Becker's Thesis and Three Models of Preference Change

Richard Bradley
Department of Philosophy, Logic and Scientific Method
London School of Economics
Houghton Street
London WC2A 2AE
r.bradley@lse.ac.uk

June 2, 2008

Abstract

This paper examines Becker's thesis that the hypothesis that choices maximise expected utility relative to fixed and universal tastes provides a general framework for the explanation of behaviour. Three different models of preference revision are presented and their scope evaluated. The first, the classical conditioning model, explains all changes in preferences in terms of changes in the information held by the agent, holding fundamental beliefs and desires fixed. The second, the Jeffrey conditioning model, explains them in terms of changes in both the information held by the agent and changes in her prior beliefs, holding her fundamental desires fixed. The final model, that of generalised conditioning, allows for explanations in terms of changes in the values of all three variables.

1 De Gustibus Non Est Disputandum

Gary Becker famously described the economic method as "... the combined assumptions of maximising behaviour, market equilibrium and stable preferences, used relentlessly and unflinchingly" [2]. As a definition of economics, this characterisation is undoubtedly too narrow: for instance, many economists work on non-market interactions and/or use models which assume bounded rationality. It may have more legitimacy as a description of prevailing practice in neoclassical economics, at least with regard to the assumption of stable preferences, but even in this respect Becker's claim is a little misleading, or can mislead if one doesn't pay attention to actual economic modelling, because economists do not typically assume that *all* preferences are stable. Rather some subset of agents' preferences, namely those defined with respect to a set of final outcomes, are fixed by the model, while others are allowed to vary (in principle at least). An agent's preferences over her option set, for instance, are bound to vary if the information set she uses to determine the expected utilities of these options changes for some reason. And however important models assuming perfect information or fixed information sets may have been historically no economist would regard them as exhaustive of their discipline. Furthermore, most economists would regard the assumption of stable final preferences as a methodological heuristic that simplifies the modelling problem and renders it susceptible to mathematical treatment rather than as a substantial empirical hypothesis. The thought would be that if a model is constructed sensibly the outcomes that it treats as final will be ones with respect to which preferences can be assumed to be sufficiently stable for the purposes at hand.[1]

[1] See Hausman [10] for further discussion of these issues.

There remains the question of whether there is some level of description of possible outcomes or states of the world with respect to which agents' preferences are truly invariant. Both positive and negative views on this question abound. Classical Utilitarianism seems to assume, for instance, that preferences for pain over pleasure should be both universal and stable: this is what makes quantities of pleasure an appropriate currency for moral accounting. Where arguments arise, and this surely is what matters in practical terms, is whether what makes for pleasure and pain stays the same. If classical Utilitarians tended to assume that it did, we find in Mill's idea of the cultivation of taste a recognition of their socio-historical variability. The tension between these views has outlived the dispute amongst Utilitarians. Who of us does not recognise both the force of the empirical evidence for disparate social and biological determinants of preference and the important normative role that the idea of ultimate preferences plays in practical deliberation and the evaluation of potential institutional arrangements?

Becker is, in many ways, the modern standard bearer for the classical Utilitarian position:

> "... one does not argue over tastes for the same reason that one does not argue over the Rocky Mountains - both are there, will be there next year, too, and are the same to all men" - (Becker and Stigler [1, p. 24]).

The tastes referred to here are what might be called *fundamental* preferences - i.e. preferences over the set of ultimate ends that Becker terms commodities - in contrast to *derived* preferences (and in particular those that are directly revealed in the choices that agents make) whose values derive from their relationship to the ultimate ends. The distinction between fundamental and instrumental value is a very common one, of course, not only in economics, but in the other social sciences and philosophy as well, and can be filled out with different specifications of the fundamental ends and of the nature of the relation between the derived values and the fundamental ones - the connection may be conceived as causal, evidential or even symbolic, for instance. What is distinctive about Becker's position is that it adds not only a detailed view about the connection between ends (commodities) and means (the production function), but also the hypothesis that fundamental preferences are both invariant and universal.

Appearances in the quotation not withstanding, Becker is not making a substantial or empirical claim about tastes, but rather a methodological one; the essence of which is that assuming fixed tastes disciplines the manner in which expected utility theory is applied to the explanation of behaviour. The explanatory strategy he advocates closely resembles that used by Bayesians to explain belief change; indeed the two are closely connected. Bayesians think of an agent's beliefs as being jointly determined by her prior beliefs and the information she holds and changes in her beliefs as being explained by changes in information, holding the priors fixed (in a sense which will be made more precise later on). Similarly Becker thinks that changes in the observed behaviour of an agent are to be explained in terms of changes in the factors determining her expectations of utility - information, stocks of personal and social capital, time, prices, etc. - relative to fixed tastes. The main advantage of this approach is that it tells researchers how to frame their explanations in a non ad hoc way i.e. in a manner which does not require the postulation of unverifiable taste changes.

It is not my intention in this paper to examine these methodological claims in their entirety. Rather my focus will be in the prior question of whether (or to what extent) Becker's claim that all variation in behaviour can be described in terms of changes in expectations relative to fixed tastes can be vindicated. To this end I will state his hypothesis in a somewhat more abstract language and then consider three different models of preference revision. All three conceive of an agent's choices, and more generally her preferences, as being (at least partially) determined by her beliefs and desires and her beliefs as being (at least partially) determined by the information she holds. These dependencies can be represented diagrammatically as in Figure 1.

