Framing as Path-Dependence

Natalie Gold and Christian List[1]

Oxford University Discussion Paper in Economics – Number 124

12 October 2002

Abstract. A "framing" effect occurs when an agent's choices are not invariant under changes in the way a choice problem is formulated, e.g. changes in the way the options are described (violation of description invariance) or in the way preferences are elicited (violation of procedure invariance). In this paper we examine precisely which classical conditions of rationality it is whose non-satisfaction may lead to framing effects. We show that (under certain conditions), if (and only if) an agent's initial dispositions on a set of propositions are "implicitly inconsistent", her decisions may be "path-dependent", i.e. dependent on the order in which the propositions are considered. We suggest that different ways of framing a choice problem may induce the order in which relevant propositions are considered and hence affect the decision made. This theoretical explanation suggests some observations about human psychology which are consistent with those made by psychologists and provides a unified framework for explaining violations of description and procedure invariance.

Keywords. Framing, preference reversal, path-dependence, rationality, deductive closure.
JEL classification. D11, D80.

The choices that people make are sometimes sensitive to the way in which the question is put. They may depend on the way in which options are described or on the way in which preferences are elicited. They are not always "description invariant" or "procedure invariant". In logicians' language, two choice problems may be "extensionally equivalent", and yet lead to different choices. If we take a descriptive expression from a proposition and substitute a different expression that designates the same object this should, ideally, not affect the truth value an agent assigns to the proposition. And yet, empirically, it sometimes does. These phenomena are called "framing effects". Psychologists have offered accounts of decision making that might explain why violations of description invariance or procedure invariance occur, but such framing effects are offensive to a logician's account of rationality. In this paper we use a logician's framework to examine exactly which classical conditions of rationality it is whose non-satisfaction may lead to framing effects. We show that (under certain conditions), if (and only if) an agent's initial dispositions on a set of propositions are *implicitly inconsistent* (as defined below), her decisions may be *path-dependent*, i.e. dependent on the order in which the propositions are considered. We suggest that different ways of framing a choice problem may induce the order in which relevant propositions are considered and hence affect the decision made. This theoretical explanation suggests some observations about human psychology which are consistent with those made by psychologists and provides a unified framework within which we can see the similarities between explanations of violations of description and procedure invariance.

[1] N. Gold, Balliol College, Oxford OX1 3BJ, U.K.; natalie.gold@balliol.oxford.ac.uk; C. List, Nuffield College, Oxford OX1 1NF, U.K.; christian.list@nuffield.oxford.ac.uk.

1. Violations of Description Invariance

An early experimental demonstration of framing effects, where the description of the options affected the choices that subjects made, is given by Tversky and Kahneman (1981). They asked subjects to imagine that the US was threatened by an unusual disease that was expected to kill 600 people and that they had to make a choice between two alternative vaccination programmes. Two groups were presented with the same choice problem but in different forms. The first group were told:

> If Program A is adopted, 200 people will be saved.
> If Program B is adopted, there is 1/3 probability that 600 people will be saved, and 2/3 probability that no people will be saved.

A second group were told that:

> If Program C is adopted, 400 people will die.
> If Program D is adopted, there is 1/3 probability that no-one will die, and 2/3 probability that 600 people will die.

In the first group, 72% of subjects opted for Program A but in the second group 78% of subjects chose Program D. Although A is extensionally equivalent to C and B is extensionally equivalent to D, changing the description of the options from one in terms of "lived saved" to one in terms of "lives lost" changed the modal preference.

In response to this and other findings in the field of decision making under uncertainty, Kahneman and Tversky (1979) developed prospect theory to explain the pattern of people's choices. Prospect theory suggests that decision makers code outcomes as gains or losses relative to some reference point and then, in their evaluation of the outcomes, are risk averse over gains but risk loving over losses. The way a decision problem is framed determines the reference point. In the above example, the phrasing "saved" in the first formulation of the problem highlights a gain so respondants are risk averse and the phrasing "die" in the second highlights a loss so they are risk loving. But although the original examples of framing involved risk, this is actually an unnecessary complicating factor. There is other evidence that changes in modal preference can be brought about in decisions which do not involve any uncertainty, simply by manipulating subjects' reference point and therefore what they regard as a gain or a loss. These results can then be explained by the theory of loss aversion, namely that individuals regard gains and losses differently (Tversky and Kahneman 1991). In fact, there are even more general framing effects, not involving gains and losses. For instance, when asked to judge the quality of beef, subjects' evaluations depend on whether it is described as '75% lean' or '25% fat'. These framing effects might all be described as "valence framing effects". Regardless of the presence of risk or reference points, in each case the different frames cast the same critical information in either a positive or a negative light. This leads to the suggestion that it is the positive or negative encoding of information that affects choice (Levin et al. 1998).[2]

2. Violations of Procedure Invariance

Framing effects are often thought of as occurring when choices are not invariant under changes of the way in which the options in the choice set are described. But there are also well documented violations of procedure invariance, where choices are affected by the way in which the preference over the options is elicited. In these experiments, by changing the method of elicitation of preferences, the same individual can be induced to make inconsistent choices. We will show that violation of the same axioms of rationality is responsible for both violations of description invariance and violations of procedure invariance.

