

FRAMING AS PATH DEPENDENCE

NATALIE GOLD

University of Oxford

CHRISTIAN LIST

London School of Economics

A framing effect occurs when an agent's choices are not invariant under changes in the way a decision problem is presented, e.g. changes in the way options are described (violation of description invariance) or preferences are elicited (violation of procedure invariance). Here we identify those rationality violations that underlie framing effects. We attribute to the agent a sequential decision process in which a "target" proposition and several "background" propositions are considered. We suggest that the agent exhibits a framing effect if and only if two conditions are met. First, different presentations of the decision problem lead the agent to consider the propositions in a different order (the empirical condition). Second, different such "decision paths" lead to different decisions on the target proposition (the logical condition). The second condition holds when the agent's initial dispositions on the propositions are "implicitly inconsistent," which may be caused by violations of "deductive closure." Our account is consistent with some observations made by psychologists and provides a unified framework for explaining violations of description and procedure invariance.

1. INTRODUCTION

The decisions that people make are sometimes sensitive to the way in which the decision problem is presented. They may depend on the way in which the options are described or on the way in which people's preferences are elicited. They are not always *description invariant* or *procedure invariant*. In a logician's language, two decision problems may

We are grateful to Luc Bovens, Robin Cubitt and three anonymous reviewers for *Economics and Philosophy* for very helpful comments and suggestions.

be *extensionally equivalent* and yet lead to different decisions. 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*. Likewise, if we elicit an agent's preferences over some options in two different ways this should, ideally, not affect the order in which the agent ranks the options. Yet again, empirically, it sometimes does. As we argue here, these phenomena can also be seen as framing effects. Psychologists have offered accounts of decision making to explain why violations of description invariance or procedure invariance occur, but 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 violation may lead to framing effects. We attribute to the agent a sequential decision process in which the agent considers a *target* proposition and several *background* propositions. We suggest that the agent exhibits a framing effect if and only if two conditions are met, one logical and one empirical. The logical condition states that the agent's decision on the target proposition is path dependent, i.e. it varies with the order in which the propositions are considered. This condition is satisfied if and only if the agent's initial dispositions on the propositions are *implicitly inconsistent* – which may be caused by violations of *deductive closure* – in a sense defined below. The empirical condition states that different presentations of the decision problem lead the agent to consider the propositions in different such orders. We suggest that different ways of framing a decision problem may indeed have this effect. This theoretical explanation is consistent with some observations made by psychologists and provides a unified framework within which we can see similarities between violations of description and procedure invariance.

2. 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 a disease that was expected to kill 600 people and that they had to make a choice between two alternative vaccination programs. Two groups were presented with the same decision problem but in different forms. The first group was 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 one will be saved.

The second group was told:

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.

Note that A and C are extensionally equivalent. They denote the same vaccination program, where precisely 200 people will be saved and 400 will die. Likewise, B and D are extensionally equivalent. They denote the same vaccination program, where there is a 1/3 probability that 600 people will be saved and no one will die and a 2/3 probability that no one will be saved and 600 people will die. However, in the first group, 72% of subjects opted for Program A but, in the second group, 78% of subjects chose Program D. Changing the description of the options from one in terms of "lives 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 respondents 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 that do not involve any uncertainty, simply by manipulating subjects' reference points 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 or 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 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).¹

¹ There is an evolutionary story about why we might encode positive and negative stimuli differently. We know that *avoidance* and *approach* responses are processed differently and that signals of pain and danger are accorded priority in neural processing (Kahneman and

3. VIOLATIONS OF PROCEDURE INVARIANCE

Framing effects are often identified with violations of description invariance, the phenomenon that choices are not invariant under changes in the way in which the options are described. But there are also well-documented violations of procedure invariance, where choices are affected by the way in which preferences over the options are elicited. By changing the method of preference elicitation, the same agent can be induced to make inconsistent choices. We suggest that a violation of the same rationality conditions is responsible for both violations of description invariance and violations of procedure invariance. For this reason, we prefer to call both phenomena framing effects.

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 0.80	Win \$9.00 with probability 0.20
Lose \$1 with probability 0.20	Lose \$0.50 with probability 0.80

Both gambles have an expected value of \$1.40. Subjects were asked which gamble they preferred to play (a qualitative *choice* task) and also, in a different stage of the experiment, what price they would sell the right to play each gamble 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 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: 203). However, the pattern of choices was that subjects 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 gambles such as the above, of participants who chose the P gamble over the \$ gamble, 81% specified a higher price for the \$ gamble than for the P gamble and, what is more,

Varey 1991). If negative stimuli were, on the whole, things that led to the death of our pre-human ancestors, such as predators or poisonous plants, whereas positive stimuli resulted in only incremental fitness increases then there may have been particular benefit to having either instinctive avoidance mechanisms or rapid reactions to negative stimuli. As a part of these asymmetric response mechanisms, we may also have evolved to encode the initial positive or negative stimuli differently.

some turned into “money pumps” continuously giving more money to the experimenters to switch between the gambles without ever playing them (Lichtenstein and Slovic 1973).²

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 occurs 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, for instance choice versus valuation. In the example of gambles, the attributes might be the maximal payoff and the probability of winning that maximal payoff. An example of violations of procedure invariance not involving 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 attributes, whereas for the other 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 overall value. 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 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 the responses to the matching task was that only 34% would have rated this candidate as better. This leads to the *prominence hypothesis* that the prominent attribute will weigh more heavily in choice than in matching (Tversky *et al.* 1988).

Psychologists have suggested two rival hypotheses to explain how elicitation procedures affect preferences: the *strategy compatibility hypothesis* and the *scale compatibility hypothesis* (Fischer and Hawkins 1993). On the strategy compatibility hypothesis, different modes of preference elicitation

² These preference reversal effects cannot be explained merely as endowment effects, as Lichtenstein and Slovic (1971) showed them to be robust over different methods of eliciting the valuation including methods where the subject does not start from a position of “owning” the bet.

induce different heuristics. Specifically, choice tasks are held to induce lexicographic reasoning, i.e. a focus on a prominent attribute that is considered lexicographically prior to other attributes. Valuation tasks, in contrast, are held to induce explicit trade-offs between different attributes. According to this hypothesis, in the above example, when subjects are asked to choose their preferred gamble the probability of winning is the prominent attribute. When asked for their monetary valuations of the gambles, they trade off the probability of winning against the maximal payoff and prefer the \$ gamble. On the scale compatibility hypothesis, choices involving multiple attributes of the options are always made using the same heuristic, but different modes of preference elicitation change the weights assigned to these attributes. According to this hypothesis, in the above example, when subjects are asked to choose their preferred gamble the probability of winning is the attribute with the greater weight. When they are asked for their monetary valuations of the gambles the maximal payoff is the attribute with the greater weight.

