European Economic Review 42 (1998) 523–542

Bounded Rationality in Strategic Interaction Contexts

# Recent developments in modeling unforeseen contingencies

Eddie Dekel[a], Barton L. Lipman[b,*], Aldo Rustichini[1c]

[a] *Economics Department, Northwestern University, Evanston, IL, USA*
[b] *Department of Economics, Social Science Centre, University of Western Ontario, London, Ont., N6A 5C2, Canada*
[c] *CORE, Louvain-La-Neuve, Belgium*

**Abstract**

We survey recent models of unforeseen contingencies, discussing both epistemic and decision–theoretic approaches. We also briefly comment on the hurdles which remain for applying these models to contract theory. © 1998 Elsevier Science B.V. All rights reserved.

*JEL classification:* D81; D84

*Keywords:* Unforeseen contingencies; Incomplete contracts

## 1. Introduction

Many writers have suggested that the nature of contracting, firm structure, and political constitutions cannot be well understood without taking account of the role of *unforeseen contingencies* – loosely, possibilities that the agent does not

* Corresponding author. E-mail: blipman@julian.uwo.ca.
[1] Also at: CentER, Tilburg, The Netherlands.


*PII* S0014-2921(97)00114-1

'think about' at the time he makes a decision. In any complex situation, real people do not consider all of the many possible situations which may arise. As a result, contracts typically assign broad categories of rights rather than calling for very specific actions as a function of what might occur. Similarly, firms are designed to decide what to do rather than simply programmed to implement some given set of actions. Finally, laws, especially sweeping ones such as constitutions, are intentionally left vague to allow adaptation to circumstances as they arise.[2] While the importance of unforeseen contingencies seems clear, unfortunately, it is not obvious how we might model them.

We survey recent attempts to provide such a model. While progress has been made, it is too early to say that a useful and convincing model has been found. Each of the papers we discuss has formalized an aspect of the problem in a useful way; none of them is entirely successful. We emphasize that several of the papers we discuss have contributions that go beyond our narrow focus and our criticisms may not carry over to other aspects of these papers. In particular, our brief summaries oversimplify the work discussed and so are no substitute for the originals.

Before discussing the approaches taken, some clarifications are useful. First, as we use the term, unforeseen contingencies are *not* events that the agent has considered but assigned zero probability. This notion does not require a new model. One way to view the difference is that with an unforeseen contingency, an 'uninformative' statement – such as 'event $x$ might or might not happen' – can change the agent's decision. Second, an unforeseen contingency is not necessarily one the agent *could not* conceive of, just one he *doesn't* think of at the time he makes his choice. In the context of contracting, we think that the agent could typically include the unforeseen contingencies if he took enough time, but that he does not get around to doing so.

Finally, it is important to clarify what we want in a model of unforeseen contingencies. Not every situation with unforeseen contingencies requires a model to understand. For example, consider the problem of the route to take to work. Of course, there is a wide variety of considerations that could, in principle, be relevant to the decision: likely construction sites, ambulance traffic, convenient places to stop for gas, etc. Realistically, no one considers many of these factors in making a choice of route. On the other hand, as long as the agent has a good estimate of the length of time for each route, it does not matter which factors are foreseen and which are not. There are no decisions that would be affected by the foreseeability or unforeseeability of these contingencies.

In contracting, unforeseen contingencies matter more. Consider, for example, the choice between long-term and short-term contracts. An important advantage

[2] See, for example, the discussion in Hayek (1960) (Chapter 12).

of short-term contracts is that it is easier to anticipate the relevant contingencies for the near future than for the distant future. Hence, the extent of unforeseen contingencies is crucial to the choice between these options. What we want is a model which can allow us to understand this kind of problem. Note that the zero-probability approach will not work. If the only sense in which some distant future contingencies are not foreseen is that they are given zero probability, then agents will perceive zero costs to excluding them. Hence they will see no 'foreseeability' advantage to short-term contracts.

There are two main approaches to modeling unforeseen contingencies. The *epistemic approach* focuses on the beliefs of an agent, trying to describe how unforeseen contingencies fit in. The *decision theoretic approach* starts from the agent's preferences and asks what view of the world we can attribute to the agent that would be consistent with them. This approach doesn't necessarily seek an accurate description of the agent's beliefs, but instead takes an 'as-if' approach, saying the agent behaves as if these were his beliefs. We discuss each of these approaches, focusing our discussion entirely on the single agent case for simplicity and because, in some cases, we know little about the multi-agent case.

## 2. Epistemic approaches

Epistemic approaches begin with the standard model of knowledge, generalizing it to allow for unforeseen contingencies.[3] The standard model has a set of *states of the world* $\Omega$. A state specifies everything relevant to the problem – every fact about the world and everything the agent knows about these facts. This knowledge is represented via a partition of $\Omega$, denoted $\Pi$. Let $\pi(\omega)$ be the event in $\Pi$ which contains $\omega$. The interpretation is that at state $\omega$, the agent knows only that the true state is in $\pi(\omega)$.