One example of a violation of procedure invariance is a preference reversal phenomenon originally reported by Lichtenstein and Slovic (1971). Subjects were asked to evaluate pairs of gambles of comparable expected value. One gamble, the P gamble, offers a high probability of winning a relatively small amount of money. The other gamble, the $ gamble, offers a low probability of winning a larger prize. For instance, one of the pairs was:

P gamble	$ gamble
Win $2 with probability .80	Win $9.00 with probability .20
Lose $0.50 with probability .20	Lose $0.50 with probability .80

Both the gambles above have an expected value of $1.40. Subjects were asked both which gamble they preferred to play (a qualitative "choice" task) and also, in a different stage of the experiment, if they owned the right to play the gamble, how much they would be willing to sell that for (a quantitative "valuation" task). As Lichtenstein and Slovic said, "We say that option A is preferred to option B if option A is selected when B is available *or* if A has a

[2] One can tell a plausible evolutionary just-so story about why we might encode positive and negative stimuli differently. If negative stimuli were, on the whole, things that led to the death of our pre-human ancestors, e.g. predators, poisonous plants, whereas positive stimuli resulted in only incremental fitness increases then there may have been particular benefit to having either instinctive avoidance

higher reservation price than B. The standard analysis of choice assumes that these procedures give rise to the same ordering. This requirement – called procedure invariance – seldom appears as an explicit axiom but it is needed to ensure that the preference relation is well defined.” (Lichtenstein and Slovic 1971, p.203) However, the pattern of choices was that subjects said that they preferred to play the P gamble but gave the $ gamble a higher selling price. When the experimenters conducted a further study in a casino they found that, for the above gambles (with positive expected value), of participants who chose the P gamble 81% gave a higher selling price to the $ gamble and, what is more, some of them turned into "money pumps" continuously giving more money to the experimenters to switch between the gambles without ever playing them (Lichtenstein and Slovic 1973). Similar reversals were found when subjects were asked, first, to rate the attractiveness of gambles (valuation) and, then, which gamble they preferred to play (choice). The one that scored higher on the attractiveness rating was not always the one they said they preferred to play (Ordonez et al. 1995).

Again, although the original examples of violations of procedure invariance concern preference reversals over gambles, this effect does not rely on the presence of risk. The effect is operative in a whole class of tasks where there are two options, each of which is assessed in terms of more than one attribute, and where there are two different modes of preference elicitation, choice versus valuation. In the example of gambles, the attributes might be the maximal payoff and the probability of winning the maximal payoff. An example of violations of procedure invariance which does not involve uncertainty is given by the comparison of “choice” and “matching”, the latter being a type of valuation task. There are two options, with two relevant attributes each. In the matching task, for one of the options subjects are given the value of both of these attributes, whereas for the other option they are given the value of only one. They are then asked to supply the value of the second attribute that would make the two options equal in value overall. For instance, subjects are asked to consider two candidates for an engineering job, X and Y, who are each assessed on two different attributes, technical knowledge and human relations. The matching task might consist of giving the subjects candidate X’s scores for both technical knowledge and human relations but only one of candidate Y’s scores, e.g. on technical knowledge, and asking what score on the other attribute, human relations, would make the two candidates equally suitable for the job. (In fact, there are four possible matching tasks depending on which of the four items of information is withheld.) From subjects’ responses in the matching task we should be able to predict the decisions subjects would make in the choice task, where they are given the values

mechanisms or rapid reactions to negative stimuli. So we may have developed asymmetric response mechanisms to positive and negative stimuli and as a result encode the two categories differently.

of all attributes, i.e. both X and Y's scores for both technical knowledge and human relations. However, in an experiment, in the choice task 65% of subjects chose the candidate who scored higher on the more prominent attribute, technical knowledge, whereas the inference from those given a matching task was that only 34% would have rated this candidate as better. This leads to the *prominence hypothesis* that the prominent attribute will, for whatever reason, weigh more heavily in choice than in matching (Tversky at al. 1988).

There are two ways in which psychologists suggest that framing can affect the construction of preferences, called strategy compatibility and scale compatibility (Fischer and Hawkins 1993). Strategy compatibility is the hypothesis that different heuristics are used depending on the mode of preference elicitation. For instance, choice tasks might induce qualitative, lexicographic reasoning (i.e. a focus on a prominent attribute that is considered lexicographically prior to other attributes), whereas valuation tasks might induce quantitative assessment and explicit trade-offs between different attributes. Scale compatibility is the hypothesis that choices always involve the same heuristic where multiple conflictual attributes of the options are adjudicated, but different modes of preference elicitation lead to the assignment of different weights to different attributes. According to this hypothesis, in the above example, when subjects are asked to choose their preferred gamble, the probability of winning the maximal payoff is the attribute with the greater weight, whereas when they are asked for their monetary valuations of the gambles, the maximal payoff is the attribute with the greater weight.

3. A Simple Model

We seek to explain framing effects by attributing to the agent a sequential decision process in which the agent considers multiple propositions. Specifically, in our model the agent considers not only the "target proposition" (on which the agent has to make a decision), but also multiple other "background propositions", which may be logically connected to the target proposition, and which describe the "run-up" or "context" to the agent's decision on the target proposition. The background propositions may include factual propositions on which the agent might have beliefs that are relevant to his or her decision on the target proposition. They may further include normative propositions whose resolution (e.g. acceptance or rejection) may entail a particular stance on the target proposition. In short, the background propositions include all those propositions that the agent may consider in the process leading up to his or her decision on the target proposition.