4. A SIMPLE MODEL OF DECISION MAKING

We seek to explain framing effects by attributing to the agent a sequential decision process in which the agent considers multiple propositions. For this purpose, we redevelop an earlier model of sequential decisions over multiple propositions by List (2004). Our redevelopment differs from the original model in two key respects. First, we represent propositions in *predicate* logic, whereas List (2004) represents propositions in *propositional* logic. The extension to predicate logic allows us to represent preferences over options, relations between options and other considerations in a single unified framework. Second, we focus on decisions made by an individual agent, whereas List's (2004) main emphasis is on collective decisions.

4.1 The key idea

We represent an agent's binary choice between x and y as the agent's assignment of a truth-value to a ranking proposition of the form " x is strictly preferred to y ," formally xPy . This ranking proposition is the *target proposition*, on which the agent has to make a decision. But we assume that the agent considers not only the target proposition, but also certain *background propositions*, which represent the "run-up" or "context" to the agent's decision on the target proposition. They may include factual propositions, on which the agent may have beliefs that are relevant to her decision on the target proposition, and normative propositions whose resolution (i.e. acceptance or non-acceptance) may entail a particular stance on the target proposition. In short, among the background propositions are all those propositions that the agent may consider in the process leading up to her decision on the target proposition.

4.2 The language of predicate logic

We first introduce the language of predicate logic in which the propositions are formalized. (For a precise definition of first-order predicate logic, see Hamilton 1988.) The language includes:

- atomic propositions with zero-place predicates, e.g. P, Q, R, \dots ;
- atomic propositions with one-place predicates, e.g. Aa (“ a has property A ”);
- atomic propositions with two-place predicates, e.g. aEb (“ a stands in relation E to b ”), aPb (“ a is strictly preferred to b ”);
- compound propositions with logical connectives or quantifiers, e.g. $\neg P$, $(P \wedge Q)$, $\forall x(Ax \rightarrow Bx)$, $\forall x\forall y((Ax \wedge \neg Ay) \rightarrow xPy)$.

Propositions take the truth-values “true” or “false.” Truth-value assignments satisfy the standard properties defined in classical first-order predicate logic (again see Hamilton 1988). A set of propositions, S , is *logically consistent* if there exists at least one truth-value assignment under which all propositions in S are true. A set of propositions S *logically entails* a proposition ϕ , if, for *all* truth-value assignments, [if all propositions in S are true, then so is ϕ].

4.3 The propositions considered by the agent

We assume that the agent considers a (finite non-empty) set of propositions, denoted X , from the language just introduced, where X contains both the target proposition and the relevant background propositions.³ If agents were asked for their reasons for making a certain choice, they might refer to those normative and factual propositions they would assent to in the run-up to making the choice. So our notion of background propositions considered by the agent parallels the notion of reasons for choice discussed in philosophy and psychology.

4.4 The agent’s initial dispositions

How do we represent the agent’s attitudes toward the propositions in X ? We assume that, for each proposition ϕ in X , the agent has an *initial disposition* on that proposition. Her *initial disposition* on ϕ is the judgment (acceptance/non-acceptance) she *would* make on ϕ if she *were* to consider ϕ in isolation, with no reference to other propositions. Note that an initial

³ For technical reasons, we assume that (i) X contains neither tautologies nor contradictions; (ii) X includes proposition-negation pairs (i.e. for every ϕ in X , $\neg\phi$ is also in X); (iii) for each ϕ in X , we identify $\neg\neg\phi$ with ϕ . A *tautology* is a proposition which is true under *all* truth-value assignments. A *contradiction* is a proposition which is false under *all* truth-value assignments.

disposition is a *counterfactual* notion. Saying that an agent has an initial disposition on ϕ does not carry any implications as to whether she has *in fact* considered the proposition. The agent's initial dispositions are represented by an *acceptance function* δ assigning a value of 0 or 1 to each proposition ϕ in X , where $\delta(\phi) = 1$ means that the agent has an initial disposition to accept ϕ , and $\delta(\phi) = 0$ means that the agent has an initial disposition not to accept ϕ .

4.5 The decision path

A *decision path* is the order in which the agent considers the propositions in a sequential decision process. Formally, a decision path is represented by a one-to-one function Ω from the numbers $1, 2, 3, \dots, l$ into the set X , where l is less than or equal to the number of propositions, k , in X .⁴ We interpret $\Omega(1), \Omega(2), \dots, \Omega(l)$ as the first, second, \dots , last propositions considered by the agent. Typically, the last proposition considered in the path, $\Omega(l)$, is the target proposition, while the preceding propositions, $\Omega(1), \Omega(2), \dots, \Omega(l - 1)$ are the background propositions. But our model also allows that sometimes the target proposition may occur earlier in the path. A decision path Ω is *exhaustive* if it reaches all propositions in X (i.e. $l = k$), and *non-exhaustive* if it reaches some but not all propositions in X (i.e. $l < k$). A decision path can be interpreted in more than one way: as the temporal order in which the agent considers the propositions, as the order of how focal the propositions are for the agent, or how much weight the agent assigns to the propositions.

4.6 The criteria for the acceptance or non-acceptance of propositions in a sequential decision process

Suppose that the agent considers the propositions one-by-one along a given decision path. When does she accept a proposition and when not? Suppose that proposition ϕ is under consideration. By assumption, the agent has an initial disposition on ϕ . There are two possibilities: either this initial disposition is logically consistent (and perceived so by the agent) with the agent's previously accepted propositions, or it is not. If it is, the agent can make a decision (acceptance/non-acceptance) on the new proposition according to her initial disposition. If it is not, the agent requires a method for resolving the logical conflict. Under the *modus ponens rule*, used below, the agent resolves this conflict by accepting the logical implications of previously accepted propositions and overruling her initial

⁴ The assumption that Ω is one-to-one means that, for all a and b in the domain of Ω , $\Omega(a) = \Omega(b)$ implies $a = b$. The requirement that a decision path be a one-to-one function ensures that no proposition occurs more than once in the path.

disposition on the new proposition.⁵ Attributing the modus ponens rule to an agent is useful for explaining framing effects, as that rule captures the notion that the run-up to the agent's decision on a proposition may *constrain* that decision. We suggest that different decision frames induce different run-ups to an agent's decision on the target proposition.