To understand how the standard model rules out unforeseen contingencies requires a more general model in which it is a special case. The best-known way to do this is using what Geanakoplos (1989)[4] called *possibility correspondences*. He suggested that we replace the partition with a function $P : \Omega \rightarrow 2^{\Omega}\backslash\{\emptyset\}$ where in state $\omega$, $P(\omega)$ is the set of states the agent thinks are possible. In other words, we drop the requirement that the sets $\{P(\omega)|\omega \in \Omega\}$ form a partition of $\Omega$. In the case where they do (and where $\omega \in P(\omega)$), we say that $P$ is *partitional* and otherwise that it is *nonpartitional*.

For our purposes, it is useful to work with a *knowledge operator*. Roughly, a possibility correspondence says what the agent believes possible at each

[3] See Geanakoplos (1989) or Dekel and Gul (1997) for more detailed discussions of these topics.
[4] See also Samet (1990) and Shin (1993).

state, while a knowledge operator rewrites this giving for each event $E$, the states at which the agent knows $E$. Formally, given an event $E \subseteq \Omega$, let $K(E) = \{\omega \in \Omega | P(\omega) \subseteq E\}$. To understand this, suppose that $P(\omega) \subseteq E$. Then every state the agent considers possible is in $E$, so, in this sense, the agent 'knows' that the true state must be in $E$. Put differently, in every state the agent considers possible at $\omega$, $E$ is true. Hence at $\omega$, the agent knows $E$ is true. Letting $\neg E$ denote the complement of $E$, $\neg K(E)$ is the set of states where it is *not* true that the agent knows $E$ to be true.

One necessary condition for a possibility correspondence to be partitional, called *negative introspection*, is $\neg K(E) \subseteq K \neg K(E)$. In words, this says that if the agent does not know $E$ is true, he must know that he does not know that $E$ is true. To understand this condition, first suppose there are no unforeseen contingencies, so the agent recognizes every possibility. Then if the agent does not know some statement is true, he should recognize this fact. Since he knows all the possibilities, he can mentally go through the list of all possibilities, checking off the ones he knows to be true, the ones he knows to be false, and the ones he is unsure about. For example, consider a fan's knowledge of how various NBA teams fared against one another during the 1996–1997 regular season. Since a fan, by definition, knows all possible pairs of teams, he recognizes the fact that he, for example, does not know how Vancouver did against Sacramento.

On the other hand, suppose the possibility that the NBA sets up a franchise in Toulouse is an unforeseen contingency for him. Then does he know that the NBA is going to do this? No, of course not. Does he recognize his lack of knowledge about this specific point? Again, of course not. Hence negative introspection fails, so partitional models cannot capture unforeseen contingencies.

Geanakoplos gives the following example to suggest that possibility correspondences can provide a useful model of unforeseen contingencies. Suppose a scientist performs an experiment which is expected to be routine. If the ozone is disintegrating – a possibility which has never occurred to her – then the experiment will yield an unexpected outcome, leading her to discover the disintegration of the ozone layer. On the other hand, if the ozone is not disintegrating, the 'expected' result will occur and the scientist will never realize that there was a chance that the ozone might be disintegrating. To model this, let $\Omega = \{a, b\}$. In $a$, the ozone is disintegrating, while in state $b$, it is not. Let the possibility correspondence be $P(a) = \{a\}$, but $P(b) = \Omega$. Then, as in the story, if the ozone is disintegrating, the scientist discovers this, while otherwise, she learns nothing. Intuitively, in state $b$, she fails to recognize that the state cannot be $a$ because she fails to foresee the possibility of getting the 'unexpected' result. This 'mistake' is reflected in a failure of negative introspection. In particular, $K(\{a\}) = \{a\}$, so $\neg K(\{a\}) = \{b\}$. However, $K \neg K(\{a\}) = K(\{b\}) = \emptyset$. Hence $\neg K(\{a\}) = \{b\} \not\subseteq \emptyset = K \neg K(\{a\})$, a violation of negative introspection.

Intuitively, at $b$, the agent does not know that $\{a\}$ is true and does not know that she does not know. In this sense, at $b$, the agent has failed to foresee the possibility of $a$.

On the other hand, the example is puzzling in several respects. In state $b$, the scientist is supposed to not recognize the possibility that the ozone is disintegrating. Yet she considers state $a$ (the one where the ozone is disintegrating) to be possible. How can she be said to consider it possible and yet not recognize its meaning?

Modica and Rustichini (1994) offer another critique. They suggest we define the agent to be unaware of (to fail to foresee) an event $E$ at state $\omega$ whenever such a failure of negative introspection occurs. That is, letting $U(E)$ denote the set of states where the agent is unaware of $E$, they define $U(E) = \neg K(E) \cap \neg K \neg K(E)$. They argue that if an agent is unaware of the possibility that $p$ is true, then she should be unaware of the possibility that it is false. Hence $U$ should be *symmetric* in the sense that $U(E) = U(\neg E)$. This assumption is violated in the example. As noted, at state $b$, the agent does not know and does not know that she does not know $\{a\}$. Symmetry, then, would dictate that she not know and not know that she does not know $\{b\}$. It is true that at state $b$, the agent does not know $b$ since $K\{b\} = \emptyset$. However, $K \neg K\{b\} = K\Omega = \Omega$, so she does know that she does not know that $b$ is true.