The model of a sequential decision process follows List (2002). We first need to give a few preliminary definitions. Let X be a set of propositions, including the "target proposition" and relevant "background propositions".[3] The propositions are formalized in terms of first-order predicate calculus.[4] In particular, the set X may include

- atomic propositions with zero-place predicates (no logical connectives), e.g. P, Q, …;
- atomic propostions with one-place predicates (no logical connectives), e.g. Aa ("a has property A");
- atomic propositions with two-place predicates (no logical connectives), including ranking propositions, e.g. aPb ("a is strictly preferred to b");
- compound propositions (with logical connectives and/or quantifiers), e.g. $(P \wedge Q)$, $\forall x(Ax \rightarrow Bx)$, $\forall x\forall y((Ax \wedge \neg Ay) \rightarrow xPy)$.

We define a *decision-path* to be a *one-to-one* mapping $\Omega : \{1, 2, ..., l\} \rightarrow X$, where $l \leq k = |X|$.[5] We interpret $\Omega(1)$, $\Omega(2)$, $\Omega(3)$, ..., respectively, as the first, second, third ... propositions considered by the agent. A decision-path Ω is *complete* if $l=k$ (i.e. it reaches all propositions in X), and *incomplete* if $l<k$ (i.e. it reaches some but not all propositions in X).

A decision-path can be interpreted in (at least) two different ways: (i) as the *temporal* order in which the agent considers the propositions; or (ii) as the order of how *focal* the propositions are for the agent, or how much *weight* the agent assigns to the propositions.

In this model of a sequential decision process, the agent considers the propositions, one by one, in the order represented by a given decision-path. An agent's acceptance or rejection of each proposition in that sequence is determined by two criteria: (i) initial dispositions, and (ii) a conflict resolution rule.

Initial dispositions. For each proposition ϕ in X, an agent's *initial disposition* on ϕ is the judgment (acceptance/rejection) the agent *would* make on ϕ *if* he or she *were to consider* ϕ in isolation, with no reference to other propositions (particularly previously considered ones). Note that an initial disposition is a *counterfactual* notion. Saying that an agent has an initial disposition on ϕ does not carry any implications as to whether the agent has *in fact* considered the proposition. The agent's initial dispositions on the propositions in X are represented by an

[3] For technical reasons, (i) we assume X to be finite; (ii) we assume that X contains neither tautologies nor contradictions; (iii) we assume that X includes proposition-negation pairs (i.e. for every proposition ϕ in X, $\neg\phi$ is also in X); (iv) for each ϕ in Φ, we identify $\neg\neg\phi$ with ϕ.

[4] For an exposition of the formalism of first-order predicate calculus, see Hamilton (1988).

[5] The assumption that Ω is one-to-one means that, for all a and b in the domain of Ω, $\Omega(a)=\Omega(b) \Rightarrow a=b$. The requirement that a decision-path be a one-to-one mapping ensures that no proposition occurs twice in the path.

acceptance/rejection function $\delta: X \rightarrow \{0, 1\}$. For each ϕ in X, $\delta(\phi) = 1$ means that the agent has an initial disposition to accept ϕ, and $\delta(\phi) = 0$ means that the agent has an initial disposition to reject ϕ.

A conflict resolution rule. There are sometimes conflicts between an agent's initial disposition on a new proposition and previously accepted propositions; specifically, the agent's initial disposition on a new proposition may be logically inconsistent (and perceived so by the agent) with his or her previously accepted propositions. A *conflict resolution rule* is a method of resolving such conflict. The *modus ponens rule* – which is used in our model below – resolves conflict by accepting the logical implications of previously accepted propositions and overruling the initial disposition on the new proposition.

Attributing the modus ponens rule to an agent turns out to be particularly useful for explaining framing effects, in so far as that rule captures the notion that the "run-up" to the agent's decision on a certain proposition may constrain the agent's decision on that proposition. The "run-up" to the decision on a target proposition is precisely what is induced by a decision frame, and what may vary from frame to frame. Other conflict resolution methods are conceivable, particularly the *modus tollens rule*, which resolves conflict by accepting the initial disposition on the new proposition and revising previously accepted propositions; but our model will explicitly use the modus ponens rule.

We can define a *modus ponens decision process* in terms of a formal procedure.

- Consider the propositions, one by one, in the order determined by a given decision-path, say proposition ϕ_1 (:= $\Omega(1)$) at time 1, proposition ϕ_2 (:= $\Omega(2)$) at time 2, and so on.
- At each time t, when proposition ϕ_t is being considered, determine whether or not previously accepted propositions in the sequence have a logical implication for the acceptance or rejection of ϕ_t.

  If yes:

  Ignore the initial disposition ($\delta(\phi_t)$) on ϕ_t.

  Accept ϕ_t if previously accepted propositions imply ϕ_t.

  Reject ϕ_t if previously accepted propositions imply the negation of ϕ_t.