4.7 A modus ponens decision process

We can now define a *modus ponens decision process*, following List (2004):

- Consider the propositions along a given decision path Ω : proposition $\phi_1 := \Omega(1)$ in step 1, proposition $\phi_2 := \Omega(2)$ in step 2, ..., proposition $\phi_l := \Omega(l)$ in step l .
- For each step $t = 1, 2, \dots, l$, let Φ_t be the set of all propositions accepted in steps 1, 2, ..., t . Define Φ_t inductively as follows (adding step 0):
 $t = 0$: Φ_0 is the empty set.
 $t > 0$: Proposition ϕ_t is considered. There are two cases:

Case I: The set of previously accepted propositions Φ_{t-1} entails ϕ_t or it entails $\neg\phi_t$.

$$\text{Then } \Phi_t := \begin{cases} \Phi_{t-1} \cup \{\phi_t\} & \text{if } \Phi_{t-1} \text{ entails } \phi_t \\ \Phi_{t-1} \cup \{\neg\phi_t\} & \text{if } \Phi_{t-1} \text{ entails } \neg\phi_t. \end{cases}$$

Case II: The set of previously accepted propositions Φ_{t-1} entails neither ϕ_t nor $\neg\phi_t$.

$$\text{Then } \Phi_t := \begin{cases} \Phi_{t-1} \cup \{\phi_t\} & \text{if } \delta(\phi_t) = 1 \\ \Phi_{t-1} & \text{if } \delta(\phi_t) = 0. \end{cases}$$

For a given acceptance function δ and a given decision path Ω , a modus ponens decision process produces an *outcome set*, defined as $M(\delta, \Omega) := \Phi_t$.⁶ In our subsequent examples, the decision process ends when the agent decides on the target proposition, but nothing hinges on the decision path being non-exhaustive.

⁵ Other methods are conceivable. Under the *modus tollens rule*, for example, the agent resolves the conflict by accepting her initial disposition on the new proposition and revising previously accepted propositions.

⁶ The reason for *not* defining $\Phi_t := \Phi_{t-1} \cup \{\neg\phi_t\}$ if $\delta(\phi) = 0$ in case II is to allow separate consideration of $\neg\phi_t$ at a different step from ϕ_t in the decision path. If δ is incomplete (defined below), this definition allows *indecisive* outcome sets, where neither ϕ nor $\neg\phi$ is in $M(\delta, \Omega)$.

5. RATIONALITY CONDITIONS AND THEIR VIOLATION

We first state four rationality conditions which an agent's initial dispositions may or may not satisfy and then consider possible violations of these conditions.

5.1 Rationality conditions

Suppose the agent's initial dispositions are represented by the acceptance function δ .

Completeness. For all propositions ϕ in X , the agent has a disposition to accept either ϕ or its negation, i.e. $\delta(\phi) = 1$ or $\delta(\neg\phi) = 1$.

Weak Consistency. For all propositions ϕ in X , the agent does not have a disposition to accept both ϕ and its negation, i.e. not both $\delta(\phi) = 1$ and $\delta(\neg\phi) = 1$.

Strong Consistency. The propositions in X that the agent has a disposition to accept can all be simultaneously true, i.e. the set $\{\phi \in X : \delta(\phi) = 1\}$ is logically consistent.

Deductive Closure. For any logically consistent set of propositions Ψ and any proposition ϕ , if the agent has dispositions to accept all propositions in Ψ and Ψ entails ϕ , then the agent also has a disposition to accept ϕ , i.e. if $[\delta(\psi) = 1$ for all ψ in $\Psi]$ and Ψ entails ϕ , then $\delta(\phi) = 1$.

The conditions are logically interdependent. Strong consistency implies weak consistency. The conjunction of weak consistency and deductive closure implies strong consistency.

5.2 Rationality violations

Three types of rationality violations are particularly important for our analysis:

Deductive closure violations. An agent's initial dispositions *violate deductive closure* with respect to a proposition ϕ in X if there exists a logically consistent set of propositions Ψ such that the agent has dispositions to accept all propositions in Ψ , Ψ entails ϕ , and yet the agent has no disposition to accept ϕ , i.e. $[\delta(\psi) = 1$ for all ψ in $\Psi]$ and Ψ entails ϕ and $\delta(\phi) = 0$.

Explicit inconsistencies. An agent's initial dispositions are *explicitly inconsistent* with respect to a proposition ϕ in X if the agent has dispositions to accept both ϕ and $\neg\phi$, i.e. $\delta(\phi) = 1$ and $\delta(\neg\phi) = 1$. (So explicit inconsistencies are violations of weak consistency.)

Implicit inconsistencies. An agent's initial dispositions are *implicitly inconsistent* with respect to a proposition ϕ in X if there exist two logically consistent sets of propositions Ψ_1 and Ψ_2 such that the agent has dispositions to accept all propositions in Ψ_1 and all propositions in Ψ_2 (i.e. $\delta(\psi) = 1$ for all ψ in $\Psi_1 \cup \Psi_2$), but Ψ_1 entails ϕ and Ψ_2 entails $\neg\phi$. (By Lemma A1 in the appendix, implicit inconsistencies are violations of strong consistency.)

Implicit inconsistencies can occur in two different ways. First, they occur when the agent's initial dispositions are explicitly inconsistent, so the agent accepts both a proposition and its negation. Second, they occur when some propositions that the agent has a disposition to accept *entail* the negation of what is *entailed* by other propositions that the agent has a disposition to accept, although the agent does not have a disposition to accept those implications themselves. To clarify this with an example, imagine an agent with initial dispositions to accept P , $(P \rightarrow Q)$, $\neg Q$, and no other propositions. These initial dispositions are implicitly inconsistent (they violate strong consistency): the set of propositions accepted by the agent has two logically consistent subsets, $\Psi_1 = \{P, (P \rightarrow Q)\}$ and $\Psi_2 = \{\neg Q\}$, such that Ψ_1 entails Q and Ψ_2 entails $\neg Q$. But the agent's initial dispositions are *not* explicitly inconsistent (they satisfy weak consistency): she does not have a disposition to accept a proposition and its negation simultaneously. In a slight abuse of language, all explicit inconsistencies are also implicit inconsistencies, but not all implicit inconsistencies are also explicit inconsistencies. Note that, if weak consistency is satisfied, deductive closure violations may cause implicit inconsistencies.

Lemma 1. (List 2004) Suppose that the agent's initial dispositions over the propositions (represented by δ) are complete and weakly consistent. Then, for any ϕ in X , the agent's initial dispositions are implicitly inconsistent with respect to ϕ if and only if the agent's initial dispositions are not deductively closed with respect to one of ϕ or $\neg\phi$.⁷

6. LOGICAL AND EMPIRICAL CONDITIONS FOR A FRAMING EFFECT

We first introduce the concept of path dependence in a modus ponens decision process and state necessary and sufficient conditions for path dependence. We then identify logical and empirical conditions for a framing effect.

⁷ A proof is stated in the appendix.

6.1 Path dependence

A modus ponens decision process is *path dependent* with respect to some proposition ϕ in X if there exist at least two decision paths with mutually inconsistent outcomes on ϕ , i.e. there exist two paths Ω_1 and Ω_2 such that, under Ω_1 , ϕ is accepted (i.e. ϕ is in $M(\delta, \Omega_1)$) and, under Ω_2 , $\neg\phi$ is accepted (i.e. $\neg\phi$ is in $M(\delta, \Omega_2)$).⁸ Note the following necessary and sufficient condition for path dependence.