Where the three model differ is with regard to the kinds of changes that they countenance. The first, the **classical conditioning** model, attempts to explain all changes in derived beliefs and preferences in

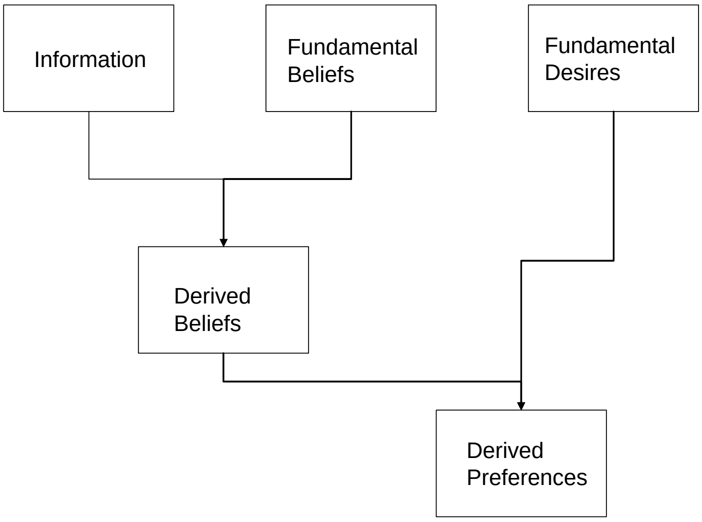

Figure 1: The Derivation of Preference

terms of changes in the information held by the agent, holding fundamental or prior beliefs and desires fixed. The second, the **Jeffrey conditioning** model, attempts to explain them in terms of changes in both the information held by the agent and changes in her prior beliefs, holding her fundamental desires fixed. The final model, that of **generalised conditioning**, allows for explanations in terms of changes in the values of all three variables.

My aim in this paper will be to examine the scope of the each model, attempting to specify in precise terms their domain of application. I will argue that the first model does not provide an adequate basis for explaining all preference change and hence that the fixed tastes hypothesis must be framed within a more general account of belief change than the classical Bayesian one. My conclusions with regard to the adequacy of the second model are much more tentative. I will argue that it is possible that all preference change can be explained in terms of Jeffrey conditioning, but only if the space of prospects is sufficiently rich. Finally, in the main result of the paper, I will prove that the third model is completely general in the sense that any change in an agent's attitudes can be explained within the framework that it provides.

2 Framework

States of Mind In the approach taken here we think of an agent's preferences for prospects as being determined, and hence explained, by her state of mind - her degrees of partial belief and desire. Prospects are the basic objects of agents' attitudes, both cognitive and conative. (Philosophers would typically refer to them as propositions, while economists tend to distinguish between objects of belief - events - and objects of desire - consequences. Nothing is at stake here in the choice of vocabulary). Prospects are modelled as subsets of a set of possible fundamental states of the world (or possible worlds), $\Omega = \{\omega_1,\omega_2,...\}$. The states in Ω play the role of the ultimate ends of agents; states whose descriptions are maximally specific with respect to all matters of concern to them and hence whose desirabilities are unconditional. The desirabilities of all (other) prospects, on the other hand, are conditioned by their relation to the fundamental states; roughly by the extent to which they render each of the fundamental possibilities more or less likely.

Let $\Gamma = \{A, B, ...\}$ be the set of all prospects (i.e. the set of all sets of possible states of the world), with Ω being the necessary prospect and $\varnothing$ being the impossible one. $\Gamma - \varnothing$ will be denoted by Γ' and $A \cap B$ shortened to AB. The agent's preferences over prospects are represented by a two-place 'at least as preferred as' relation $\succeq$ on Γ' which we will assume to be transitive, but not necessarily complete. The relations of strict preference, $\succ$, and indifference, $\approx$, are related to $\succeq$ in the usual way, i.e. $x \approx y$ means that $x \succeq y$ and $y \succeq x$, while $x \succ y$ means that $x \succeq y$ but not $y \succeq x$. Informally we may think of the agent's preferences, and their evolution over time, as providing the observations that require explanation.

The state of mind of an opinionated agent will be represented, as they are in Jeffrey [12], by a pair of real-valued functions $\langle p, v \rangle$, where p is a probability measure on Γ of her degrees of belief and v a normalised desirability measure on Γ' of her degrees of desire satisfying, for all $X, Y \in \Gamma'$ such that $XY = \varnothing$:

$$\begin{aligned} \text{(Normality) } v(\Omega) &= 0 \\ \text{(Averaging) } v(X \cup Y) &= \frac{v(X).p(X) + v(Y).p(Y)}{p(X) + p(Y)} \end{aligned}$$

The states of mind of less opinionated agents can be represented by sets of such pairs. For simplicity, however, we will work with a single pair at a time.

Normality reflects the thought that one should be inclined neither to promote nor to hinder something that is inevitable. But none of the conclusions of the paper depend on the normalisation of v with respect to the necessary prospect that Normality requires and it is assumed mainly to simplify the mathematics. The Averaging axiom plays a much more substantial role and captures the idea that the desirability of a prospect depends on the desirability of the various specific ways it could come to be realised weighted by the conditional probability of these ways given the prospect in question. Thus the desirability of taking a bus depends on the desirabilities of reaching my destination on time and of failing to do so, weighted by the probabilities of reaching or not reaching it on time, given that I take the bus.

A state of mind $\langle p, v \rangle$ explains an agent's preferences whenever it is the case that, for all X, Y in the domain of $\succeq$:

$$X \succeq Y \Rightarrow v(X) \geq v(Y)$$

The more information we hold about an agent's preferences, the more constraints they place on the class of states of mind that explain them. Under certain assumptions about the preference relation $\succeq$ - completeness and Bolker's averaging and impartiality conditions (see Bolker [3] and [4]) - both the existence of a state of mind explaining someone's preferences can be formally demonstrated and its uniqueness up to particular class of transformations. Since we make weaker assumptions about the nature of preferences, the existence of explanatory states of mind poses no problems here. They will not typically be unique however and very different representations of the agent's state of mind may be consistent with what we know about her preferences, even if they satisfy all the aforementioned conditions. To avoid the complications that this gives rise to, we will simply assume here that explanatory state of minds are unique up to a choice of scale for measuring her degrees of desire. Indeed, since the choice of zero has already been settled by the normalisation of v with respect to Ω, the choice of unit for v is the only remaining free parameter that we will need to worry about.

We assume throughout that Ω is countable and that all worlds have non-zero prior probability. In this case we can express the probability and desirability of any $X \in \Gamma'$ in the following manner:

$$\begin{aligned} p(X) &= \sum_{i \in X} p(\omega_i) && (1) \\ v(X) &= \sum_{i} v(\omega_i).p(\omega_i|X) && (2) \end{aligned}$$

These equations give formal expression to the idea that the agent will regard prospects as probable and desirable to the degree that they render one or another fundamental state, with given prior probability and desirability, more or less likely.