  If no:

  Accept ϕ_t if the initial disposition on ϕ_t is positive, i.e. if $\delta(\phi) = 1$.

  Reject ϕ_t if the initial disposition on ϕ_t is negative, i.e. if $\delta(\phi) = 0$.

We consider four rationality conditions which the initial dispositions of an individual may, or may not, satisfy.

Completeness. An agent's initial dispositions are *complete* if, for any proposition ϕ, the agent has a disposition to accept the proposition ϕ or its negation $\neg\phi$ (formally, $\delta(\phi)$=1 *or* $\delta(\neg\phi)$=1).

Weak Consistency. An agent's initial dispositions are *weakly consistent* if the agent never has a disposition to accept a proposition ϕ and its negation $\neg\phi$ simultaneously (formally, *not both* $\delta(\phi)$=1 *and* $\delta(\neg\phi)$=1).

The define a stronger notion of consistency, we need to introduce the notion of *semantic consistency*. A subset Y of X is *semantically consistent* if there exists a consistent assignment of truth-values which makes all the propositions in Y simultaneously true.

Strong Consistency. An agent's initial dispositions are *strongly consistent* if it is possible for all the propositions which the agent has a disposition to accept to be simultaneously true (formally, the set $\{\phi \in X : \delta(\phi)=1\}$ is semantically consistent).

Deductive Closure. An agent's initial dispositions are *deductively closed* if, whenever the agent has a disposition to accept a set of propositions Ψ and Ψ implies another proposition ϕ, then the agent also has a disposition to accept ϕ (formally, *if* [$\delta(\psi)$=1 for every ψ in Ψ] *and* [Ψ implies ϕ], *then* $\delta(\phi)$=1).

An agent's initial dispositions are not deductively closed with respect to ϕ if there exists a subset Ψ of X such that [$\delta(\psi)$=1 for every ψ in Ψ] *and* [Ψ implies ϕ], *but* $\delta(\phi)$=0.

The four conditions are not all logically independent from each other. Strong consistency implies weak consistency. The conjunction of completeness, weak consistency and deductive closure implies strong consistency.

Violations of strong consistency by an agent's initial dispositions will be called *implicit inconsistencies*. Lemma A1 in the appendix shows that such violations occur if and only if there exist two semantically consistent subsets Ψ_1 and Ψ_2 of X and a proposition ϕ in X such that [$\delta(\psi)$=1 for every ψ in $\Psi_1 \cup \Psi_2$] *and* [Ψ_1 implies ϕ] *and* [Ψ_2 implies $\neg\phi$]. We then say that the agent's initial dispositions are *implicitly inconsistent with respect to* ϕ.

Violations of weak consistency by an agent's initial dispositions will be called *explicit inconsistencies*. Such violations occur if and only if there exists a proposition ϕ in X such that $\delta(\phi)=1$ *and* $\delta(\neg\phi)=1$. We then say that the agent's initial dispositions are *explicitly inconsistent* with respect to ϕ.

Informally, an agent's initial dispositions are explicitly inconsistent if the agent has a disposition to accept both a proposition and the negation of that proposition. An agent's initial dispositions are implicitly inconsistent if some of the propositions which the agent has a disposition to accept *imply* the negation of what is *implied* by other propositions which the agent has a disposition to accept. In that case, the agent may or may not have a disposition to accept those *implications* themselves. If the agent has such a disposition, the inconsistency is not merely implicit, but also explicit. If the agent has no such disposition, on the other hand, the inconsistency is merely implicit, but not explicit. All explicit inconsistencies are also implicit inconsistencies, but not all implicit inconsistencies are also explicit inconsistencies.

Suppose, for example, an agent has an initial disposition to accept the propositions P, $(P\rightarrow Q)$ and $\neg Q$, but no other propositions. Then the agent's initial dispositions are implicitly inconsistent (and thus violate strong consistency): the set of propositions accepted by the agent has two semantically consistent subsets, namely $\Psi_1 = \{P, (P\rightarrow Q)\}$ and $\Psi_2 = \{\neg Q\}$, such that Ψ_1 implies Q and Ψ_2 implies $\neg Q$. However, the agent's initial dispositions are *not* explicitly inconsistent (and thus they satisfy weak consistency): there exists no proposition such that the agent has a disposition to accept the proposition and its negation simultaneously.

Our main theorem (see also List 2002) shows that an implicit inconsistency in an agent's initial dispositions is a necessary and sufficient condition for the occurrence of path-dependencies in a modus ponens decision process. The theorem follows immediately from lemma A2 in the appendix.

Theorem 1. Suppose the agent uses a modus ponens decision process. Then

(i) there exist (at least) two alternative decision-paths such that, under one path, ϕ is accepted and, under the other, $\neg\phi$ is accepted

<u>if and only if</u>

(ii) the agent's initial dispositions are implicitly inconsistent with respect to ϕ.

Implicit inconsistencies (i.e. condition (ii)) occur when the agent's initial dispositions are explicitly inconsistent, or when they satisfy weak consistency, but an inconsistency is "hidden" by a violation of deductive closure. Note the following lemma:

Lemma 1. Suppose the agent's initial dispositions over the propositions satisfy completeness and weak consistency. Then

(i) the agent's initial dispositions are implicitly inconsistent with respect to ϕ

<u>if and only if</u>

(ii) the agent's initial dispositions are not deductively closed with respect to ϕ.

