshaken. Consequently, he determined that he must have made some error in one of his initial choices. Re-examining the choices Savage applies independence hence reducing both decisions to the choice (0.11 chance of 500) vs (01. chance of 2500). Then, "Finally, consulting my purely personal taste, I find that I would prefer the gift of $500,000 and, accordingly, that I prefer Gamble 1 to Gamble 2 and (contrary to my initial reaction) Gamble 3 to Gamble 4." (Savage, 1954, p103). On Savage's own account, his initial choice of g4 over g3 was an error which the application of normative reasoning allowed him to detect and correct.

This account seems unproblematic in relation to the argument of S54. In that context, Savage is explicitly concerned with developing a normative theory which may be used to 'police decisions'. The

11

fact that real decisions may depart from the normative is a pre-requisite for there to be any interesting policing role for his axioms. However, the example provides a counter argument to the proposition that being normatively committed to a decision principle implies conforming behaviour and the example seems especially compelling when the person concerned is an eminent decision theorist violating a principle which is the most important in his own normative theory.

2. Framing Effects

One possible defence against this example would be to suggest that Savage's mistake lay not in his choice, but in his adoption of independence as normative. An alternative interpretation of the Allais paradox is that it shows that independence is not a compelling normative principle (indeed, this is Allais' own reading). Someone who took this line would have the option of arguing that the Allais paradox does not demonstrate any inconsistency between compelling normative principles and choice behaviour. That argument, however, could be quickly dismissed by pointing to violations of *description invariance*. By description invariance I mean the proposition that preferences between choice alternatives should not be altered by redescriptions of the alternatives which leave the objective choice set unchanged. I take this to be a very basic proposition implicit in normative decision theories (so basic, that it is rarely explictly stated as a proposition). There is, however, considerable evidence that very minor changes in the presentation or "framing" of decisions can have dramatic impacts upon the choices of decision makers: such effects are failures of *description invariance*. Here is one famous example due to Tversky and Kahneman (1981) in which two groups of subjects -call them groups I and II - were presented with the following cover story:

> "Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programmes to combat the disease have been proposed. Assume that the exact scientific estimate of the consequences of the programmes are as follows:"

Each group then faced a choice between two policy options. One group chose between A and B:

> "If programme A is adopted, 200 people will be saved.
> If programme 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."

12

The other chose between C and D:

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

The two pairs of options are stochastically equivalent. The only difference is that the group I description presents the information in terms of *lives saved* while the information presented to group II is in terms of lives lost. Tversky and Kahneman found a very striking difference in responses to these two presentations: 72 percent of subjects preferred option A to option B while only 22 percent of subjects preferred C to D. It seems hard to deny that different behaviour in the two problems is normatively unsound. Yet similar patterns of response were found amongst groups of undergraduate students, university faculty and practicing physicians. Real people it seems, do not in general behave consistently with normative principles, even those that find very common ascent.

Market selection and the FS-twist

Taken at face value, the evidence from simple choice experiments seems to tell against the FS-twist. Yet, one might accept that it is possible to construct experimental situations where people violate normative principles, but raise doubts about whether such violations would be common features of the wider world. One reason for such doubt might be connected with Friedman's famous 'natural selection' argument. In F53, he argues that part of what may explain the success of the profit maximisation hypothesis is that market forces will promote the survival (demise) of firms using strategies clo bsest to (furthest from) profit maximisation. The predictive success of the theory of profit maximisation then relies partly on features of market processes, not explicitly modelled in the economic theories of firms and markets.

Analogous arguments could be constructed to suggest that EUT might have better predictive success in a market place, relative to laboratory tests, perhaps by virtue of the existence of market mechanisms that punish deviations from it. Indeed, widely rehearsed 'money pump' arguments seem to

13

have just this character. Such arguments typically work by suggesting that agents who fail to satisfy normative decision principles lay themselves open to exploitation. Consider, for example, an agent who has non-transitive preferences of the form $q >\sim r$, $r >\sim s$ and $s >\sim q$. Suppose they are initially endowed with q. Repeatedly trading according to these binary preferences could lead them to: (i) pay to swap q for s; (ii) pay to swap s for r; and (iii) pay to swap r for q, and so on. This agent has then been money pumped because they have paid for several trades only to find themselves returned to their original endowment (minus some cash).