Theorem 1. (List 2004) For any proposition ϕ in X , a modus ponens decision process is path dependent with respect to ϕ if and only if the agent's initial dispositions (represented by δ) are implicitly inconsistent with respect to ϕ .⁹

The conjunction of theorem 1 and lemma 1 yields a necessary and sufficient condition for path dependence in the case where the agent's initial dispositions are complete and weakly consistent.

Theorem 2. (List 2004) Suppose that the agent's initial dispositions over the propositions (represented by δ) are complete and weakly consistent. A modus ponens decision process is path dependent with respect to ϕ if and only if the agent's initial dispositions (represented by δ) violate deductive closure with respect to one of ϕ or $\neg\phi$.¹⁰

6.2 Framing

An agent exhibits a framing effect if two different presentations of a decision problem lead her to make two different decisions on the target proposition. In our model, the agent makes decisions using a modus ponens decision process, based on an acceptance function δ representing her initial dispositions. We suggest that each presentation of a decision problem leads the agent to focus on the background propositions in a particular order, thus inducing a corresponding decision path. In particular, different presentations may induce different decision paths. In terms of our model, an agent therefore exhibits a framing effect if and

⁸ The present use of *path dependence* follows that in List (2004), where it is also compared with the standard social-choice-theoretic use of that concept (e.g. Plott 1973). Typically, path dependence refers to the phenomenon that, in pairwise choices over multiple alternatives, the winning alternative may differ depending on the order in which the choices are made. However, if we identify pairwise choices with pairwise ranking propositions, the standard concept of path dependence can be shown to be a special case of the concept as used here. The relation between preferences and propositions is discussed in List and Pettit (2004).

⁹ A proof is stated in the appendix.

¹⁰ A proof is stated in the appendix.

only if two conditions are met, one logical and one empirical:

The logical condition. There exist two decision paths such that, under one, the agent accepts the target proposition and, under the other, she accepts its negation (or some set of propositions which entail that negation).¹¹

The empirical condition. There exist two ways of presenting the decision problem to the agent that, empirically, lead the agent to use these two decision paths.¹²

Note that the logical condition alone is insufficient for a framing effect: There may exist two *logically possible* decision paths generating different decisions on the target proposition in a modus ponens decision process, and yet there may not exist two corresponding presentations of the decision problem that *empirically* lead the agent to use these paths. (This is related to the more general point, discussed below, that not all logically possible decision paths are empirically feasible.) Theorem 1 leads to the following result:

Result 1. The logical condition for a framing effect is satisfied if and only if the agent's initial dispositions are implicitly inconsistent with respect to the target proposition.

Further, theorem 2 leads to the following result:

Result 2. If the agent's initial dispositions are complete and weakly consistent, the logical condition for a framing effect is satisfied if and only if the agent's initial dispositions violate deductive closure with respect to the target proposition or its negation.

While an implicit inconsistency with respect to the target proposition (or a deductive closure violation with respect to the target proposition or its negation) is necessary and sufficient for the *logical condition* of a framing effect, it is *only* necessary, but *not* sufficient, for the *actual occurrence* of a framing effect, as the empirical condition must also be met.

When is the empirical condition met? More research is needed on this question. However, from a theoretical point of view, we can distinguish between two classes of decision paths: (1) decision paths that are empirically feasible (i.e. that an agent would use under certain empirical conditions) and (2) decision paths that are not empirically

¹¹ This is a logical condition once the set of propositions X , the acceptance function δ and the use of a modus ponens decision process are given. Of course, it remains an empirical matter what the set of propositions X is, what the agent's acceptance function is, and whether the agent uses a modus ponens decision process.

¹² This is an empirical condition as it depends on empirical facts about the agent's psychology and particularly her responsiveness to external stimuli.

feasible (i.e. that an agent would not use under any empirical conditions). Within class (1), there might be a secondary distinction between (1a) decision paths that can be induced externally by some presentation of the decision problem and (1b) decision paths that cannot be induced externally in this manner. An important question for further research on presentations and decision paths is whether or not class (1b) is empty. For the empirical condition to be met in addition to the logical condition, the two logically possible decision paths leading to opposite decisions on the target proposition must both fall into class (1a).¹³

What decision paths are empirically infeasible and thus fall into class (2)? One might think that class (2) includes decision paths in which some factual propositions come *after* some normative ones *and* the modus ponens rule leads the agent to overrule her views on these factual propositions in light of some of the normative propositions accepted earlier. However, it is an empirical question whether people will sometimes revise their factual beliefs on the basis of their normative beliefs, and we cannot rule out *a priori* that they might sometimes do this. Clearly, the question of how large or small class (1) (particularly class (1a)) is compared to class (2) has important implications for the question of how empirically prevalent the phenomena represented by our model are.

But *when* the empirical condition is met, our model suggests that framing effects are caused by certain implicit inconsistencies or deductive closure violations in the agent's initial dispositions.

7. VIOLATIONS OF DESCRIPTION INVARIANCE AS PATH DEPENDENCE

We can use the framework above to illuminate violations of description invariance. Take the disease problem of Kahneman and Tversky. Recall that they asked their subjects to choose between two alternative vaccination programs. For one group of subjects, the two programs were presented in terms of "lives saved" and labeled A and B. For another group, they were presented in terms of "lives lost" and labeled C and D. As noted above, A and C are extensionally equivalent, denoting the same vaccination program. Let us call this program p_{AC} . Likewise, B and D are extensionally equivalent, denoting the same vaccination program. Let us

¹³ Actually, to be precise, it might not be strictly necessary for a framing effect that *both* of the two paths fall into class (1a). It might sometimes be sufficient if one of them, say Ω_1 , falls into class (1a), whereas the other, say Ω_2 , falls into class (1b). This might be the case if Ω_2 is a "default" path that the agent uses for internal reasons (unless externally prompted to use a different path), while Ω_1 is a path that the agent can be externally induced to use by a suitable presentation of the decision problem.

call this program p_{BD} . We 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

We assume that, as a minimal condition of rationality, the agent accepts $\forall x\forall y(xPy \rightarrow \neg yPx)$. We further assume that the agent has initial dispositions to accept the following (factual) propositions (which are true of the programs):

- (1) Program A/C saves some lives (200) with certainty, i.e. Qp_{AC} .
- (2) Program B/D involves the risk that no one will be saved (with probability 2/3 no one will be saved), i.e. $\neg Qp_{BD}$.
- (3) Program A/C consigns some people to death (400) with certainty, i.e. Rp_{AC} .
- (4) Program B/D offers the chance that no one will die (with probability 1/3 no one will die), i.e. $\neg Rp_{BD}$.

We also assume that the agent has initial dispositions to accept the following two (normative) propositions:

- (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 preferable 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 preferable to x , i.e. $\forall x\forall y((Rx \wedge \neg Ry) \rightarrow yPx)$.