The conjunction of theorem 1 and lemma 1 immediately implies the following theorem.

Theorem 2. Suppose the agent's initial dispositions over the propositions satisfy completeness and weak consistency. Suppose the agent uses a modus ponens decision process. Then

(i) there exist (at least) two alternative decision-paths such that, under one path, ϕ is accepted and, under the other, $\neg\phi$ is accepted

<u>if and only if</u>

(ii) the agent's initial dispositions are not deductively closed with respect to ϕ.

4. Violations of Description Invariance as Path-Dependence

We can use the framework above to illuminate violations of description invariance. Take the unusual disease problem of Kahneman and Tversky introduced above, where presentation 1 is given in terms of "lives saved" and presentation 2 is given in terms of "lives lost". Let us define three predicates:

Qx : x saves some lives *with certainty* (and does *not* involve a *risk* that *no-one* will be saved).
Rx : x consigns some people to death *with certainty* (and does *not* involve the *chance* that *no-one* will *die*).
xPy : x is strictly preferred to y.

As a general rule of rationality, we assume that the agent accepts $\forall x \forall y(xPy \rightarrow \neg yPx)$.

Let a, b, c, d denote programs A, B, C, D in the Kahneman and Tversky problem, respectively. Note that (extensionally) a and c denote the same program, and b and d denote the same program, and hence, on reflection, the agent would accept the propositions $a = c$ and $b = d$.

By the definition of the four programs, the agent has initial dispositions to accept the following (factual) propositions:

(1) Program A saves some lives (200) with certainty;
i.e. Qa.

(2) Program B involves the *risk that no-one will be saved* (with probability 2/3 no-one will be saved);
i.e. $\neg Qb$.

(3) Program C consigns some people to death (400) with certainty;
i.e. Rc.

(4) Program D offers the *chance that no-one will die* (with probability 1/3 no-one will die);
i.e. $\neg Rd$.

We also suppose that the agent has initial dispositions to accept the following two general (normative) principles:

(5) It is *not* worth taking the *risk that no-one will be saved*. Formally, if program y involves the *risk that no-one will be saved*, whereas program x saves some lives with certainty, then x is preferrable to y;
i.e. $\forall x \forall y((Qx \wedge \neg Qy) \rightarrow xPy)$.

(6) It is unacceptable to consign some people to death with certainty. Formally, if program x consigns some people to death with certainty, whereas program y offers the *chance that no-one will die*, then y is preferrable to x;
i.e. $\forall x \forall y((Rx \wedge \neg Ry) \rightarrow yPx)$.

Note that the agent's initial dispositions are incomplete. While the agent has initial dispositions over factual propositions concerning the alternatives, such as (1) to (4), and over general normative principles such as (5) and (6), the agent may not have initial dispositions over specific ranking propositions such as aPb or cPd, as the agent may be unable to consider these in isolation, i.e. without having considered relevant factual and normative background propositions such as (1) to (6).

It can easily be proved that the agent's initial dispositions are implicitly, but not explicitly, inconsistent. Let $\Psi_1 = \{Qa, \neg Qb, \forall x \forall y((Qx \wedge \neg Qy) \rightarrow xPy)\}$ and $\Psi_2 = \{Rc, \neg Rd, \forall x \forall y((Rx \wedge \neg Ry) \rightarrow yPx), a = c, b = d\}$. Then Ψ_1 implies aPb, and Ψ_2 implies bPa

(which implies $\neg aPb$). Further, the agent's initial dispositions violate deductive closure: for instance, aPb is implied by a set of propositions which the agent has an initial disposition to accept (e.g. by the set Ψ_1), and yet the agent has no initial disposition to accept aPb itself. As the agent's initial dispositions are implicitly inconsistent with respect to the ranking proposition aPb, theorem 1 immediately implies that there exist two alternative decision-paths such that, under one path, aPb is accepted and, under the other, $\neg aPb$ is accepted. We will now see that this result can be used to explain the Kahneman and Tversky framing problem.

In that problem, the agent is being asked to make a decision between the alternative programs. So the target proposition to be considered by the agent is the ranking proposition aPb in the first presentation of the decision problem, and the ranking proposition cPd in the second presentation. Suppose now that the agent engages in a modus ponens decision process over the propositions. We suggest that the two different presentations of the problem, in terms of "lives saved" and "lives lost", induce two different decision paths or run-ups to the target proposition, respectively:

Path 1: $\forall x \forall y((Qx \wedge \neg Qy) \rightarrow xPy)$ at time 1, Qa at time 2, $\neg Qb$ at time 3, aPb at time 4
Path 2: $\forall x \forall y((Rx \wedge \neg Ry) \rightarrow yPx)$ at time 1, Rc at time 2, $\neg Rd$ at time 3, dPc at time 4.