One might argue that if agents were exploitable in this way, we should expect to see other people exploiting them. The lack of evidence of money pumps in operation in the world might then be taken as evidence that on the whole people must have preferences conforming at least roughly with normative presumptions most of the time. I think these arguments unconvincing as attempts to suggest empirical support for EUT axioms not least because careful theoretical analysis seems to defeat the claim – central to money pump arguments - that agents will be money-pumpable unless their preferences satisfy EUT. Contrary to this view, Cubitt and Sugden (2001) show that agents can have preferences which satisfy few, if any, of the standard normative criteria, yet still be money-pump proof.

But even if money pump arguments fail as a convincing blanket defence of EUT, it is not entirely unreasonable to suppose that some market mechanisms might actually promote conformity with EUT. While it is currently hard to find much direct support for that conjecture, there is evidence that at least some preference anomalies do appear to decay in the context of some market mechanisms. One of the clearest examples relates to contingent valuation methodology: a standard methodology for assigning money values to non-marketed goods. The two most widely used valuation measures are willingness to pay (WTP) and willingness to accept (WTA). Standard preference theory based on indifference curve analysis implies that these two measures should generate similar valuations, but in practice WTA is typically much bigger that WTP. But further research has shown that the gap between WTP and WTA

tends to be eroded as a consequence of exposure to some market mechanisms, particularly the second-price variant of the Vickrey auction (see, for example, List and Shogren, 1999; Shogren et al, 2001).

While the existing literature suggests that more than one causal mechanism may lie behind the convergence of WTA and WTP in repeated auctions, there is some basis for thinking that part of the explanation involves subjects learning to behave more consistently with principles of rationality. For example, one factor sometimes cited as a possible account of the WTA/WTA disparity is that bidders may approach a market with a tendency to behave "cautiously" by underbidding when buying and over-asking when selling relative to their true values. Such a tendency may be rationally justified in some market contexts, for instance in markets where prices are negotiated through some process of haggling. In Vickrey auctions, however, revealing true value is a weakly dominant strategy and bidders operating cautiously may miss valuable opportunities to trade (e.g. by asking too high and then failing to sell at a price which is above their value). A possible explanation of the convergence of WTP and WTA is that missed opportunities of this sort cause agents to reflect on their behaviour and consequently adjust towards the optimal bidding strategy. Some support for this explanation of convergence has been found in a study conducted by Loomes, Starmer and Sugden (2003).

In principle, then, selection reasoning is a genre of argument that might work to forge, what seems to me, a missing link in the FS-twist argument of FS52 by giving some account of how normatively appealing principles may come to be manifest in market behaviour even though they are widely violated in simple choice experiments. While Friedman's use of this kind of reasoning in F53 suggests that he may have been sympathetic to this strategy, the text of F52 provides no obvious indication that Friedman and Savage had such a link in mind in this context. Even if they did, the link would remain weak in the absence of more explicit (theoretical and/or empirical) support for selection processes in markets specifically promoting behaviour consistent with EUT.

Normative notions in descriptive dialogues

Let us take stock. FS52 seek to argue that because EUT can be restated in terms of normatively appealing axioms, that provides a source of indirect support for the model as an empirical theory of behaviour. This is what I have called the FS-twist. I think that is a questionable claim for at least three sets of reasons. First, the connection from norm to decision, at least in FS52, seems mysterious. Second, while it might be possible to articulate some connection, say by specifying a model of decision process with that feature, my impression is that the literature which does seek to model decision processes would typically point in the opposite direction: in general, models of decision process provide many varied reasons for thinking that behaviour will deviate from normatively appealing criteria[14]. Thirdly, I have pointed to experimental evidence against the presumption that normative appeal implies empirical validity. At the end of the last section I did concede that selection arguments might in principle provide a missing step. The link made in that way, however, would remain an extremely tenuous one, absent a clearer theoretical statement of it and an assessment of its empirical plausibility.