Changes of Mind Changes in an agent's preferences are to be explained by changes in her state of mind. In the models we consider below, a change in the agent's state of mind is viewed as a two-step process:

Stage 1 The agent changes her attitude to a particular prospect (or more generally across a partition of the space of prospects).

Stage 2 She adjusts her attitudes to all other prospects in order to restore consistency.

Formally a model of revision maps each prior opinionated state of mind, $\langle p, v\rangle$, to a posterior one, $\langle p^*, v^*\rangle$, as a function of the constraints yielded by stage 1. The process inducing the initial change is not itself modelled: these might include sensory experience, private or public deliberation, reception of a message from a reliable information source, or even hypnosis.

Methodological Theses We are now in a position to express formally various theses of interest concerning attitudes to fundamental states. Let $\langle p, v\rangle$ and $\langle p^*, v^*\rangle$ respectively be the prior and posterior states of mind of an agent. Then for all $\omega, \omega' \in \Omega$ such that $p^*(\omega), p^*(\omega') > 0$:

1. Invariance of Fundamental Desire (IFD): $v^*(\omega) > v^*(\omega') \Leftrightarrow v(\omega) > v(\omega')$
2. Invariance of Fundamental Belief (IFB): $p^*(\omega) > p^*(\omega') \Leftrightarrow p(\omega) > p(\omega')$

The first thesis, IFD, is one half of Becker's methodological position on tastes. The second thesis, IFB, is a parallel claim about fundamental beliefs and is central to classical Bayesian thinking. It is these two theses, and particularly the first, that will be investigated here. But it is worth noting that for each of them there is a corresponding hypothesis about the universality of fundamental attitudes. Let $I = \{1, ..., n\}$ be a set of individual agents and $\langle p_i, v_i\rangle$ be the state of mind of the ith member of I at a given time. Then for all $i, j \in I$ and for all $\omega, \omega' \in \Omega$ such that $p_j(\omega), p_j(\omega'), p_i(\omega), p_i(\omega') > 0$, one might postulate that:

3. Universality of Fundamental Desire (UFD): $v_j(\omega) \geq v_j(\omega') \Leftrightarrow v_i(\omega) \geq v_i(\omega')$
4. Universality of Fundamental Belief (UFB): $p_j(\omega) \geq p_j(\omega') \Leftrightarrow p_i(\omega) \geq p_i(\omega')$

Universality theses, like the invariance ones, are commonly made on methodological or heuristic grounds and can have very powerful consequences. Some examples: Aumann's use of a common prior assumption (UFD) in his 'Agreeing to Disagree' theorem; Harsanyi's similarity postulate (UFB) to support his derivation of Utilitarianism; and, of course, Becker's own use of universal tastes in his explanation of cultural difference. Although we shall not examine the universality theses here, much of the discussion of the invariance ones carries over to them.

3 Classical Conditioning

Classical conditioning is a way of revising your attitudes when you learn that some proposition A is true or receive the information that A. If an agent revises by classical conditioning on A then her new attitudes are such that for all prospects X such that $p(AX) \neq 0$:

$$\begin{aligned} p^*(X) &= p(X|A) = \frac{p(AX)}{p(A)} \\ v^*(X) &= v(X|A) = v(AX) - v(A) \end{aligned}$$

The functions $p(\cdot|A)$ and $v(\cdot|A)$ are, respectively, a probability measure and a desirability measure of the agent's degrees of conditional belief and conditional desire given the truth of A. An agent's conditional attitude to any prospect, given that A, is *not* the attitude she will have to it in the event that the condition is realised or discovered to be true, but her current attitude to it on the supposition that A is true. So what classical conditioning on the truth of A essentially consists in is the adoption of one's current conditional attitudes, on the supposition that A, as one's new attitudes. Thus, if I currently prefer taking white wine to red to a dinner party, conditional on the supposition that fish is to be served, and red to white on the supposition that meat is to be, then once I have phoned my hosts and settled the issue of what we will be eating, I should simply adopt as my preferences over wine, the conditional preference that was based on the correct supposition.

Revision by classical conditioning accords with IFB as it leaves the ordering induced by fundamental beliefs unchanged upon receipt of the information. For if $p(\omega|A), p(\omega'|A) > 0$ then $p(\omega|A) \geq p(\omega'|A) \Leftrightarrow p(\omega) \geq p(\omega')$. Classical conditioning also accords with Becker's Thesis, as preferences amongst the elements of the fundamental partition do not change. For if $p^*(\omega), p^*(\omega') > 0$:

$$\begin{aligned} v(\omega|A) &> v(\omega'|A) \\ &\Leftrightarrow v(\omega) - v(A) > v(\omega') - v(A) \\ &\Leftrightarrow v(\omega) > v(\omega') \end{aligned}$$

Indeed, with one qualification, we can explicitly express the agent's new state of mind in terms of her old desirabilities plus her new degrees of belief. The qualification derives from the requirement that the measure of the agent's degrees of desire is normalised with respect to the set of fundamental states, Ω. Given the renormalisation factor $k = \sum_i v(\omega_i).p^*(\omega_i)$, which roughly expresses the desirability gain (relative to the agent's prior expectations) derived from the truth of A, we have:

$$\begin{aligned} v^*(X) &= \sum_i v(\omega_i).p^*(\omega_i|X) - k \\ &= \sum_i v(\omega_i).p(\omega_i|AX) - k \end{aligned}$$

i.e. the agent's new degrees of desire are obtained from her old by averaging her old fundamental desires with her old conditional degrees of belief, given the new information that A. Or more pithily: new degrees of desire are old conditional expectations, given new information, for old fundamental desirability.