Modica and Rustichini show that the requirement of symmetry (together with some other assumptions) eliminates unforeseen contingencies in a possibility-correspondence model. Hence, they conclude, dropping negative introspection alone does not provide an adequate model of unforeseen contingencies.

Dekel et al. (1997a) reject the example for a different reason. Recall that $\neg K\{a\} = \{b\}$ and $\neg K \neg K\{a\} = \Omega$. Hence at $b$, the agent does not know that $a$ is true nor does she know that she does not know $a$. On the other hand, $K \neg K \neg K\{a\} = K\Omega = \Omega$. That is, she does know that she does not know that she does not know $a$. This form of knowledge is certainly odd; however, it seems to suggest that she has positive knowledge of $a$ at some level and hence should *not* be said to be unaware of the possibility.

The Dekel–Lipman–Rustichini critique uses knowledge operators, not possibility correspondences. Above, we derived knowledge operators from possibility correspondences. However, not every knowledge operator can be so derived. It is well known that a knowledge operator derived from a possibility correspondence satisfies *monotonicity* ($E \subseteq F$ implies $K(E) \subseteq K(F)$) and *necessitation* ($\mathrm{K}(\Omega) = \Omega$). Dekel–Lipman–Rustichini show that any knowledge operator satisfying *either* monotonicity *or* necessitation cannot give a nontrivial model of unforeseen contingencies. Specifically, consider an unawareness operator $U : 2^{\Omega} \to 2^{\Omega}$ where, again, $U(E)$ is interpreted as the set of states where the agent is unaware of the possibility of $E$. Dekel–Lipman–Rustichini put three axioms on $U$ and $K$:

1. $U(E) \subseteq U(U(E))$. That is, if the agent is unaware of the possibility of $E$, he must be unaware of the possibility of being unaware of $E$.
2. $U(E) \subseteq \neg K(E) \cap \neg K \neg K(E)$. In other words, unawareness can only occur when negative introspection is violated.
3. $KU(E) = \emptyset$. The agent never knows he is unaware of a specific event $E$. A person may know that he is unaware of *something*, but should not know which event he is unaware of.

Dekel–Lipman–Rustichini show that given these three axioms, necessitation implies there is no unawareness ($U(E) = \emptyset$ for all $E$), while monotonicity implies there can only be unawareness when the agent has *no* knowledge ($U(E) \subseteq \neg K(F)$ for all $E$ and $F$). To see this, note that the first axiom says $U(E) \subseteq UU(E)$. Applying the second axiom to the event $U(E)$ gives $U(U(E)) \subseteq \neg K \neg K(U(E))$, so $U(E) \subseteq \neg K \neg K(U(E))$. Rewriting the third axiom, $\neg KU(E) = \Omega$. Hence, $\neg K \neg KU(E) = \neg K\Omega$, so $U(E) \subseteq \neg K(\Omega)$. But necessitation says $K(\Omega) = \Omega$ or $\neg K(\Omega) = \emptyset$ and so implies $U(E) = \emptyset$. Monotonicity says $K(F) \subseteq K(\Omega)$ or $\neg K(\Omega) \subseteq \neg K(F)$. Hence it implies $U(E) \subseteq \neg K(F)$.

Dekel–Lipman–Rustichini also show that no ‘standard state-space model’ can deliver a nontrivial model of unforeseen contingencies. A rough intuition for this result is that in standard state-space models, states play two distinct roles: they are the analyst’s descriptions of ways the world might be and they are also the agent’s descriptions of ways the world might be. If the agent is unaware of some possibility, ‘his’ states should be less complete than the analyst’s. Hence, any model which does not explicitly distinguish between the agent’s descriptions and the analyst’s will fail to capture unforeseen contingencies.

One response is to consider models which do make this distinction. Dekel–Lipman–Rustichini sketch one way this might be done using an approach based on Modica and Rustichini (1993) which also has similarities to Fagin and Halpern (1988) and to three-valued logics.[5] Pires (1994) (Chapter 1) also proposed a model that makes this distinction.

Alternatively, we may give up on ‘realistic’ models of the agent’s view of the world and seek an ‘as if’ model – one which predicts the agent’s behavior correctly, even if it is not descriptively accurate. Epistemic models describe the agent’s view of the world and use this to describe his behavior. Decision-theoretic models start with behavior and ask what model would generate this. Given a model that generates the ‘right’ kind of behavior, we can predict correctly even if the model is not literally true.

[5] See Fagin et al. (1995) (Chapter 9).

## 3. Decision-theoretic models

First, we explain the standard Savage (1954) model and how it rules out unforeseen contingencies. As before, we have a set of states of the world, $\Omega$. Again, a state is a complete description of a relevant situation, where we define 'complete' below. We also have a set of *consequences* $X$. A consequence is a complete description of the outcome of a choice and determines how the agent 'feels about' the outcome. The options available to the agent are *acts*, functions from $\Omega$ to $X$. The set of all such functions is denoted $F$. The agent is assumed to have a preference relation $\succeq$ on $F$. Implicitly, then, the agent is assumed to view each possible action in terms of how its consequence varies with the state. This is the meaning of state $\omega$ being 'complete': for any action, the agent knows exactly the consequence of that action in state $\omega$.