The argument made by FS52, is of more than passing historical interest however. Indeed, it seems to me that developments in the economic theory of risk behaviour, over recent decades, have been influenced in significant ways by a widely presumed (but I think mis-conceived) connection between normative principles and actual behaviour. For example, since the late 1970s, numerous alternatives to EUT have been proposed (largely motivated by empirical counter examples such as the Allais' examples)[15]. In these developments, it seems clear that economists have paid almost no attention to modelling phenomena that are inconsistent with principles that most economists take to be normatively compelling. For instance, although there is evidence that people violate principles like monotonicity and description invariance, economists have not been busy attempting to explain these findings.

---

[14] For example, in Payne, Bettman and Johnson's model, agents may draw on a wide variety of heuristics based on tradeoffs between 'accuracy' and 'effort'. But while here is an economising motive, there is no reason to expect the degree if coherence in choice implied by EUT. Similarly, in Gigerenzer and Selten (2000), agents are conceived of as adopting 'fast and frugal' heuristics.

16

Moreover, theories which have had the potential to explain violations of normative principles have been heavily criticised in the literature. For example, most would agree that the principle of monotonicity is a normatively appealing property in a theory of risky decision making. Simply put, monotonicity means that objective improvements to a prospect (e.g. increasing at least some of its payoffs holding others constant) should not reduce its attractiveness. Some non-expected utility theories, however, do predict failures of monotonicity, though in these cases the prediction has usually been a by-product of the model: an unintended consequence of the theoretical structure adopted to explain some other phenomenon.

One example is the model that extends EUT by attaching simple decision weights to probabilities such that agents are assumed to maximise a function of the form: $V(\mathbf{p}) = \Sigma w(p_i)u(x_i)$. This is essentially the type of value function assumed in prospect theory of Kahneman and Tversky (1979) and they show that the Allais paradoxes can be explained by non-linearities in $w(p_i)$. But this strategy for explaining the Allais examples has the side-effect that predicted choices among at least some prospects will violate monotonicity. This has generally been regarded as a fatal flaw sufficient to damn theories of this type. Arguments to this effect have been made by Fishburn (1978), Machina (1982), and Quiggin, (1982), each of them heavyweights in the arena of choice theory.

No doubt conscious of these potential criticisms, Kahneman and Tversky propose an editing heuristic which eliminates dominated options, so long as they are detected (the *dominance heuristic*). This strategy for inducing monotonicity, however, has attracted further criticism from some economists. Quiggin (1982, 1993), for example, has criticised the approach on two counts. First, he argues that by appropriate specification of the preference function the dominance heuristic can be rendered redundant. Second, he criticises the Kahneman and Tversky strategy for imposing monotonicity because it has the further spin-off effect that the theory then admits violations of transitivity in pairwise choice. Quiggin (1982, p327) describes this an "undesirable result".

---

[15] I review some of these developments in Starmer (2000)

What is the basis of this latter claim?   It is not I contend, based on sound empirical evidence. First of all there is well established evidence - Tversky (1969) produced some of the earliest - that cyclical choice is a robust and reasonably general phenomenon (for later evidence see Loomes et al (1989,91)). It is true that the form of intransitivity predicted by prospect theory is of an unusual kind and distinct from other forms of intransitivity. That, however, presents a useful opportunity for testing a novel prediction of prospect theory.   Starmer (1999) reports an experiment which tests for the specific form of intransitivity implied by prospect theory and finds very strong evidence of it.