Under our assumptions, the agent's initial dispositions are incomplete (though this is not a crucial requirement, as discussed below). The agent has initial dispositions to accept some factual propositions, such as (1) to (4), and some normative propositions, such as (5) and (6). But she does not have initial dispositions to accept the ranking propositions $p_{AC}Pp_{BD}$ or $p_{BD}Pp_{AC}$, as she is unable to accept either of these in isolation, without considering relevant factual and normative background propositions such as (1) to (6).

It is easy to see that the agent's initial dispositions are implicitly, but not explicitly, inconsistent with respect to the ranking proposition $p_{AC}Pp_{BD}$. Let

$\Psi_1 = \{Qp_{AC}, \neg Qp_{BD}, \forall x\forall y((Qx \wedge \neg Qy) \rightarrow xPy)\}$ and $\Psi_2 = \{Rp_{AC}, \neg Rp_{BD}, \forall x\forall y((Rx \wedge \neg Ry) \rightarrow yPx)\}$. Then the agent has initial dispositions to accept the propositions in each of Ψ_1 and Ψ_2 , where Ψ_1 entails $p_{AC}Pp_{BD}$, and Ψ_2 entails $p_{BD}Pp_{AC}$ (which implies $\neg p_{AC}Pp_{BD}$). But there is no proposition such that the agent has a disposition to accept both the proposition and its negation. The agent's initial dispositions also violate deductive closure, as $p_{AC}Pp_{BD}$ is entailed by Ψ_1 and yet the agent has no disposition to accept $p_{AC}Pp_{BD}$ itself.

We argue that these properties of the agent's initial dispositions can be used to explain the framing phenomenon identified by Kahneman and Tversky. In the decision problem given to Kahneman and Tversky's subjects, the target proposition is the ranking proposition $p_{AC}Pp_{BD}$ (or its opposite $p_{BD}Pp_{AC}$). As we have just seen, the agent's initial dispositions are implicitly inconsistent with respect to that proposition, so (by result 1) the agent satisfies the logical condition for a framing effect. Does she also satisfy the empirical condition?

We suggest that the two different presentations of the decision problem, in terms of "lives saved" and "lives lost," induce two different decision paths. The presentation in terms of "lives saved" may induce a decision path starting with factual and normative propositions about saving lives, i.e. propositions (1), (2) and (5). When the agent is prompted to think about "lives saved," she may consider these propositions first, they may become more focal, or they may receive more weight.

Path 1: Qp_{AC} in step 1, $\neg Qp_{BD}$ in step 2, $\forall x\forall y((Qx \wedge \neg Qy) \rightarrow xPy)$ in step 3, $p_{AC}Pp_{BD}$ in step 4.

For parallel reasons, the presentation in terms of "lives lost" may induce a decision path starting with factual and normative propositions about losing lives, i.e. propositions (3), (4) and (6).

Path 2: Rp_{AC} in step 1, $\neg Rp_{BD}$ in step 2, $\forall x\forall y((Rx \wedge \neg Ry) \rightarrow yPx)$ in step 3, $p_{BD}Pp_{AC}$ in step 4.

The outcome set of a modus ponens decision process under path 1 is $\{Qp_{AC}, \neg Qp_{BD}, \forall x\forall y((Qx \wedge \neg Qy) \rightarrow xPy), p_{AC}Pp_{BD}\}$. The outcome set under path 2 is $\{Rp_{AC}, \neg Rp_{BD}, \forall x\forall y((Rx \wedge \neg Ry) \rightarrow yPx), p_{BD}Pp_{AC}\}$. In step 4, under each path, the agent accepts a ranking proposition: $p_{AC}Pp_{BD}$ under path 1 and $p_{BD}Pp_{AC}$ under path 2. This enables the agent to make a choice over the alternative programs, i.e. A/C is chosen under path 1, and B/D is chosen under path 2. The outcomes of the two paths are mutually inconsistent, as the propositions $p_{AC}Pp_{BD}$ and $p_{BD}Pp_{AC}$ cannot both be accepted under the minimal rationality condition introduced above.¹⁴ (Note that both decision

¹⁴ 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

paths are non-exhaustive in that they reach some but not all relevant propositions and stop once the target proposition is reached.)

This result is not dependent on the fact that the agent's initial dispositions are incomplete (although incompleteness would seem to be a realistic assumption here). 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 $p_{AC}Pp_{BD}$ and $\neg p_{BD}Pp_{AC}$. Such an assumption might be motivated by the supposition that the agent has already considered the decision problem under the first presentation. The modified initial dispositions still satisfy weak consistency but (by lemma 1) are implicitly inconsistent because they violate deductive closure with respect to the ranking proposition $p_{BD}Pp_{AC}$. So now (by result 2) the agent satisfies the logical condition for a framing effect. The agent might then satisfy the empirical condition for similar reasons as above.

8. VIOLATIONS OF PROCEDURE INVARIANCE AS PATH DEPENDENCE

We now suggest that violations of procedure invariance can also be understood as path dependence. Take the preference reversal phenomenon identified by Lichtenstein and Slovic. Let p and d denote the P gamble and the \$ gamble, respectively. We define four predicates:

- xEy : x has a higher expected payoff than y .
- xSy : x has a higher probability of winning the maximal payoff than y .
- xTy : x has a larger maximal payoff than y .
- xPy : x is strictly preferred to y .

As before, we assume that, as a minimal condition of rationality, the agent accepts $\forall x\forall y(xPy \rightarrow \neg yPx)$. We further assume that the agent has initial dispositions to accept the following (factual) propositions (which are true of the gambles):

- (1) The \$ gamble has a larger maximal payoff than the P gamble, i.e. dTp .

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, as the agent does in our model. 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: 457–458).

- (2) The P gamble has a higher probability of winning the maximal payoff than the \$ gamble, i.e. pSd .
- (3) Neither gamble has a higher expected payoff than the other, i.e. $(\neg pEd \wedge \neg dEp)$.¹⁵

We also assume that the agent has initial dispositions to accept the following two (normative) propositions:

- (4) For two gambles with the same expected payoff, the one with the higher probability of winning the maximal payoff is preferable, i.e. $\forall x \forall y ((\neg xEy \wedge \neg yEx) \rightarrow (xSy \rightarrow xPy))$.
- (5) For two gambles with the same expected payoff, the one with the larger maximal payoff is preferable, i.e. $\forall x \forall y ((\neg xEy \wedge \neg yEx) \rightarrow (xTy \rightarrow xPy))$.