In this setup, Savage derives a subjective expected-utility representation of the preferences. Specifically, he shows that given any preference $\succeq$ satisfying his axioms, there is a function $u: X \to \mathbf{R}$ and a probability measure $\mu$ on $\Omega$ such that[6]

$$f \succeq g \text{ iff} \sum_{\omega \in \Omega} \mu(\omega) u(f(\omega)) \geqslant \sum_{\omega \in \Omega} \mu(\omega) u(g(\omega)).$$

We state this well–known result to make clear the nature of the exercise: the agent's preferences (behavior) are used to derive a representation of his 'view of the world'.

Savage was troubled by the complete state-space assumption and proposed the notion of *small worlds*, incomplete descriptions of the world which the agent uses as if they were complete. The idea, slightly modified, is to treat the agent's state space as a partition of the true space, so that the states of the world the agent perceives are subsets of the real state space. We let $S$ denote the agent's perceived state space and $\Omega$ the true state space.

To see the idea, suppose there are two sources of uncertainty: whether it rains and whether there is a revolution in Haiti. This gives four possible states:

$$\Omega = \{(\text{rain, revolution}), (\text{rain, no revolution}),$$
$$(\text{no rain, revolution}), (\text{no rain, no revolution})\}$$

Consider an agent who is unaffected by or has not considered the possibility of revolution. He considers only whether or not it rains, so $S = \{(\text{rain}), (\text{no rain})\}$. The 'state' (rain) corresponds to the event $\{(\text{rain, revolution}), (\text{rain, no revolution})\}$ and (no rain) is analogous. Thus, the state space as seen by the

[6] We have written this for the finite state space case for simplicity. To be precise, though, Savage's axioms require an infinite state space. See, for example, Gul (1992) for axioms for the finite case.

agent is a partition of the true state space. The variation within an event of this partition is either irrelevant or unforeseen by the agent.

How would we model decisions of an agent in this context? Suppose the agent can use his money to buy an umbrella, invest it in Haiti, or save it. We use the following numerical representation of the consequences as a function of the real state. This is a description of objective reality – the agent's perceptions come next.

| State | Consequence if Buy | Consequence if Save | Consequence if Invest |
|---|---|---|---|
| (Rain, revolution) | 5 | 0 | – 100 |
| (Rain, no revolution) | 5 | 0 | 10 |
| (No rain, revolution) | 6 | 8 | – 100 |
| (No rain, no revolution) | 6 | 8 | 10 |

Recall that the agent sees the possibilities as rain versus no rain. The consequences of buying the umbrella or saving money are clear since they only depend on this. In this sense, these acts are measurable with respect to his awareness. Hence,

| State | Perceived consequence if Buy | Perceived consequence if Save | Perceived consequence if Invest |
|---|---|---|---|
| (Rain) | 5 | 0 | $x$ |
| (No rain) | 6 | 8 | $x'$ |

where we need to determine $x$ and $x'$.[7] Many recent papers have dealt with this in one way or another. Some of them give ways to answer the question 'what should $x$ and $x'$ be?', while others reformulate the problem. All, however, can be seen as addressing the basic problem of how to model the agent's perception of this kind of 'nonmeasurable' act.[8]

Before turning to these, we note Fishburn's (1970) perspective on the point. In Savage, states and consequences are given and acts are constructed from them. Fishburn suggested thinking of the acts and consequences as given and the states as constructed. In the context of our example, we would take buy, invest, and save as the given acts and the real numbers for the set of consequences. Then for each possible payoff function – each specification of a consequence for each

[7] Savage's proposal, that $x$ and $x'$ be acts, was not motivated by unforeseen contingencies and does not seem particularly appropriate for this problem. Hence, we do not discuss his views on this further.

[8] This problem is not special to decision-theoretic models. Adding decision-making to an epistemic model will also require an answer to this question.

act – we would have a state. That is, every vector of three numbers (giving the payoffs to the three actions in some order) would be a state. By construction, this state space is complete. This state space is a natural description of how a 'fully rational' person should make choices when she is aware that her knowledge of the true state space is incomplete. Such an individual does not care about the 'real' states per se, only about how well she does, how she feels as a consequence of her choice. That is, the payoff relevant contingencies are 'how does each choice make me feel?'

However, we cannot appeal to Savage's theorem with this state space. Savage assumes the agent has preferences over all functions from states to consequences. If we begin with our three actions, the state space we constructed would be the set of all three–tuples of real numbers. Hence, to apply Savage requires the agent to rank rank all functions from $\mathbf{R}^3$ to $\mathbf{R}$ – a large set to analyze for the choice between three elements! More to the point, most of these functions would be entirely artificial. We could never observe an agent's choices among such options and so could not test the theory, even in principle. If we could carry out a similar construction but in a way which makes a connection to observable choices, this objection would be eliminated. As we will see, some recent papers take this approach.