Notwithstanding this evidence, the economics literature continues to be dominated by models built on transitivity and other empirically questionable but normatively attractive principles.   In light of this, I am inclined to the view that economists' reactions to evidence, in the theory of choice under risk, have been mediated by certain theoretical pre-commitments.   In particular, a pre-commitment to preference theories which satisfy normatively appealing criteria such as monotonicity etcetera. This pre-commitment has, in my view, delimited certain problems as interesting (i.e. those that appeared at least potentially soluble in terms of the pre-committed assumptions) and others as uninteresting or outside the realm of economic enquiry.

If it is granted that such a pre-commitment has been at work, is that bad thing?   I am not certain that it is. Some philosopher's of science, among them Kuhn and Lakatos, provide strong arguments for thinking that some kinds of theoretical pre-commitments may be pre-requisites for healthy, ongoing, scientific enquiry.   On a Lakatosian interpretation for instance, principles of rational choice like monotonicity and transitivity, might be taken as part of the hard core of the research programme of choice under risk.   What is questionable, however, is any tendency to mistake such a pre-commitment for evidence in favour of it.

Conclusion: reflections on positive economics

Friedman's 1953 essay is a remarkably rich and enduring piece. And, while it is long, and complex, and deals with many subtle issues, it sets an apparently clear line on certain methodological principles: positive economics is fundamentally empirical and should involve confronting theories with evidence from the world; the predictive ability of a theory (in its intended domain of application) is the key test of a theory's success; attempting to assess the realism of assumptions is not a meaningful test of any theory.

While F53 is primarily an exercise is abstract normative methodology, the papers which have been the primary focus of this discussion (FS48 and FS52) are *methodology in action* in so far as they involve explicit attempts to evaluate a particular theory, specifically, EUT. The contrast between the abstract and applied methodology is both striking and interesting. The abstract F53 essentially bids the economist, qua positive scientist, ask one primary question of a proposed theory: *how well does it fit the facts in its intended domain of application*? When it comes to the theory of risk, however, while Friedman and Savage explicitly endorse the key principles of positive economics (a la F53), its application appears complicated by the scarcity of a crucial ingredient: that is, direct evidence bearing on the predictive success of EUT in domains of intended application. This looks like a significant hurdle to get over if Positive Economics is to work in the domain of risky choice behaviour. I suggest that both FS48 and FS52 can be read as attempts to respond this, albeit in rather different ways. I have focussed on the argument of F52 not least because that paper involves strategies that might seem to present marked contrast with the prescripts of F53 such as: appeals to indirect evidence; to casual observation; to introspection and to the intuitive appeal of axioms.

The simple fact that economics in action (that is the business of constructing and evaluating theories of economic behaviour) turns out to be more complex and messy than Positive Economics a la F53 need not necessarily undermine faith in its key methodological assertions. All must surely accept that

theorising necessarily involves simplification and abstraction from complex reality whether it be theory about the ideal behaviour of an economy, or theory about the ideal behaviour of economists, or whatever. Yet the persuasiveness of Friedman's methodology might be reduced to the extent that his own economics in action were plainly at odds with the prescriptions of F53.

An obvious question then is whether the FS-twist ultimately coheres with the Positive Economics of F53? One difficulty in assessing this arises from the fact that the FS-twist is not fully articulated in FS52: that is, the paper does not provide a clear account of why we should expect behaviour to conform with normatively appealing axioms. I have argued that there may be different ways of forging such a link and I have discussed two particular possibilities: one way would be to develop a theory of choosing mind, another would be to invoke a theory of selection. I have suggested that the second of these routes would provide a reading of FS52 which coheres much more naturally with F53. On this reading, does the FS-twist then provide a satisfactory empirical defence of EUT? I think not as it stands, because it begs a further question about whether there are, in fact, selection mechanisms which operate in domains of interest to promote conformity between behaviour and normative decision principles. I take it that this, in turn, is a question susceptible to positive economic analysis and not something simply to be taken on faith.

References

Appleby, L. and C. Starmer (1987), "Individual Choice Under Uncertainty: A Review of Experimental Evidence, Past and Present, in J. Hey and P. Lamert (eds) Blackwell.