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 to accept the ranking propositions dPp or pPd , and they are implicitly, but not explicitly, inconsistent. If we let $\Psi_1 = \{dTp, (\neg pEd \wedge \neg dEp), \forall x \forall y ((\neg xEy \wedge \neg yEx) \rightarrow (xTy \rightarrow xPy))\}$ and $\Psi_2 = \{pSd, (\neg pEd \wedge \neg dEp), \forall x \forall y ((\neg xEy \wedge \neg yEx) \rightarrow (xSy \rightarrow xPy))\}$, then Ψ_1 entails dPp , and Ψ_2 entails pPd (which implies $\neg dPp$). The agent's initial dispositions also violate deductive closure with respect to dPp and pPd .

We argue that these properties of the agent's initial dispositions lead to an explanation of the Lichtenstein and Slovic problem, analogous to our explanation of the Kahneman and Tversky problem. We take the target proposition to be the ranking proposition dPp . This requires some explanation. Recall that Lichtenstein and Slovic asked their subjects which gamble they preferred to play and, separately, what price they would sell each gamble for. So subjects had to do two different tasks, a choice task and a valuation task. The use of a single target proposition in our explanation – i.e. the ranking proposition dPp – rests on the theoretical

¹⁵ In the following, we do not mean to imply that the agent actually calculates the expected payoff of the gambles and then disregards the result of the calculation when making her choice. Of course, when two gambles have the same expected payoff, then an expected payoff maximising agent might use another criterion as a tie breaker. But, although Lichtenstein and Slovic used pairs of gambles with comparative expected payoffs, not all the pairs consisted of gambles with exactly the same expected payoff. It is arguably psychologically more realistic that the agent would notice that the gambles have similar expected payoffs, without making precise calculations of them. The proposition that the gambles have the same expected payoff $(\neg pEd \wedge \neg dEp)$, is stronger than this. If anything, this suggests that the preference reversal phenomenon might occur in a wider class of cases than those suggested by the application of the model in this section.

stipulation that the same ranking proposition governs not only the choice of which gamble to play but also the valuation of the gambles.¹⁶ We suggest that using a single target proposition not only leads to a more parsimonious explanation of the Lichtenstein and Slovic problem than using two different ones (one for choice and one for valuation) but also highlights the analogy to the Kahneman and Tversky problem. As we have seen, the agent's initial dispositions are implicitly inconsistent with respect to the target proposition dPp , so, again (by result 1), the agent satisfies the logical condition for a framing effect. We now turn to the empirical condition.

Again, we suggest that the two different presentations of the decision problem, in terms of "which gamble is preferable to play" and "what price to sell each gamble for," induce two different decision paths. There is evidence that, when assessing gambles, *choices* between gambles are determined primarily by the gambles' probabilities, while *valuations* (both selling prices and buying prices) depend mainly on the payoffs (Slovic and Lichtenstein 1968). In a study of preference reversals, Schkade and Johnson (1989) monitored the time spent by subjects looking at probabilities and payoffs and found that subjects who exhibited preference reversals spent more of their time looking at payoffs when pricing gambles than when choosing between them.¹⁷ In line with this evidence, we suggest that asking "which gamble is preferable to play" may induce a decision path starting with the factual and normative propositions about the probability of winning (together with noticing that the two gambles have a similar expected payoff), i.e. propositions (2), (3) and (4).

Path 1: $(\neg pEd \wedge \neg dEp)$ in step 1, pSd in step 2, $\forall x\forall y((\neg xEy \wedge \neg yEx) \rightarrow (xSy \rightarrow xPy))$ in step 3, pPd in step 4.

¹⁶ In fact, in Lichtenstein and Slovic's experimental design, the choice task was comparative (the subject had to choose between the two gambles), whereas the valuation task was non-comparative (the subjects were asked not for a *relative* valuation of the two gambles, but for an *absolute* valuation of each gamble). Our theoretical explanation, however, represents both tasks as comparative, so we implicitly assume that relative and absolute valuation are ordinarily equivalent. So, our theoretical representation rests on the hypothesis that the (preference) reversal phenomenon identified by Lichtenstein and Slovic is driven primarily by the choice/valuation distinction between the tasks, rather than the comparative/non-comparative distinction. In particular, our explanation suggests that a (preference) reversal phenomenon will occur even in two comparative tasks, where probability of winning is salient in one and monetary value in another.

¹⁷ This is consistent with psychological theory. Advocates of the strategy compatibility hypothesis specifically suggest that the probability of winning is the prominent attribute in choice. The strategy compatibility hypothesis and the scale compatibility hypothesis both suggest that the monetary component of the gambles is the more important attribute in valuation.

For parallel reasons, asking “what price to sell each gamble for” may induce a decision path starting with the factual and normative propositions about the maximal payoff, i.e. propositions (1), (3) and (5).

Path 2: $(\neg pEd \wedge \neg dEp)$ in step 1, dTp in step 2, $\forall x\forall y((\neg xEy \wedge \neg yEx) \rightarrow (xTy \rightarrow xPy))$ in step 3, dPp in step 4.

The outcome set of a modus ponens decision process under path 1 is $\{(\neg pEd \wedge \neg dEp), pSd, \forall x\forall y((\neg xEy \wedge \neg yEx) \rightarrow (xSy \rightarrow xPy)), pPd\}$. The outcome set under path 2 is $\{(\neg pEd \wedge \neg dEp), dTp, \forall x\forall y((\neg xEy \wedge \neg yEx) \rightarrow (xTy \rightarrow xPy)), dPp\}$. In step 4, under each path, the agent accepts a ranking proposition: pPd under path 1 and dPp under path 2. 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, the agent would sell the \$ gamble at a higher price than the P gamble. When represented in terms of the same target proposition – i.e. dPp – the two outcomes are mutually inconsistent, as the propositions pPd and dPp cannot both be accepted under the minimal rationality condition introduced above. (As before, both decision paths are non-exhaustive in that they reach some but not all relevant propositions and stop once the target proposition is reached.)

Again, the result does not depend 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 (maybe as a result of a previous choice), the modified initial dispositions still violate deductive closure with respect to dPp and result 2 yields an explanation of the framing problem similar to the one given here.

9. DISCUSSION

An agent’s decisions are sometimes not invariant under changes in the way in which the decision problem is framed, be it the way in which the options are described or the way in which preferences are elicited. We have presented a model of decision making where an agent considers a number of background propositions in the run-up to making a decision on the target proposition, and we have suggested that framing effects may occur when the agent’s initial dispositions on the relevant propositions are implicitly inconsistent. Our approach highlights similarities between violations of description invariance and procedure invariance, in that it represents both as the result of different frames inducing different decision paths that lead to different decisions on the target proposition.

One can identify two traditions in research on decision making: value-based and reason-based (Shafir, Simonson, and Tversky 1993). Value-based accounts typically specify a numerical value function over alternative

outcomes and represent choice as the maximisation of value. Reason-based accounts identify how reasons and arguments influence choice, generally without the use of formal models. Our model falls into the reason-based tradition, formalising some of the ideas underlying such accounts. Our notion of background propositions 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 propositions they would assent to in the run-up to making that choice. As with formal value-based models, we attribute to the agent a preference relation representing the agent's choice. In our model, however, the preference relation is formalized not in terms of a utility function, but in terms of ranking propositions, where those ranking propositions are considered in a *modus ponens* decision process along with other background propositions. The decision process generates a set of accepted propositions. Choices are based on the ranking propositions accepted in that process.