The Rigidity Condition Adopting $p(\cdot|A)$ and $v(\cdot|A)$ as your new degrees of belief and desires is demonstrably the correct thing to do just in case your the effect of interaction with the environment is confined to your becoming certain of the truth of some prospect A, and your conditional desires, given the truth of A, do not change as a result i.e. for all $X \in \Gamma'$ such that $p(AX) \neq 0$:

1. *Certainty*: $p^*(A) = 1$, $v^*(A) = 0$
2. *Rigidity*: $v^*(X|A) = v(X|A)$

This is proved as Theorem 10 in Bradley [7]. We now show the equivalence of Rigidity to a couple of other useful conditions (proof in the appendix).

Theorem 1 *The following are equivalent:*
(i) $\forall(X \in \Gamma' : p(AX) \neq 0)$, $v^*(B|A) = v(B|A)$;
(ii) $\forall(X \in \Gamma' : p(AX) \neq 0)$, $v^*(XA) - v(XA) = v^*(A) - v(A)$ *and* $\frac{p^*(XA)}{p(XA)} = \frac{p^*(A)}{p(A)}$;
(iii) $\forall \omega \in A$, $v^*(\omega) - v(\omega) = v^*(A) - v(A)$ *and* $\frac{p^*(\omega)}{p(\omega)} = \frac{p^*(A)}{p(A)}$.

The important question, of course, is whether and when we can expect these conditions to be satisfied. Arguably Rigidity should hold whenever A describes all and everything that is learnt by the agent as a result of interaction with the environment and all changes to the agent's partial attitudes are rational effects of her learning that A. For suppose that the agent's new conditional degrees of desire given A were not the same as her old ones. Since the truth of A is not itself a reason to change one's conditional desires given A, something more than A must have been learnt, such as that A's being true has previously unforeseen consequences. But that is contrary to the supposition that A is all that is learnt.

The notion of a reason for a change in attitude is left rather vague by this informal argument. It can be sharpened however by showing that an agent whose conditional desires do not satisfy the Rigidity condition under the imagined circumstances is vulnerable to a money pump. For the purposes of the exercise let us suppose that the truth of prospects can be bought and sold in some market so that, in an appropriate currency, $v(X)$ and $v^*(X)$ give the fair prices for the agent, before and after learning that A, of the prospect of X. Suppose firstly that the agent commits herself to a revision policy in case of learning that A such that for some X, $v(X|A) \neq v^*(X|A)$. There are two cases:

i.) $v(X|A) > v^*(X|A)$. In this case the agent can be sold the option of XA for $v(XA)$ and the option of A can be bought from her for $v(A)$. Once the option of A has been exercised the option of XA can be bought from the agent for $v^*(XA)$. By assumption $v(XA) - v(A) = v(X|A) > v^*(X|A) = v^*(XA)$. So in this case she is $v^*(X|A) - v(X|A) > 0$ poorer.

ii) $v^*(X|A) > v(X|A)$. In this case the option of XA can be bought from the agent for $v(XA)$ and the option of A sold for $v(A)$. Once the option of A has been exercised the option of AX can sold back to the agent for $v^*(XA)$. By assumption $v^*(XA) = v^*(X|A) > v(X|A) = v(XA) - v(A)$. So in this case she is $v^*(XA) - v(XA) + v(A) > 0$ poorer.

It follows that in any case in which the agent commits to a revision policy that fails to satisfy the Rigidity condition she will find herself open to a sure loss.

Money pump arguments, like their close relatives the Dutch Book arguments, show that failure to satisfy some condition or other renders the agent vulnerable to exploitation. It does not follow without further argument that rigidity of conditional attitude is a requirement of rationality under the given circumstances. After all one can render oneself invulnerable to money pumps by simply not declaring a belief revision policy. Indeed this would seem to be a sensible precaution since there are cases in which one's attitudes may change as a result of interaction with the environment but not (entirely) because of the information that one acquires during it simply because the manner in which something is learnt has some non-rational effect on one's attitudes. If, for instance, one learns of the consequences of excessive alcohol consumption by doing the drinking oneself or of the presence of a poisonous snake in the house by standing on it, there is every possibility that other attitudes will be altered in the process and in a manner not representable as a conditioning on what has been learnt. Unless the manner in which information is acquired can be controlled somehow (as perhaps it is in scientific experiments), it would be unwise to commit oneself to a revision policy in the manner required by the money-pump argument.

The Scope of Classical Conditioning The money-pump argument, as well as its informal predecessor, really only succeeds if we restrict its application to cases in which learning that A is the *only* immediate effect of the interaction with the environment on one's state of mind. The cases described in the previous paragraph suggest however that some revisions have non-informational roots and consequently that they cannot be described in terms of learning the truth of some proposition. The mark of attitude change of this kind is the failure of conditions akin to what van Fraassen [14] dubbed the Reflection Principle. Suppose that I learn somehow that I will come to adopt $p^*(X)$ as my degree of belief in X some time in the future. Then the Reflection Principle requires me to adopt $p^*(X)$ as my current degree of belief in X as well. A similar condition on preference would require that one adopts as one's current preferences any future preferences that one expects to acquire. Neither versions are convincing as conditions of rationality, even though failure to satisfy them does in fact render one vulnerable to either a Dutch Book or a Money Pump. Consider:

Example 2 *A twenty-something socialist, seeing the truth in Briand's claim that the "man who is not a socialist at twenty has no heart, but if he is still a socialist at forty he has no head", believes that he will no longer believe in socialism when he is older. He does not however believe that his change in belief will bring him nearer to the truth. On the contrary, he suspects that Briand's observation is true because of a natural capacity for rationalisation and self-justification that accompanies steady absorption into bourgeois life and its material comforts. So he feels no compulsion to give up his socialist beliefs now, even if he recognises that he will do so one day.*

There is something tragic about the condition of someone who knows that her attitudes will change for the worse, by her present lights. But given this, far from it being rational for her to adopt her anticipated attitudes, it would be positively irrational to do so. Reflection principles only have force with respect to future attitudes that one believes will be acquired by a reliable learning process such as conditioning on information received. The fact that we do not regard such principles as universally binding shows that we do not believe classical conditioning to exhaust the ways in which we can change our mind. Consequently to represent all belief change in terms of information acquisition requires that we regard such beliefs about ourselves to be in error.