Modica et al. (1997) give a simple approach to nonmeasurable acts. In terms of the example, they suggest that if the agent only considers rain or no rain, he is implicitly assigning a default truth value to revolution or no revolution. The simplest case would be that the agent implicitly assumes there will not be a revolution and so $x = x' = 10$. While this approach seems quite natural, it has one drawback: in some settings, it is equivalent to standard expected utility where the agent puts zero probability on 'unforeseen' states. As shown by Modica, Rustichini, and Tallon, this equivalence does not always hold – for example, it breaks down with 'large' penalties for bankrupcy. On the other hand, it appears difficult to use this approach to study incomplete contracts.

Taking a different approach, Ghirardato (1996) suggests that acts should be correspondences, not functions.[9] That is, an act will be a function from $S$ to $2^X \backslash \{\emptyset\}$. For brevity, let $Y = 2^X \backslash \{\emptyset\}$. The interpretation is that $f(s)$ is the set of consequences possible in 'state' $s$. In our example, we might assume the agent views the acts as

| State | Consequences if Buy | Consequences if Save | Consequences if Invest |
|---|---|---|---|
| (Rain) | $\{5\}$ | $\{0\}$ | $\{-100, 10\}$ |
| (No rain) | $\{6\}$ | $\{8\}$ | $\{-100, 10\}$ |

[9] In independent and contemporaneous work, Mukerji (1997) gives a mathematically similar treatment, but with a different interpretation.

Ghirardato first generalizes the Savage axioms to correspondences.[10] Clearly, this gives a utility function $u: Y \to \mathbf{R}$ and a probability measure $\mu$ on $S$ such that

$$f \succeq g \text{ iff } \sum_{s\in S} \mu(s)u(f(s)) \geqslant \sum_{s\in S} \mu(s)u(g(s)).$$

Recall that $f(s)$ and $g(s)$ are subsets of $X$, not points in $X$. To write the utility of a set in terms of the utility of the points in it, Ghirardato adds an axiom which says that the agent prefers the best consequence in $y$ to the set $y$ and prefers the set $y$ to the worst consequence in $y$. This implies that there is a function $\alpha: Y \to [0,1]$ such that

$$u(y) = \alpha(y) \min_{x\in y} u(x) + [1-\alpha(y)] \max_{x\in y} u(x)$$

(where we abuse notation by writing $u(x)$ instead of $u(\{x\})$). Hence, his representation has act $f$ evaluated by

$$\sum_{s\in S} \mu(s)\left[\alpha(f(s)) \min_{x\in f(s)} u(x) + [1-\alpha(f(s))] \max_{x\in f(s)} u(x)\right],$$

where $\alpha(f(s))$ is part of the representation, just as the probabilities and utility function are. He suggests that $\alpha$ be interpreted as a measure of pessimism/optimism.

Ghirardato's representation generalizes nonadditive probability models. To see this, suppose we set $\alpha(y) = 1$ for all $y$. In this case, the agent's evaluation of acts is given by an expected value (with additive probabilities) of a minimum. Gilboa and Schmeidler (1994) have shown that nonadditive probability models can always be restated in such terms. This connection may seem surprising. After all, the standard axiomatizations for nonadditive probabilities are based on relaxing the sure-thing principle, while Ghirardato uses the standard sure-thing principle adapted to correspondences. More to the point, why should the sure-thing principle not apply in the presence of unforeseen contingencies? In our example, if an agent is deciding between two acts which are affected in exactly the same way by the 'state' (rain), why should not he make his choice entirely on the basis of comparing the acts in the 'state' (no rain)?

This connection comes from the fact that probabilities are placed on *sets* of consequences, not individual consequences. These probabilities are 'translated'

[10] Ghirardato does *not* require the 'accuracy' of perception we use in the example, nor does he require the agent to rank *all* possible acts (only a natural subset of them).

into probabilities on individual consequences, but not in a way which makes the distribution over consequences additive. Intuitively, the agent does *not* think 'rationally' about sets of consequences in the sense that if (rain) yields a set of consequences $y$ for some act, the agent does not assign probabilities directly to the individual consequences in $y$. As a result, the effective probabilities on consequences may not be additive. One could interpret Ghirardato as completing the state space in the manner suggested by Fishburn (1970) but in such a way that the probability distribution over the completed space is nonadditive.

Ghirardato's approach has several nice features, but also one important gap (as noted by Ghirardato): it takes the subjective representation of acts as given. Traditionally, one begins with an *objective* description of the agent's environment and derives a *subjective* perception of this environment. In Savage, states and consequences are interpreted as objective in the sense that an outside observer could construct these sets. Hence, the outside observer can construct the acts and observe the agent's preferences over acts by witnessing the agent's choices among these options. The utilities and probability beliefs are the only subjective variables and Savage's theorem tells how these are to be derived from observing the agent's preferences. In Ghirardato's approach, acts are exogenously given correspondences, even though we interpret them as the agent's subjective view of his actions. The crucial unanswered question, then, is how we can identify the agent's view from his preferences. Without an answer to this question, it is not clear whether the theory has any testable implications. It is only fair to note that one could make a similar critique of Savage in that he also assumes that the representation of objects in the world as acts describes the agent's perceptions of these objects. Of course, this assumption is precisely what the work on unforeseen contingencies attempts to relax.