Unlike value-based accounts of choice our model does not represent choices explicitly in terms of numerical trade-offs. However, as mentioned above, violations of description or procedure invariance also occur in situations where there is no risk, so it might be considered as an asset of our account that it can explain a family of effects within a unified framework. Also, it appears that people wish to justify their decisions by saying that they chose for a (single) reason, even to the extent of constructing and selecting choice situations such that there is always a dominant reason for choice (Montgomery 1983). Under an interpretation where the ordering of the propositions given by the decision path represents the importance accorded to various considerations, our model has much in common with lexicographic proposals such as *take the best* (Gigerenzer *et al.* 2000) and *elimination by aspects* (Tversky 1972), where options are explicitly compared on one dimension at a time. We have argued that people who exhibit framing effects endorse reasons that fail certain tests of consistency. Our model focuses on the formal properties of these reasons; it does not specify or restrict their content. We do not claim that people never think numerically or make trade-offs. An agent's reasons may include explicit calculations of risk and expected value. In a reason-based approach such calculations can simply be seen as reasons for certain choices (Schafir, Simonson, and Tversky 1993).

Our model captures some features of decision making that psychologists have observed and that reason-based accounts seek to incorporate. The decision path leading up to the target proposition represents the context in which a choice is made. The path used depends on the formulation of the question, and it may end when the target proposition is reached, even when other background propositions have not been considered. What matters for the decision, in our model, are

the particular propositions (reasons) that occur in the decision path, but not other propositions outside that path even if these also seem relevant from the perspective of an external observer. This is akin to the notion of *concrete thinking*, whereby decision-makers use only surface information (Slovic 1972). Judgments are based only upon explicitly stated probabilities and payoffs. Underlying characteristics of the alternatives do not exert any significant influence, and information that is derivable from, but not explicitly included in, the presentation of the decision problem tends to be ignored (Fischhoff, Slovic, and Lichtenstein 1978). Psychologists have noted a tendency for “considerations that are out of sight to be out of mind” (Slovic, Fischhoff, and Lichtenstein 1988: 153).

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: 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 our account, caused by inconsistencies in an agent’s initial dispositions, *implicit* inconsistencies are sufficient, while *explicit* ones are not necessary. Explicit inconsistencies are of course special cases of implicit inconsistencies. The fact that some subjects do not change their choices even when the inconsistency is made explicit to them suggests that some people are willing to hold explicitly inconsistent beliefs (Ordonez *et al.* 1995). But even an agent who checked her initial dispositions to rule out explicit inconsistencies might not be aware of implicit inconsistencies in these dispositions. More computationally demanding checking would be required to discover implicit inconsistencies, as the agent would have to compute all the logical implications of all the (sets of) propositions she has dispositions to accept. The logical condition for a framing effect is met if the agent has a disposition to accept a set of propositions (or reasons) that have normative force for her, but that have different *implications* for some target proposition and thus violate *strong* consistency (although they may be *weakly* consistent). It is certainly conceivable that the totality of reasons that have normative force for an agent may violate strong consistency in this manner. Now the empirical condition for a framing effect is met if different presentations of a decision problem make different reasons with different such implications salient. To the extent that these two conditions are satisfied, people will make different choices in extensionally equivalent choice situations.

APPENDIX

The proofs in this appendix follow List (2004).

Lemma A1. Recall that X contains neither tautologies nor contradictions. The following holds: (i) $\{\phi \in X : \delta(\phi) = 1\}$ is logically inconsistent if and only if

(ii) there exist two logically consistent subsets $\Psi_1, \Psi_2 \subseteq X$ and a proposition $\phi \in X$ such that $[\delta(\psi) = 1$ for all $\psi \in \Psi_1 \cup \Psi_2]$ and Ψ_1 entails ϕ and Ψ_2 entails $\neg\phi$.

Proof of lemma A1.

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

(ii) implies (i): Suppose that (ii) holds. Since Ψ_1 entails ϕ and Ψ_2 entails $\neg\phi$, $\Psi_1 \cup \Psi_2$ is logically inconsistent. But $\{\phi \in X : \delta(\phi) = 1\} \supseteq \Psi_1 \cup \Psi_2$. Therefore $\{\phi \in X : \delta(\phi) = 1\}$ is also logically inconsistent. ■

Lemma A2. Suppose that the agent uses a modus ponens decision process. Then, for any $\phi \in X$, (i) there exists a decision path Ω such that ϕ is accepted in a modus ponens decision process for Ω (i.e. $\phi \in M(\delta, \Omega)$) if and only if (ii) there exists a logically consistent subset $\Psi \subseteq X$ such that $[\delta(\psi) = 1$ for all $\psi \in \Psi]$ and Ψ entails ϕ .

Proof of lemma A2. Let ϕ be any proposition in X .

(i) implies (ii): Suppose that (i) holds. Let Ω be a decision path such that $\phi \in M(\delta, \Omega)$. Choose t such that ϕ is accepted in step t in the decision process under path Ω . Let

$$\Psi = \{\psi \in X : \delta(\psi) = 1 \text{ and } \psi \text{ is accepted at some step } m < t \text{ under } \Omega\}.$$

The fact that ϕ is accepted in step t implies that either $\delta(\phi) = 1$ or Ψ entails ϕ . If $\delta(\phi) = 1$, then $\Psi = \{\phi\}$ has the properties required by (ii); Ψ is logically consistent, as, by assumption, ϕ is not a contradiction. If Ψ entails ϕ , then Ψ has the properties required by (ii); Ψ is logically consistent, as $\Psi \subseteq M(\delta, \Omega)$, which is logically consistent.

(ii) implies (i): Suppose that (ii) holds. Construct Ω as follows. Let $t = |\Psi \cup \{\phi\}|$. On $\{1, 2, \dots, t\}$, let Ω be any bijective mapping from $\{1, 2, \dots, t\}$ onto $\Psi \cup \{\phi\}$ such that $\Omega(t) = \phi$. To make Ω exhaustive, we extend the definition of Ω as follows. On $\{t+1, \dots, k\}$ (where $k = |X|$), let Ω be any bijective mapping from $\{t+1, \dots, k\}$ onto $X \setminus (\Psi \cup \{\phi\})$. Then Ω has the properties required by (i). ■

Proof of theorem 1. Let ϕ be any proposition in X . Let (i) and (ii) denote the left and right sides of the biconditional, respectively.