Bernoulli, Daniel. (1954), "Exposition of a New Theory on the Measurement of Risk," (original 1738) *Econometrica*, 22, pp. 23–26.

R. Cubitt and R. Sugden (2001), "On money pumps", *Games and Economic Behavior*, 2001, Vol. 37, pp. 121-160.

Fishburn, P. C. (1978), "On Handa's 'New Theory of Cardinal Utility' and the Maximization of Expected Return," *J. Polit. Econ.*, 86, pp. 321–24.

Friedman, M. (1953), "The Methodology of Positive Economics", in *Essays in Positive Economics* Chicago University Press.

Friedman, F. and L. Savage (1948), "The utility Analysis of Choices Involving Risk", *Journal of Political Economy, LVI, pp.279-304.*

Friedman, F. and L. Savage (1952), "The Expected-Utility Hypothesis and The Measurability of Utility", *Journal of Political Economy,* LX, pp.463-474.

Kahneman, D. and A. Tversky (1979), "Prospect Theory: An Analysis of Decision under Risk," *Econometrica,* 47:2, pp. 263–91.

Gigerenzer, G. and Selten, R., eds. (2000). *The Adaptive Tool Box.* Cambridge, MA/London: MIT Press.

List, J. and J. Shogren (1999), "Price Information and Bidding Behavior in Repeated Second-Price Auctions", *American Journal of Agricultural Economic*, 81, 942-49.

Loomes, G., C. Starmer and R. Sugden (1989), "Preference Reversal: Information-Processing Effect or Rational Non-transitive Choice?" *Economic Journal*, 99, pp. 140–51.

Loomes, G., C. Starmer and R. Sugden (1991), "Observing Violations of Transitivity by Experimental Methods," *Econometrica*, 59, pp. 425–39.

Loomes, G., C. Starmer and R. Sugden (2003), "Do anomalies disappear in repeated markets?", *Economic Journal*, 113, C153-66.

Machina, M. (1982), "'Expected Utility' Theory without the Independence Axiom," *Econometrica,* 50, pp. 277–323.

Neumann, J von and O.Morgenstern. (1947), *The Theory of Games and Economic Behaviour*. 2nd ed. Princeton: Princeton U. Press.

Payne, J., J. Bettman, and E. Johnson. (1993), *The Adaptive Decision Maker*. Cambridge U. Press.

Quiggin. (1982), "A Theory of Anticipated Utility," *J. Econ. Behavior Org.*, 3:4, pp. 323–43.

——— (1993), *Generalized Expected Utility Theory*. Dordrecht: Kluwer.

Samuelson, P. A. (1952), "Probability, Utility, and the Independence Axiom", *Econometrica*, 20, pp. 670-8.

Savage, L. (1954), *The Foundations of Statistics*. NY: Wiley.

Shogren, J., S. Cho, C. Koo, J. List, C. Park, P. Polo and R. Wilhelmi (2001), "Auction mechanisms and the measurement of WTP and WTA", *Resource and Energy Economics*, 23, 97-109.

Starmer, C. (1999), "Cycling with Rules of Thumb: An Experimental Test for a New Form of Non-Transitive Behaviour," *Theory and Decision*, 46, pp. 141–58.

Starmer, C. (2000), "Developments in non-expected utility theory: the hunt for a descriptive theory of choice under risk", *Journal of Economic Literature*, XXXVIII, 332-382.

Sugden, R. (1991), "Rational Choice: A Survey of Contributions from Economics and Philosophy," *Economic Journal*, 101, pp. 751–85.

Tversky, A. (1969), "Intransitivity of Preferences," *Psych. Rev.*, 76, pp. 31–48.

Tversky, A. and D. Kahneman. (1981), "The Framing of Decisions and the Psychology of Choice," *Science*, 211, pp. 453–58.

——— (1986), "Rational Choice and the Framing of Decisions," *J. Bus.*, 59:4, pt. 2, pp. s251–78.