Skiadis (1997) gives one approach to answering this question. In terms of our example, he derives a subjective representation of the $x$ and $x'$ above. That is, the numbers we plug in for $x$ and $x'$ become part of what we derive as our representation of the agent's preferences, just as a utility function is part of a representation of preferences. Roughly, Skiadis' view is that consequences are another word for utilities and thus are not observable either. In this sense, he endorses the extension of our critique of Ghirardato to a critique of Savage. Skiadis derives a subjective notion of what the act 'is' – that is, a subjective notion of how the act's payoffs vary with the state. He does *not* assume the existence of a set of consequences. Instead, 'acts' are points in a set, $F$, with no structure whatsoever.

To get more out of the model, one must put more in. He considers preferences over (act, event) pairs instead of acts. That is, instead of $f \succsim g$, we write $(f, E) \succsim (g, E')$ where $E$ and $E'$ are subsets of $S$. The interpretation is that the agent prefers choosing $f$ when the event $E$ occurs to choosing $g$ when $E'$

occurs.[11] Skiadis' axioms yield a representation where the payoff to $(f, E)$ is

$$u(f, E) = \sum_{s \in E} \frac{\mu(s)}{\mu(E)} u(f,\{s\}) \tag{1}$$

for some probability distribution $\mu$ on $S$. In effect, $u(f, E)$ is the conditional expected utility of $f$ given event $E$ where $u(f, \{s\})$ is the utility of the 'consequence' of choosing $f$ in state $s$. So we have derived 'consequences' from the preferences.

For intuition, let us return to the example with our three actions, buy, save, and invest. As noted, in Skiadis' treatment, these are simply points in a set. The representation in effect must give values for the unknowns $a$ through $f$ in the following table:

| State | Utility if Buy | Utility if Save | Utility if Invest |
|---|---|---|---|
| (Rain) | $a$ | $b$ | $c$ |
| (No rain) | $d$ | $e$ | $f$ |

as well as a probability for the state (rain). The preference order over act–event pairs tells us how to order $a$ through $f$ directly. Hence, we only need to be able to choose the right magnitudes to make Eq. (1) hold.

Note that this avoids assuming what should be derived in Ghirardato. On the other hand, the solution is incomplete. To see why, compare the table above to our original table giving the true acts. Intuitively, $c$, for example, can be thought of as the agent's subjective 'expectation' of the utility of 'invest' given rain – that is, the expectation over the possibilities of revolution versus no revolution. That is, Skiadis does not give us the 'completed' state space that Fishburn suggests, but only certain aggregated variables from it. Ghirardato took as exogenous the numbers the agent saw as possible values for this utility; Skiadis derives a subjective expectation over this set of values. Ghirardato gives more information about the act, while Skiadis provides a derivation for this subjective quantity. Ideally, though, we would want a subjective derivation of the set Ghirardato postulates.

Why does having the set of numbers instead of the expectation over the set matter? Clearly, there is more information in the former than the latter. Furthermore, in many applications, this additional information is critical. For example, consider the problem of long-term versus short-term contracts. As noted earlier,

[11] At this point, we oversimplify Skiadis' work in two ways. First, he does not literally consider a preference over act-event pairs. However, as he explains, his approach is essentially equivalent to this. Second, a large part of Skiadis' contribution is to consider preferences which are not 'separable' across states and his interpretation of the preference relation is more subtle than we make it sound in order to allow this. Since our interest lies elsewhere, we ignore this aspect of Skiadis' work.

the obvious advantage of short-term contracts is that it is easy to foresee contingencies for the near future than the distant future. Thus the agent would be 'more certain' about a given expectation if it concerns the near future. To model this seems to require getting at a more detailed representation of what the agent thinks is 'inside' these events and lies behind these 'expected utilities'.

Kreps (1979, 1992) gives an alternative approach which derives a subjective representation of this set of numbers.[12] That is, Kreps uses the agent's preferences to derive a subjective representation of what the rest of the state-space looks like, tying Fishburn's construction to observables. Kreps obtains this additional information about the agent by considering her preferences for flexibility. As Kreps puts it,

> What behavior ... do we look at to discern whether the individual does indeed anticipate that unforeseen things may occur? How can we recognize behavior that reflects a self-aware inability to predict all future contingencies? The answer ... is that we ask the individual how much flexibility he wishes to preserve in the future. That is, knowing that unforeseen events may happen, the individual is unwilling today to precommit to all actions he will later wish to take, even contingent on the contingencies he can foresee. He will wish instead to preserve for himself some ability to adapt to those events ....