At time 4, under each path, the agent accepts a ranking proposition, aPb under path 1, and dPc under path 2, which enables the agent to make a choice over the alternative programs, A under path 1, and D under path 2. Both decision-paths are incomplete in that they reach some but not all the relevant propositions which the agent has dispositions over and stop once the target proposition is reached. The inconsistency between the two alternative outcomes would become explicit if each of the two decision-paths were extended and the agent were to consider the propositions $a = c$ and $b = d$ next. Given the acceptance of $a = c$ and $b = d$, the two paths lead to mutually inconsistent outcomes, aPb in path 1 and bPa (which implies $\neg aPb$) in path 2.[6]

[6] In the original Kahneman and Tversky experiments, the two different presentations of the decision problem were given to two different groups of subjects, where the majority of one group preferred A to B, and the majority of the other D to C. Hence there was no opportunity for the same subject to reveal inconsistent preferences under the two alternative presentations. However, Kahneman and Tversky's claim is that "[i]ndividuals who face a decision problem and have a definite preference (i) might have a different preference in a different framing of the same problem, (ii) are normally unaware of alternative frames and of their potential effects on the relative attractiveness of options, (iii) would wish their preferences to be independent of frame, but (iv) are often uncertain how to resolve detected inconsistencies" (Tversky and Kahneman 1981, pp. 457-458).

Note that this result is not dependent on the fact that the agent's initial dispositions are incomplete. This can be illustrated by making the agent's initial dispositions complete, for instance by assuming that (in addition to the initial dispositions specified above) the agent has initial dispositions to accept *aPb* and *cPd*. This can be motivated by the assumption that the agent has already considered the decision problem under the first frame. The modified initial dispositions still satisfy weak consistency but are implicitly inconsistent because they violate deductive closure with respect to the ranking proposition *dPc*. Now theorem 2 immediately implies that there exist two alternative decision-paths such that, under one path, *dPc* is accepted and, under the other, $\neg dPc$ is accepted. We can then give an explanation of the Kahneman and Tversky framing problem similar to the one given in this section.

5. Violations of Procedure Invariance as Path-Dependence

Violations of procedure invariance can also be understood as path-dependence. Take the preference reversal problem of Lichtenstein and Slovic introduced above. Let us define four predicates:

Exy : *x* has a higher expected payoff than *y*.
Sxy : *x* has a higher probability of winning the maximal payoff than *y*.
Txy : *x* has a larger maximal payoff than *y*.
xPy : *x* is strictly preferred to *y*.

As before, as a rule of rationality, we assume that the agent accepts $\forall x \forall y (xPy \rightarrow \neg yPx)$.

Let *p* and *d* denote the P gamble and the $ gamble, respectively.

By the definition of the two gambles, the agent has initial dispositions to accept the following (factual) propositions:

(1) The $ gamble has a larger maximal payoff than the P gamble;
i.e. *Tdp*.
(2) The P gamble has a higher probability of winning the maximal payoff than the $ gamble;
i.e. *Spd*.
(3) Neither gamble has a higher expected payoff than the other;
i.e. $(\neg Epd \wedge \neg Edp)$.

We also suppose that the agent has initial dispositions to accept the following two general (normative) principles:

(4) For two gambles with the same expected payoff, the one with the higher probability of winning the maximal payoff is preferrable;
i.e. $\forall x \forall y((\neg Exy \wedge \neg Eyx) \rightarrow (Sxy \rightarrow xPy))$.

(5) For two gambles with the same expected payoff, the one with the larger maximal payoff is preferrable;
i.e. $\forall x \forall y((\neg Exy \wedge \neg Eyx) \rightarrow (Txy \rightarrow xPy))$.

As in the Kahneman and Tversky problem of the previous section, the agent's initial dispositions are incomplete, as the agent does not have initial dispositions over the ranking propositions *dPp* or *pPd*, and they are implicitly, but not explicitly, inconsistent. If we let $\Psi_1 = \{Tdp, (\neg Epd \wedge \neg Edp), \forall x \forall y((\neg Exy \wedge \neg Eyx) \rightarrow (Txy \rightarrow xPy))\}$ and $\Psi_2 = \{Spd, (\neg Epd \wedge \neg Edp), \forall x \forall y((\neg Exy \wedge \neg Eyx) \rightarrow (Sxy \rightarrow xPy))\}$, then Ψ_1 implies *dPp*, and Ψ_2 implies *pPd* (which implies $\neg dPp$). The agent's initial dispositions also violate deductive closure with respect to *dPp* and *pPd*. Again, as the initial dispositions are implicitly inconsistent with respect to these ranking propositions, theorem 1 implies that there exist two alternative decision-paths such that, under one path, *dPp* is accepted and, under the other, $\neg dPp$ is accepted. In analogy to the explanation of the Kahneman and Tversky problem, this leads to an explanation of the Lichtenstein and Slovic problem.

Again, we can take the target proposition to be the ranking proposition *dPp*. Under one presentation of the decision problem, the ranking between the two gambles determines which gamble to play, and under the other, which gamble to sell for a higher price. Suppose the agent engages in a modus ponens decision process over the propositions. Again, we suggest that the two different presentations of the decision problem, in terms of "which gamble is preferable to play" and "which price to sell each gamble for", induce two different decision paths or run-ups to the target proposition, respectively:

Path 1: $\forall x \forall y((\neg Exy \wedge \neg Eyx) \rightarrow (Sxy \rightarrow xPy))$ at time 1, $(\neg Exy \wedge \neg Eyx)$ at time 2, *Spd* at time 3, *pPd* at time 4.
Path 2: $\forall x \forall y((\neg Exy \wedge \neg Eyx) \rightarrow (Txy \rightarrow xPy))$ at time 1, $(\neg Exy \wedge \neg Eyx)$ at time 2, *Tdp* at time 3, *dPp* at time 4.