4 Jeffrey conditioning

Our discussion thus far has left open the question whether classical conditioning is the only form of belief change that can be motivated in terms of what has been learnt from interaction with the environment. There are reasons for thinking, however, that someone's degree of belief in a particular proposition may change with reason without them being sure of either its truth or falsehood and that:

> "probabilistic judgement may be appropriate as a direct response to experience, underived from sure judgment that the experience is of such and such a character". - Jeffrey [11, p. 45].

Example 3 *I overhear a conversation in a foreign language and from the sounds of words and the mannerisms of the speakers I conclude that they are most likely, say, Spanish, but perhaps Catalan or even French. They seem to be assenting to each other's remarks by utterances of 'si', but perhaps I am mishearing. Could the 'si' be the French denial of a negated assertion? There is no hope here of producing a sentence that summarises all and only the facts learnt (and believed with probability one) in the encounter. Relevant evidence is not entirely indubitable, many of the cues never make it into consciousness (and perhaps cannot) and I don't have well-defined conditional degrees of belief for the speaker's language, given the bits of evidence that do.*

In circumstances in which the agent's degrees of belief over some partition of prospect space changes, without her being certain of the truth of any one of the elements of it, the natural generalisation of classical conditioning is a form of revision called Jeffrey conditioning. If an agent revises by Jeffrey conditioning on a partition $\{A_i\}$ then her new attitudes are such that for all prospects X :

$$\begin{aligned} p^*(X) &= \sum_i p(X|A_i).p^*(A_i) \\ v^*(X) &= \sum_i v(XA_i).p^*(A_i|X) - k \end{aligned}$$

where, as before, $k = \sum_i v(\omega_i).p^*(\omega_i) = \sum_i v(A_i).p^*(A_i)$ is a renormalisation term required to ensure that $v^*(\Omega) = 0$.

Revision by Jeffrey conditioning satisfies IFD, since preferences amongst the elements of the fundamental partition do not change. For:

$$v^*(\omega) > v^*(\omega') \Leftrightarrow \sum_i v(\omega).p^*(A_i|\omega) > \sum_i v(\omega').p^*(A_i|\omega') \Leftrightarrow v(\omega) > v(\omega')$$

Indeed once again we can explicitly express the agent's new state of mind in terms of her old desirabilities (up to normalisation) plus her new degrees of belief:

$$v^*(X) = \sum_i v(\omega_i).p^*(\omega_i|X) - k$$

i.e. the agent's new degrees of desire are obtained from her old by averaging her old fundamental desires with her new degrees of beliefs. Or more pithily: new degrees of desire are new expectations for old fundamental desirability.

On the other hand, IFB is not satisfied since fundamental belief change is not ruled out in this model of preference revision and there may be no sense in which the agent's new degrees of belief are derived from her old. Instead posterior belief is obtained by averaging the new partition probabilities with the old conditional probabilities given each element of it.

The Scope of Jeffrey Conditioning Jeffrey conditioning on the partition $\{A_i\}$ is demonstrably rational whenever the Stage 1 constraint on revision is exhausted by the requirement that $p^*(A_1) = a_1$, $p^*(A_2) = a_2$, etc., and this redistribution of belief leaves the agent's degrees of conditional desires given the A_i unchanged i.e. whenever the Rigidity condition applies to all the A_i. A proof of this claim is given in Bradley [7]. But why should an agent's conditional desires given the A_i not change when her degrees of belief for the A_i do? We can extend our earlier Money Pump argument to this more general case of probabilistic updating to provide an answer.

Consider a two-stage revision process. At the first stage, interaction with the environment induces the agent to adopt new probabilities for the elements of the partition $\{A_i\}$, without the probability of any one of them going to one. In the second stage the agent learns which of the A_i is the truth. Suppose that this process leads to a transformation of her state of mind from $\langle p, v\rangle$ to $\langle p^*, v^*\rangle$ and then to $\langle p^{**}, v^{**}\rangle$. By our previous argument for Rigidity in the context of classical conditioning, $v^{**}(\cdot|A_i) = v^*(\cdot|A_i)$ and $v^{**}(\cdot|A_i) = v(\cdot|A_i)$, since both the revisions from $\langle p^*, v^*\rangle$ to $\langle p^{**}, v^{**}\rangle$ and that from $\langle p, v\rangle$ to $\langle p^{**}, v^{**}\rangle$ fall under its scope. It follows that $v^*(\cdot|A_i) = v(\cdot|A_i)$ for any of the A_i and hence that the Rigidity condition holds for pure probabilistic shifts as well.

As before this argument presupposes that the initial impact of interaction with the environment is confined to changes to the agent's degrees of belief over the relevant partition. If there are any additional effects on the agent's desires that are by-products of the manner in which she acquires uncertain evidence about the elements of the partition, then her conditional desires given the A_i may change, as presumably might be the case in the examples given before. Given this possibility it is unwise to declare a belief revision strategy unless the manner in which the uncertain evidence is acquired can be controlled.

On the other hand the scope of the money-pump argument for Jeffrey conditioning is far greater than that for classical conditioning. For it is not only cases of receipt of uncertain evidence that can be represented by redistributions of probability over some definite partition, but a great many others including those involving receipt of probabilistic or conditional information (see Bradley [6]). It also extends to the kinds of cases we discussed before involving the breakdown of reflection principles, because the interaction with the environment inducing the redistribution of probabilities need not be interpreted in informational terms. In fact *any* revision of probabilities defined on a countable set of propositions and not involving assignment of probability to prospects with zero prior probability can be represented as an instance of Jeffrey conditioning on some partition satisfying the Rigidity condition (see van Fraassen [14] and Diaconis and Zabell [9] for further discussion of this feature).[2]

Taste and Value Change Significant though it may be that all belief change is representable in terms of Jeffrey conditioning, it does not settle the question of whether all preference change can represented in these terms. On the face of it the answer is no, for there are cases of preference revision that seem to involve pure taste or value changes. Examples include cases of conditioning or cultivation of taste by habituation - e.g. weaning infants onto cow's milk or acquiring a taste for olives - and cases of value discovery - e.g. when you learn that relationships require discretion as well as honesty, or that red wine is best drunk with cheese. In habituation cases repeated experience of something leads to a re-evaluation of it (typically unconsciously) despite the fact that any informational gains are made only in the early repetitions. One can grow tired of a foodstuff, for instance, not because of anything one learns about it, but simply because of the jading of one's palate. In cases of value discovery, some kind of learning is involved, but it seems to be of a different nature to that involved in the improvement of belief. When one learns that a particular wine is a good companion to a particular cheese (perhaps contrary to prior expectations), one does of course learn something about the two products. But what one learns about them is how they stand in relation to one's tastes; a discovery that must give rise to an improved evaluation of the products in combination, before it gives rise to a new (and improved) belief about them.