Suppose the agent writes a contract giving her action as a function of the state $s$. She knows these states are not complete and so correspond to events in the full state space $\Omega$. Typically, she will prefer not to pin down her choice precisely if she believes she will learn more about $\Omega$ in the future. For simplicity, suppose that in the future, she will learn the exact true state, $\omega$. Then her ability to choose in the future is strictly better than her ability to choose today: now she can only choose an action as a function of an event in $\Omega$, getting only approximately what she wants, while tomorrow, she can get *exactly* what she wants as a function of the *exact* state. There would be no gain if the agent thought $S = \Omega$, so this reveals her recognition of unforeseen contingencies. Further, the agent's preferences over various constraints on future choice reveal the utility possibilities she perceives. For example, if she is indifferent between constraining herself to choose from the set $x$ or from $x$ plus one more option, she she must believe the additional option cannot be useful. That is, she believes there is no state $\omega \in \Omega$ in which this option is better than everything in $x$.

Formally, we have a finite set of basic options $B$. A contract is a function $f: S \to 2^{\mathbf{B}} \setminus \{\emptyset\}$, where in state $s$, the agent must choose from the set $f(s)$. The agent has preferences $\succeq$ over the set of such contracts.

Given certain axioms, Kreps shows that there is a *subjective state space* $\Theta$ and state–dependent utility functions $U : B \times \Theta \times S \to \mathbf{R}$ such that contract $f$ is

[12] Koopmans (1964) also discussed a model like this, motivated in part by unforeseen contingencies.

evaluated by

$$V(f) = \sum_{s \in S} \sum_{\theta \in \Theta} \max_{b \in f(s)} U(b, \theta, s).$$

To interpret this, think of $\Theta \times S$ as the agent's 'estimate' of the true state space $\Omega$. That is, the agent knows that any $s$ is really an event in $\Omega$, so he creates subjective versions of these states by considering $(\theta, s)$ for various $\theta$'s. The agent doesn't know which $(\theta, s)$ will obtain, but will find out at the beginning of the next period. At this point, she will, of course, choose the best item from $f(s)$, what the contract allows in state $(\theta, s)$. To get 'expected' payoffs, then, we sum over $(\theta, s)$. This can be viewed as an aggregation procedure or as an expectation with respect to a uniform probability distribution.

This representation gives subjective versions of many of the exogenous objects in the Savage model. Savage begins with a state space $\Omega$ and a consequence space $X$ and treats the options as functions from $\Omega$ to $X$. Here, as in Skiadis, we began with the options ($B$) as simply points in a set. As in both Ghirardato and Skiadis, instead of the true state space, we have a coarsening of it, $S$. Kreps' representation derives an extension of the state space to $\Theta \times S$ and a utility function $U(b, \theta, s)$ where, as in Skiadis, we can think of $\sum_{\theta \in \Theta} U(b, \theta, s)$ as the utility of the consequence that $b$ yields in 'state' $s$.

This solves the problem of providing a subjective derivation of the set of utility consequences of a given act in a given $s$. On the other hand, the identification is not tight. For simplicity, we assume in the sequel that $S$ is a singleton so a contract, henceforth termed a *menu*, is simply a nonempty subset of $B$.

Consider the following example. Let $B = \{c, d, e\}$ and assume the agent only cares about the number of items on the menu, strictly preferring longer menus to shorter ones. It is easy to check that one Krepsian representation has $\Theta = \{\theta_1, \theta_2, \theta_3\}$ where $U(b, \theta)$ is given by

$$\begin{array}{cccc} & \theta_1 & \theta_2 & \theta_3 \\ c & 1 & 0 & 0 \\ d & 0 & 1 & 0 \\ e & 0 & 0 & 1 \end{array}.$$

It is easy to see that another representation of these preferences has state space $\Theta' = \{\theta'_1, \theta'_2, \theta'_3\}$ where

$$\begin{array}{cccc} & \theta'_1 & \theta'_2 & \theta'_3 \\ c & 2 & 1 & 0 \\ d & 0 & 2 & 1 \\ e & 1 & 0 & 2 \end{array}.$$

Furthermore, if we take the union of these two state spaces, we get a third representation of the same preferences. Hence the state space (the set of rankings of $c$, $d$, and $e$ the agent considers possible) is not identified.

Why is this a problem? To apply this model, we need to know how assumptions on the structure relate to assumptions on the preferences. By analogy, in applying Savage, we know that if we assume $u$ is concave, we are assuming the agent is risk averse. We know that assuming that $u'(100) < 10$ is meaningless since an affine transformation of $u$ could flip this inequality but cannot affect preferences. We can not make analogous statements for the Kreps model. For example, we might think that the larger is the agent's $\Theta$, the more he is concerned about unforeseen contingencies. However, the example showed that we could have two representations *of the same preferences* where one $\Theta$ is a strict subset of the other. Unless we can pin down $\Theta$, we cannot determine the relationship between it and the agent's preferences and hence cannot easily apply the model.

Two papers address this problem by incorporating lotteries into the Kreps structure. Nehring (1996) considers lotteries over menus[13] – that is, the agent chooses a probability distribution giving a menu as a 'prize'. He shows there is a unique *additive dichotomous representation*, one where a lottery over menus is evaluated by the expectation over the lottery of $\sum_{\theta \in \Theta} \max_{b \in m} U(b, \theta)\mu(\theta)$, where $U(b, \theta)$ equals 0 or 1 for all $b$ and $\theta$.