(i) implies (ii): Suppose that (i) holds. Since there exists a decision path Ω_1 such that $\phi \in M(\delta, \Omega_1)$, lemma A2 implies that there exists a logically consistent subset $\Psi_1 \subseteq X$ such that $[\delta(\psi) = 1$ for all $\psi \in \Psi_1]$ and Ψ_1 entails ϕ . Similarly, since there exists a decision path Ω_2 such that $\neg\phi \in M(\delta, \Omega_2)$, lemma A2 implies that there exists a logically consistent subset $\Psi_2 \subseteq X$ such that $[\delta(\psi) = 1$ for all $\psi \in \Psi_2]$ and Ψ_2 entails $\neg\phi$. Therefore the agent's initial dispositions are implicitly inconsistent with respect to ϕ .

(ii) implies (i): Suppose that (ii) holds. Then there exist two logically consistent subsets $\Psi_1, \Psi_2 \subseteq X$ such that $[\delta(\psi) = 1 \text{ for all } \psi \in \Psi_1 \cup \Psi_2]$ and Ψ_1 entails ϕ and Ψ_2 entails $\neg\phi$. By lemma A2, since Ψ_1 is a logically consistent subset of X such that $[\delta(\psi) = 1 \text{ for all } \psi \in \Psi_1]$ and Ψ_1 entails ϕ , there exists a decision path Ω_1 such that $\phi \in M(\delta, \Omega_1)$. Similarly, by lemma A2, since Ψ_2 is a logically consistent subset of X such that $[\delta(\psi) = 1 \text{ for all } \psi \in \Psi_2]$ and Ψ_2 entails $\neg\phi$, there exists a decision path Ω_2 such that $\neg\phi \in M(\delta, \Omega_2)$. ■

Proof of lemma 1. Suppose that δ is complete and weakly consistent. Let ϕ be any proposition in X . Let (i) and (ii) denote the left and right sides of the biconditional, respectively.

(i) implies (ii): Suppose that (i) holds. Then there exist two logically consistent subsets $\Psi_1, \Psi_2 \subseteq X$ such that $[\delta(\psi) = 1 \text{ for all } \psi \in \Psi_1 \cup \Psi_2]$ and Ψ_1 entails ϕ and Ψ_2 entails $\neg\phi$. Since δ is weakly consistent, $\delta(\phi) = 0$ or $\delta(\neg\phi) = 0$. If $\delta(\phi) = 0$, then δ violates deductive closure with respect to ϕ , since $[\delta(\psi) = 1 \text{ for all } \psi \in \Psi_1]$, Ψ_1 entails ϕ and $\delta(\phi) = 0$. If $\delta(\neg\phi) = 0$, then δ violates deductive closure with respect to $\neg\phi$, since $[\delta(\psi) = 1 \text{ for all } \psi \in \Psi_2]$, Ψ_2 entails $\neg\phi$ and $\delta(\neg\phi) = 0$.

(ii) implies (i): Suppose that (ii) holds. Case 1: δ violates deductive closure with respect to ϕ . Then there exists a logically consistent set $\Psi_1 \subseteq X$ such that $[\delta(\psi) = 1 \text{ for all } \psi \in \Psi_1]$, Ψ_1 entails ϕ and $\delta(\phi) = 0$. Since δ is complete, $\delta(\neg\phi) = 1$. Let $\Psi_2 = \{\neg\phi\}$, which is logically consistent since $\neg\phi$ is not a contradiction. Case 2: δ violates deductive closure with respect to $\neg\phi$. Then there exists a logically consistent set $\Psi_2 \subseteq X$ such that $[\delta(\psi) = 1 \text{ for all } \psi \in \Psi_2]$, Ψ_2 entails $\neg\phi$ and $\delta(\neg\phi) = 0$. Since δ is complete, $\delta(\phi) = 1$. Let $\Psi_1 = \{\phi\}$, which is logically consistent since ϕ is not a contradiction. In both cases, the existence of Ψ_1 and Ψ_2 establishes that δ is implicitly inconsistent with respect to ϕ . ■

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* 19:580–97
- Fischhoff, B., P. Slovic and S. Lichtenstein. 1978. Fault trees: sensitivity of estimated failure probabilities to problem representation. *Journal of Experimental Psychology: Human Perception and Performance* 4:330–44
- Gigerenzer, G., P. M. Todd and ABC Research Group 2000. *Simple Heuristics That Make Us Smart*. Oxford University Press
- 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
- Kahneman, D. and C. Varey. 1991. Notes on the psychology of utility. In J. Elster and J. Roemer, eds. *Interpersonal Comparisons of Wellbeing*. Cambridge University Press
- 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 P. Slovic. 1971. Reversal of preferences between bids and choices in gambling decisions. *Journal of Experimental Psychology* 89:46–55
- Lichtenstein, S. and P. Slovic. 1973. Response-induced reversals of preference in gambling: an extended replication in Las Vegas. *Journal of Experimental Psychology* 101:16–20

- List, C. 2004. A model of path dependence in decisions over multiple propositions. *American Political Science Review* 98, forthcoming
- List, C. and P. Pettit. 2004. Aggregating sets of judgments: two impossibility results compared. *Synthese* 140:207–35.
- 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* 343–69. Amsterdam: North Holland
- Ordóñez, L., B. Mellers, S.-J. Chang and J. Roberts. 1995. Are preference reversals reduced when made explicit?. *Journal of Behavioral Decision Making* 8:265–77
- Plott, C. 1973. Path dependence, rationality, and social choice. *Econometrica* 41:1075–91
- Schkade, D. and E. Johnson. 1989. Cognitive processes in preference reversals. *Organizational Behavior and Human Performance* 44:203–31
- Shafir E., I. Simonson and A. Tversky. 1993. Reason based choice. *Cognition* 49:11–36
- Slovic, P. 1972. From Shakespeare to Simon: speculations – and some evidence – about man's ability to process information. *ORI Research Monograph*: 1–19.
- Slovic, P. and S. Lichtenstein. 1968. The relative importance of probabilities and payoffs in risk taking. *Journal of Experimental Psychology Monograph Supplement* 78:1–18.
- Slovic, P., B. Fischhoff and S. Lichtenstein. 1988. Response mode, framing, and information-processing effects in risk assessment. In D. Bell, H. Raiffa, and A. Tversky, eds. *Decision making: Descriptive, normative and prescriptive interactions*. Cambridge University Press
- Tversky, A. 1972. Elimination by aspects: a theory of choice. *Psychological Review* 79:281–99
- Tversky, A. and D. Kahneman. 1981. The framing of decisions and the psychology of choice. *Science* 211:453–8
- Tversky, A. and D. Kahneman. 1991. Loss aversion in riskless choice: a reference-dependent model. *The Quarterly Journal of Economics* 106:1039–61
- Tversky, A., S. Sattath and P. Slovic. 1988. Contingent weighting in judgement and choice. *Psychological Review* 95:371–84