At time 4, under each path, the agent accepts a ranking proposition, *pPd* under path 1, and *dPp* (which implies $\neg pPd$) under path 2. The two alternative outcomes are explicitly inconsistent with each other. If asked to choose which of the two gambles to play, i.e. under path 1, the agent would choose the P gamble over the $ gamble. If asked to specify a price for which to sell each gamble, i.e. under path 2, on the other hand, the agent would sell the $ gamble at a higher price than the P gamble. As in the example of the previous section, both decision-paths are incomplete in that they reach some but not all the relevant propositions which the agent has dispositions over and stop once the target proposition is reached.

As before, the result is not dependent on the fact that the agent's initial dispositions are incomplete. If we make the agent's initial dispositions complete, for instance by assuming that the agent has the additional initial disposition to accept *pPd* (for instance, as a result of a previous choice), the modified initial dispositions still violate deductive closure with respect to *dPp*, and theorem 2 yields an explanation of the framing problem similar to the one given here.

6. Discussion

An agent's choices are sometimes not invariant under changes in the way in which the choice problem is framed, be it the way in which the options are described or the way in which preferences are elicited. We have suggested that such framing effects may occur when the agent's initial dispositions on the relevant propositions are implicitly inconsistent. The framing of the decision problem may make particular propositions salient, and thereby induce the order in which the propositions are considered. But if the agent's initial dispositions are implicitly inconsistent, then, by theorem 1, the outcome of a modus ponens decision process is dependent on that order and hence dependent on the framing. Our model highlights similarities between violations of description invariance and procedure invariance, as it models both as being the result of different frames inducing different decision-paths which then lead to different outcomes in a modus ponens decision process (given implicit inconsistencies in the agent's initial dispositions).

Kenneth Arrow said that making the same choices in extensionally equivalent decision problems is, "[a]n elemental effect of rationality, so elemental that we hardly notice it" (Arrow 1982, p.6). Contrary to Arrow, we may hardly notice that our choices are *not* always the same in such decision problems. Although violations of description or procedure invariance are, on the present account, caused by inconsistencies in our initial dispositions, they can be caused by *implicit* inconsistencies, while *explicit* ones are not necessary. But *implicit* inconsistencies are precisely the kinds of inconsistencies an agent is (typically) not explicitly aware of.

Explicit inconsistencies are of course special cases of implicit inconsistencies, i.e. sometimes an implicit inconsistency may also be an explicit one, in that the agent's initial dispositions violate not only strong consistency, but also weak consistency. The fact that some subjects do not change their choices even when the inconsistency is pointed out to them suggests that some people are willing to hold explicitly inconsistent beliefs (Ordonez et al. 1995). When an inconsistency is implicit but not explicit (i.e. when the initial dispositions satisfy weak but not strong consistency), the implicit inconsistency is a result of a violation of deductive closure. By lemma 1, if the initial dispositions satisfy completeness, then a violation of deductive closure is a necessary and sufficient condition for an implicit inconsistency.

Modelling framing as path-dependence is compatible with explanations of framing effects offered by philosophers and psychologists. The notion of normative propositions over which an agent has dispositions parallels the notion of reasons for choice discussed in philosophy and psychology. If agents were asked for their reasons for making a certain choice, they might select those normative propositions they would assent to in the run-up to making the choice. If different presentations of the problem make different normative propositions (or reasons) salient in the run-up to the choice, then all that is necessary for the existence of framing effects is that individuals have a disposition to assent to a set of propositions (or reasons) that have normative force for them, but that are not *strongly* consistent with each other (although they may be *weakly* consistent). What matters for the decision, in our model, is the particular propositions (reasons) that occur in the decision-path, but not other propositions outside the decision-path even if these seem also relevant from the perspective of an external observer. It is certainly conceivable that the totality of reasons that have normative force for an agent may not be strongly consistent. Our model finds support from the suggestion by psychologists that choice is "reason-based", with decision makers seeking and constructing reasons in order to justify their choices (Shafir, Simonson and Tversky 1993). People seem to wish to justify their decisions by saying that they chose for a (single) reason, even to the extent of constructing and selecting the situations they find themselves in so that there is always a dominant reason for choice (Montgomery 1983).

We have used a logician's framework to identify a necessary and sufficient condition for framing effects. The condition is precisely the non-satisfaction of a certain classical condition of rationality, namely the condition of strong consistency. To the extent that violations of that condition, i.e. implicit inconsistencies (or violations of other rationality conditions implying implicit inconsistencies), are prevalent, people will make different choices in extensionally equivalent choice situations.

Appendix

Lemma A1. Suppose X contains neither tautologies nor contradictions. Then

(i) the set $\{\phi \in X : \delta(\phi)=1\}$ is semantically inconsistent

<u>if and only if</u>

(ii) there exist two strongly consistent subsets Ψ_1 and Ψ_2 of X and a proposition ϕ in X such that [$\delta(\psi)=1$ for every ψ in $\Psi_1 \cup \Psi_2$] *and* [Ψ_1 implies ϕ] *and* [Ψ_2 implies $\neg\phi$].