As in the case of belief change, recognition of the possibility of taste change without an informational, or indeed credal, source is implicit in the choices we make. And again the mark of this is the violation of a reflection principle; in this case the requirement that anticipation of future desire change should lead to the adoption of the anticipated desire. Consider:

Example 4 *A chocolate lover prefers to eat only a small amount of chocolate after dinner each day because of the impact on her health of excessive chocolate consumption. Chocolate could be bought in bulk once a week or in small quantities on a daily basis. Despite the fact that the latter is both more expensive and time-consuming she prefers to buy daily, because she knows that if she has a lot of chocolate available she will not confine herself to eating a small quantity of it.*

This example, like that of the twenty-something socialist, involves an expectation of attitude changes that are not endorsed by the person expecting to undergo them. But this time it is the expected increase in her desire for chocolate, not a belief change, that would seem to motivate the choice of the more costly action. Unless we are deluded about the attitude changes we can undergo, the evidence is that not all preference revision can be described in terms of Jeffrey conditioning.

5 Generalised conditioning

The most general of the types of attitude revision to be considered is what I call generalised conditioning. This is a form of revision that is appropriate when the effects of interaction with the environment can be represented by a redistribution of probability and desirability over some particular partition $\{A_i\}$ of the space of prospects, so that the Stage 1 constraint takes the form of the requirement that $p^*(A_1) = x_1$, $p^*(A_2) = x_2$, ..., $v^*(A_1) = y_1$, $v^*(A_2) = y_2$, etc., . In which case her new attitudes are such that for all

[2] The restriction to cases not involving assignment of positive probability to prospects previously regarded as certainly false is for technical reasons alone. The restriction could be dropped if Popper-Renyi conditional probability functions were used for measuring conditional belief.

prospects X:

$$\begin{aligned} p^*(X) &= \sum_i p(X|A_i).p^*(A_i) \\ v^*(X) &= \sum_i [v(X|A_i) + v^*(A_i)].p^*(A_i|X) \end{aligned}$$

Revision by generalised conditioning can violate both IFB and IFD as revision of both fundamental belief and fundamental desire is allowed. As in Jeffrey conditioning, posterior probabilities are obtained by averaging the new probabilities for the elements of the base partition with the old conditional probabilities given these elements. Similarly, the posterior desirability of each prospect X is obtained by averaging the new desirabilities for the elements of the partition $\{XA_i\}$, where these are obtained from the new desirabilities of the A_i - given by Stage 1 - and the old conditional desirabilities of X given the A_i.

Generalised conditioning is demonstrably valid given the assumed Stage 1 constraints when the agent's conditional desirabilities, given the A_i, remain invariant i.e. whenever the Rigidity condition applies to all the A_i. This is proved as Lemma 22 in Bradley [7]. In this case moreover the Money-pump argument given in the previous section applies without restriction, because the Stage 1 conditions for the application of generalised conditioning rule out the possibility of any attitude change as a by-product on the interaction (assuming these conditions are exhaustive). This makes for a strong case in favour of the Rigidity condition within the postulated conditions of application.

Assuming that generalised conditioning is rational under the assumed circumstances, we can now ask to what extent preference change can be modelled in these terms. Both the phenomenon of habituation and that of value learning are amenable to modelling in terms of generalised conditioning, despite the fact that the processes inducing them may be quite different in nature. In fact, somewhat surprisingly, it turns out that *all* preference revision can be modelled as generalised conditioning. This is the message of the main theorem of the paper, which follows.

Theorem 5 *Let $\langle p, v\rangle$ and $\langle p^*, v^*\rangle$ be respectively an agent's prior and posterior states of mind. Then there exists some partition of Ω such that $\langle p^*, v^*\rangle$ is obtained from $\langle p, v\rangle$ by generalised conditioning on this partition. (Proof in the appendix).*

Theorem 5 shows that it is always possible to represent a change in an agent's state of mind as an instance of generalised conditioning on some partition of the space of prospects, even when changes in fundamental desires are involved. In this sense it gives a completely general framework for the representation of preference change. It does not follow that the explanation of any instance of preference change in terms of generalised conditioning is the only possible one or even the best one. Interaction with the environment could leave the agent with new conditional desires which she then uses as a basis for adjusting her unconditional ones e.g. when tasting different combinations of wine and cheese leads to new conditional preferences for the wines given the cheeses. But it does show that any such change would be equivalent to generalised conditioning on some partition; in our example this would most likely be the partition of wine-cheese pairs.

6 Reducibility to Belief Change

In this final section we return to the assessment of our version of Becker's methodological hypothesis; namely that it is possible to explain all preference change in terms of changes in belief against the backdrop of invariant fundamental desires. I have given what I take to be strong arguments against the possibility of doing so if belief change is construed in purely informational terms. But it is less clear that it is impossible to do so if we equip ourselves with the full resources of the Jeffrey conditioning models. It may be possible to describe what was previously identified as instances of pure taste change - habituation and value learning - in terms of changes in beliefs concerning factors determining the desirability of the prospects in question. Becker himself is a master of such redescriptions as is evidenced by his explanation of addiction (a habituation phenomenon) in terms of changes in stocks of personal capital and hence expectations about the benefits of consumption that derive from them, and of advertising-induced taste change (an instance of value learning) in terms of changes in the shadow prices of the advertised goods, themselves consequences of changes in the agent's beliefs about the properties of these goods (see Becker [1]).