Proof of lemma A1.

(i) implies (ii):

Suppose (i) holds. Let Y be a maximal semantically consistent subset of $\{\phi \in X : \delta(\phi)=1\}$. First, Y is non-empty, since X contains no contradictions. Second, Y is a proper subset of $\{\phi \in X : \delta(\phi)=1\}$, since $\{\phi \in X : \delta(\phi)=1\}$ itself is not semantically consistent. Choose any $\psi \in \{\phi \in X : \delta(\phi)=1\} \backslash Y$. Since Y is a *maximal* semantically consistent subset of $\{\phi \in X : \delta(\phi)=1\}$, $Y \cup \{\psi\}$ is not semantically consistent (otherwise Y would not be maximal), and hence Y implies $\neg\psi$. Now let $\Psi_1 = \{\psi\}$ and $\Psi_2 = Y$. Then Ψ_1 and Ψ_2 have the properties required by (ii).

(ii) implies (i):

Suppose (ii) holds. Since [Ψ_1 implies ϕ] *and* [Ψ_2 implies $\neg\phi$], the set $\Psi_1 \cup \Psi_2$ is semantically inconsistent. But $\{\phi \in X : \delta(\phi)=1\}$ is a superset of $\Psi_1 \cup \Psi_2$. Therefore $\{\phi \in X : \delta(\phi)=1\}$ is also semantically inconsistent. ■

Lemma A2. For any ϕ in X, (i) ϕ is accepted in a modus ponens decision process for *some* decision-path Ω if and only if (ii) there exists a (semantically consistent) subset Ψ of X such that [$\delta(\psi)=1$ for every ψ in Ψ] *and* [Ψ implies ϕ].

Proof of lemma A2.

(i) implies (ii):

Suppose (i) holds. Determine m such that $\Omega(m) = \phi$. Let $\Psi = \{\psi \in X : (\psi$ *is accepted in the modus ponens decision process at time* $t \leq m$) *and* $\delta(\psi)=1\}$. Then Ψ has the properties required by (ii).

(ii) implies (i):

Suppose (ii) holds. Define Ω as follows. Let $m = |\Psi \cup \{\phi\}|$. On $\{1, 2, \ldots, m\}$, let Ω be any bijective mapping from $\{1, 2, \ldots, m\}$ onto $\Psi \cup \{\phi\}$ such that $\Omega(m) = \phi$. To make Ω complete, we add the following definition. On $\{m+1, ..., k\}$ (where $k = |X|$), let Ω be any bijective mapping from $\{m+1, ..., k\}$ onto $X \backslash (\Psi \cup \{\phi\})$. Then Ω has the properties required by (i). ■

References

Arrow, K. 1982. 'Risk Perception in Psychology and Economics' *Economic Inquiry* 20: 1-9

Fischer, G. and S. Hawkins. 1993. 'Strategy Compatibility, Scale Compatibility and the Prominence Effect'. *Journal of Experimental Psychology: Human Perception and Performance,* 19: 580-597

Hamilton, A. G. 1988. *Logic for mathematicians*. Revised edition. Cambridge University Press.

Kahneman, D. and A. Tversky. 1979. 'Prospect Theory: an Analysis of Decision Under Risk'. *Econometrica,* 47: 263-91

Levin, I., S. Schneider and G. Gaeth. 1998. 'All Frames are not Created Equal: A Typology and Critical Analysis of Framing Effects'. *Organizational Behavior and Human Decision Processes*, 76:149-88

Lichtenstein, S. and Slovic, P. 1971. 'Reversal of Preferences Between Bids and Choices in Gambling Decisions'. *Journal of Experimental Psychology,* 89: 46-55

Lichtenstein, S. and Slovic P. 1973. 'Response-Induced Reversals of Preference in Gambling: An Extended Replication in Las Vegas'. *Journal of Experimental Psychology,* 101: 16-20

List, C. 2002. 'A Model of Path-Dependence in Decisions over Multiple Propositions'. Nuffield College Working Paper in Economics 2002-W15

Montgomery, H. 1983. 'Decision Rules and the Search for a Dominance Structure: Towards a Process Model of Decision Making'. In P.Humphreys, O. Svenson and A. Vari eds. *Analysing and Aiding Decision Processes,* 343-369. Amsterdam: North Holland.

Ordonez, L., Mellers, B., Chang, S-J. and Roberts, J. 1995. 'Are Preference Reversals Reduced when Made Explicit?'. *Journal of Behavioral Decision Making,* 8: 265-77

Shafir E., I. Simonson and A. Tversky. 1993. 'Reason Based Choice'. *Cognition,* 49: 11-36

Tversky, A. and D. Kahneman. 1991. ‘Loss Aversion in Riskless Choice: A Reference-Dependent Model’. *The Quarterly Journal of Economics,* 106: 1039-1061

Tversky, A. and D. Kahneman. 1981. 'The Framing of Decisions and the Psychology of Choice'. *Science,* 211: 453-8

Tversky, A, S. Sattath and P. Slovic. 1988. 'Contingent Weighting in Judgement and Choice'. *Psychological Review,* 95: 371-84