My own view is that it is indeed often possible to describe a given instance of preference change either in terms of a belief change or a taste change (or both) and that which of these is correct is underdetermined by the evidence we can bring to bear on the task. Consider, for instance, our previous examples of the chocolate lover and of the twenty-something socialist. In both, there is some ambiguity as to whether it is belief changes or value/desire changes that are involved. The change from socialist may be as much a matter of a change in values, or weight given to certain values, as changes in beliefs about the consequences of socialism. Equally it may be possible to explain the purchasing habits of the chocolate lover either in terms of an expectation that consuming chocolate will increase the strength of her desire to consume more, or in terms an expectation that consuming chocolate will change her beliefs about the consequences of so doing for her health.

This raises the question of whether the option of explaining a preference change in terms of changes in belief is *always* available. Clearly the answer is no when we work with a fixed countable algebra of prospects, as we have in this paper. For a revision of the desirabilities of the fundamental states cannot be derived from a belief revision on some more refined partition (since there is no such partition). On the other hand, if one is allowed to refine the space of fundamental states at will, then it seems likely that it will always be possible to satisfy Becker's demand for explanations based on invariant tastes. But refining the state space can be as ad hoc from a methodological point of view as postulating changes in fundamental desires and I see little to recommend it as a general strategy for explaining away apparent taste changes.

If a sufficiently refined space of fundamental states is picked, however, then it is possible that for many of the cases of preference change that are actually observed, it will be possible to formulate an explanation in terms of underlying belief changes. To state more precisely the conditions under which such an explanation will be available suppose that we observe a change in the agent's preferences over some partition of the state space only and that our representation of the agent's new state of mind is constrained by the information we hold about them to no greater extent than determining a measure of her degrees of belief and desire over the power set of this partition. (Of course if we hold less information about the agent's new state of mind then it becomes easier to formulate an explanation in terms of belief change that is consistent with the known facts). In this case the kind of explanation we seek will be available whenever the following condition on the agent's postulated fundamental preferences over the partition is satisfied.

Condition 6 (Richness) *Let $\{A_i\}$ be a partition of Ω and $\succ$ a strict preference relation on Ω. Then $\langle\{A_i\}, \succ\rangle$ satisfies the richness condition iff there exists, for all A_i, $\omega_i^g, \omega_i^b \in A_i$ such that $\omega_i^g \succ \Omega \succ \omega_i^b$.*

The Richness condition requires of a partition that each of its elements contains both a good and a bad possible state of the world, where 'good' and 'bad' mean more or less preferred than the status quo. Intuitively this will be the case whenever the partition in question does not involved a division of possible states in accordance with how good they are. Thus a partition containing a element defined as the set of all states that are preferred to the staus quo will violate the condition, $\{A, \neg A\}$

Theorem 7 *Let an agent's prior state of mind be represented by $\langle p, v\rangle$ and suppose that the agent's degrees of belief and desire for the elements of a partition $\{A_i\}$, satisfying the richness condition, are* *'observed' to change from $v(A_i)$ and $p(A_i)$ to $v^*(A_i)$ and $p^*(A_i)$. Then there exists a representation of her posterior state of mind probability $\langle p^*, v^* \rangle$ such that:*

$$v^*(A_i) = \sum_{\omega_j \in \Omega} v(\omega_j).p^*(\omega_j|A_i) - k$$

where $k = \sum_i v(A_i).p^*(A_i)$.

Theorem 7 shows that in some fairly typical set of conditions it will be possible to give an explanation of an observed preference change in terms of a change in underlying fundamental beliefs. These conditions are, on the one hand, that the preference change should have occurred over a relatively coarse-grained set of prospects, each of which has both good and bad potential conditions of realisation and, on the other hand, that we hold less than full information about the agent's new state of mind. In particular we don't hold any information that rules out the possibility that the postulated change in fundamental belief has occurred.

Like Theorem 5, Theorem 7 is a possibility result; a result that show what sorts of explanations of preference change are open to us. Possibility results don't settle the question of desirability, of whether an explanation should be offered in one form or another. As has long been emphasised by philosophers of mind, explanation of behaviour is in part a matter of fitting what we observe into recognisable patterns (see, for instance, Dennett [8]). And in this regard there are intelligible patterns of taste change just as much as there are of belief change. There are a number of criteria that one can appeal to in assessing the appeal of an explanation: empirical content, simplicity and plausibility are all surely relevant. For my part, I doubt that these criteria will always favour explanations that assume fixed fundamental tastes. The sorts of suppositions that will need to be made about changes to underlying beliefs in order to preserve the invariance of tastes may well be as ad hoc as the assumptions about taste changes that they are supposed to replace, and may be no more constrained by the empirical evidence.

7 Appendix

Proof of Theorem 1.
(i) $\Leftrightarrow$ (ii): From the definition of conditional desirability (equation 2) it follows immediately that:

$$v^*(X|A) = v(X|A) \Leftrightarrow v^*(AX) - v(AX) = v^*(A) - v(A)$$

And by Lemma 21 of Bradley [7], if $v^*(X|A) = v(X|A)$ then $p^*(X|A) = p(X|A)$. Hence $\forall X \in \Gamma$:

$$\frac{p^*(AX)}{p(AX)} = \frac{p^*(A)}{p(A)} \tag{3}$$

(ii) $\Rightarrow$ (iii): Suppose that $\forall X \in \Gamma$, $v^*(AX) - v(AX) = v^*(A) - v(A)$ and $\frac{p^*(AX)}{p(AX)} = \frac{p^*(A)}{p(A)}$. Then it follow that, in particular, $\forall \omega \in A$, $v^*(\omega) - v(\omega) = v^*(A) - v(A)$ and $p^*(\omega)/p(\omega) = p^*(A)/p(A)$.
(iii) $\Rightarrow$ (i): Now assume that $\forall \omega \in A$, $v^*(\omega) - v(\omega) = v^*(A) - v(A)$ and $p^*(\omega)/p(\omega) = p^*(A)/p(A)$.

Then $\forall X \in \Gamma$:

$$\begin{aligned}
v^*(X|A) &= v^*(AX) - v^*(A) \\
&= \sum_{\omega \in AX} \frac{v^*(\omega).p^*(\omega)}{p^*(AX)} - v^*(A) \\
&= \sum_{\omega \in AX} \frac{(v(\omega) - v^*(A) + v(A)).p^*(\omega)}{p^*(AX)} - v^*(A) \\
&= \sum_{\omega \in AX} \frac{v(\omega).p^*(\omega)}{p^*(AX)} - v(A) \\
&= \sum_{\omega \in AX} \frac{v(\omega).p(\omega)}{p(AX)} - v(A) \\
&= v(AX) - v(A) \\
&= v(X|A)
\end{aligned}$$

■

Proof of Theorem 5. Let $R := \{i \in \Re : v^*(\omega) - v(\omega) = i, \forall \omega \in \Omega\}$ and $S := \{j \in \Re : p^*(\omega)/p(\omega) = j, \forall \omega \in \Omega\}$. Then let $A_{ij} := \{\omega \in \Omega : v^*(\omega) - v(\omega) = i, p^*(\omega)/p(\omega) = j, \forall i \in R, j \in S\}$, so that the A_{ij} partition the set of fundamental worlds into equivalences classes of worlds with identical differences in prior and posterior desirabilities and ratios of prior and posterior probabilities. Now by the definition of conditional desirability and equation (2):

$$\begin{aligned}
v^*(X|A_{ij}) &= v^*(X|A_{ij}) - v^*(A_{ij}) \\
&= \sum_{\omega \in XA_{ij}} \frac{v^*(\omega).p^*(\omega)}{p^*(XA_{ij})} - \sum_{\omega \in A_{ij}} \frac{v^*(\omega).p^*(\omega)}{p^*(A_{ij})} \\
&= \sum_{\omega \in XA_{ij}} \frac{(v(\omega) + i).j.p(\omega)}{p^*(XA_{ij})} - \sum_{\omega \in A_{ij}} \frac{(v(\omega) + i).j.p(\omega)}{p^*(A_{ij})}
\end{aligned}$$

But by equation (1):

$$p^*(XA_{ij}) = \sum_{\omega \in XA_{ij}} p^*(\omega) = \sum_{\omega \in XA_{ij}} j.p(\omega) = j.p(XA_{ij})$$

So:

$$\begin{aligned}
v^*(X|A_{ij}) &= \sum_{\omega \in XA_{ij}} \frac{(v(\omega) + i).j.p(\omega)}{j.p(XA_{ij})} - \sum_{\omega \in A_{ij}} \frac{(v(\omega) + i).j.p(\omega)}{j.p(A_{ij})} \\
&= \sum_{\omega \in XA_{ij}} \frac{v(\omega).p(\omega)}{p(XA_{ij})} - \sum_{\omega \in A_{ij}} \frac{v(\omega).p(\omega)}{p(A_{ij})} \\
&= v(X|A_{ij}) - v(A_{ij}) \\
&= v(X|A_{ij})
\end{aligned}$$

Hence the Rigidity condition applies to the elements of $\{A_i\}$. It follows immediately that $\langle p^*, v^* \rangle$ is obtained from $\langle p, v \rangle$ by generalised conditioning on $\{A_i\}$. ■

Proof of Theorem 7. We prove the theorem by constructing $\langle p^*, v^* \rangle$ from $\langle p, v \rangle$ and the $v^*(A_i)$ and $p^*(A_i)$. For any A_i let $\omega_i^g, \omega_i^b \in A_i$ be such that $\omega_i^g \succ \Omega \succ \omega_i^b$. Define:

$$\begin{aligned}
k &= \textstyle\sum_i v(A_i).p^*(A_i) \\
p^*(\omega_i^g|A_i) &= \frac{v^*(A_i) - v(\omega_i^b) + k}{v^*(\omega_i^g) - v(\omega_i^b)} \\
p^*(\omega_i^b|A_i) &= 1 - p^*(\omega_i^g|A_i)
\end{aligned}$$

and $\forall \omega \in A_i - \{\omega_i^g, \omega_i^b\}$, $p^*(\omega|A_i) = 0$. Then:

$$v^*(A_i) = v(\omega_i^g).p^*(\omega_i^g|A_i) + v(\omega_i^b).p^*(\omega_i^b|A_i) - k$$

■

References

[1] Becker, G. (1996) *Accounting for Tastes*, Cambridge MA: Harvard University Press

[2] Becker, G. (1976) *The Economic Approach to Human Behavior*, Chicago; London : University of Chicago Press

[3] Bolker, E. (1966) "Functions Resembling Quotients of Measures", *Transactions of the American Mathematical Society* 124: 292-312

[4] Bolker, E. (1967) "A Simultaneous Axiomatisation of Utility and Subjective Probability", *Philosophy of Science* 34: 333-340

[5] Bradley, R. (1999) "Conditional Desirability", *Theory and Decision* 47: 23-55

[6] Bradley, R. (2005) "Radical Probabilism and Mental Kinematics", *Philosophy of Science*, 72: 342-364

[7] Bradley, R. (2007) "The Kinematics of Belief and Desire", *Synthese* 156: 513-535

[8] Dennett, D. (1991) "Real Patterns", *Journal of Philosophy* 88: 27-51

[9] Diaconis, P. and Zabell, S. (1982) "Updating Subjective Probability", *Journal of the American Statistical Association* 77: 822-30

[10] Hausman, D. (2006) "Consequentialism and Preference Formation in Economics and Game Theory" in *Preferences and Well-being*, ed. S. Olsaretti, Cambridge University Press, 2006, 111-129

[11] Jeffrey, R. C. (1992) *Probability and the Art of Judgement*, Cambridge University Press

[12] Jeffrey, R. C. (1983) *The Logic of Decision*, 2nd ed, Chicago, University of Chicago Press

[13] Stigler, G. J. and G. S. Becker (1977) "De Gustibus Non Est Disputandum", *The American Economic Review*, 67: 76-90

[14] van Fraassen, B. (1980) "Rational Belief and Probability Kinematics", *Philosophy of Science* 47: 165-87