Dekel et al. (1997b) consider menus of lotteries – that is, the menus can contain lotteries over $B$ – and focus on *additive EU representations*. Let $\beta$ be a typical lottery over $B$ where $\beta(b)$ is the probability the lottery gives outcome $b$. An additive EU representation has menu $m$ evaluated by $\sum_{\theta \in \Theta} \max_{\beta \in m} U(\beta, \theta)$, where for all $\theta$, $U(\cdot, \theta)$ is an expected – utility function – that is, it satisfies

$$U(\beta, \theta) = \sum_{b \in B} \beta(b) U(b, \theta).$$

They show that[14] there is a unique state space for such representations. Also, if one considers any *additive representation* – that is, drops the restriction that each $U(\cdot, \theta)$ be an expected-utility function – the expected-utility case generates the smallest possible state space. Finally, their axioms imply that every representation is additive.[15]

[13] More precisely, he considers Savage acts which give menus as consequences. However, he uses the standard approach of deriving a subjective probability distribution and replacing these acts with subjective lotteries. Hence, for our purposes, it is as if he studied lotteries over menus directly.

[14] Subject to caveats regarding infinite state spaces – see the paper for details.

[15] Nehring alludes to a similar result for his formulation.

## 4. Incomplete contracts

We now seem to have a model of unforeseen contingencies which can potentially be used to study contracts. In fact, the model looks very similar to the 'observable but unverifiable' uncertainty story used by Grossman and Hart (1986), Hart and Moore (1988), and Hart (1995) to model incomplete contracts. For brevity, we henceforth refer to this as the *GHM approach*. This approach assumes that some of the variables which are relevant to the contracting parties are observed by them but cannot be 'shown' to a court. As a result, these papers suggest, the parties cannot contract on these variables because a dispute about their realizations cannot be settled by the court.[16] Hence, these papers conclude, contracts can only allocate control rights; that is, assign the rights to make various decisions ex post. Intuitively, if some variables cannot be contracted on, one has to rely on the parties to choose appropriately in the relevant contingencies.

Similarly, Kreps assumes the agent contracts over the states as he understands them ($S$) and generates a noncontractable part of the state space ($\Theta$). This suggests interpreting this aspect of the GHM approach as generated by unforeseen contingencies.[17] In this sense, the Kreps approach has already been used to study incomplete contracts under another interpretation. This is an important conclusion since changing the interpretation of the GHM approach can change the assumptions which seem appropriate, opening up new avenues for exploration. In short, we seem to have a model of unforeseen contingencies which can deliver interesting results for contract theory; it is essentially the standard Savage model with restrictions on contracts that follow from the state space being endogenous.

Recent work by Maskin and Tirole (1997) calls this conclusion into doubt. They show that observable but unverifiable variables[18] do not justify a departure from standard contract theory. More specifically, even with observable but unverifiable variables, a mechanism can be designed which will induce the parties to reveal to the court what the values of the unverifiable variables are. Thus the fact that they cannot *prove* these facts to the court is not a problem since the court knows they will tell the truth.

Given the similarity to GHM, this suggests that the Kreps approach may not yield results different from standard contract theory either. In principle, there are ways one could introduce more realism to the Maskin–Tirole framework

[16] These papers often also use the assumption that certain actions are undescribable, but this aspect of the GHM approach is not relevant for our purposes.

[17] We emphasize that a more formal derivation along these lines would require extending the model above in several directions and it is far from clear that such an extension is possible.

[18] And undescribable actions.

and overturn their conclusion. For example, they suggest that bounded rationality may imply that the agents do not understand and so cannot use the complex mechanisms needed to enforce truth telling. On the other hand, it seems surprising that considerations other than unforeseen contingencies would be needed to generate something different from standard contract theory.[19]

In short, the literature has made significant progress in developing a model of unforeseen contingencies. But very significant problems remain to be solved before we can generate interesting conclusions for contracting from such models.

First, while the Kreps model (and its modifications) seems appropriate for unforeseen contingencies, the model is one where the preferences are state dependent and hence there are no meaningful subjective probabilities. A refinement of the model that pins down probabilities would be useful. Also, a multi-person extension of the model has not yet been developed. Both of these extensions are necessary to complete the connection between Kreps' decision-theoretic model and the GHM approach to incomplete contracts.

In a more speculative vein, one reason for the difficulty in developing interesting applications may be that the models to date are very close to the standard Savage model. Part of this is because the models have assumed that the agent is 'perfectly rational' except for unforeseen contingencies, a natural starting point. On the other hand, it may be important to weaken rationality assumptions further to obtain a more useful model.

We also suspect that part of the difficulty is that, while it is clear how the Savage model assumes away unforeseen contingencies, it is not clear what about unforeseen contingencies is not adequately captured by the modifications of Savage. To draw an analogy, prior to Savage (1954), known and unknown probabilities were seen as conceptually different and likely to lead to different behavior. Savage (1954) called this into question by showing that under his axioms, an agent facing unknown probabilities would act as if she assessed subjective probabilities and used these like known probabilities. Later, Ellsberg (1961) gave examples of intuitively plausible behavior which is inconsistent with subjective probabilities. This showed what an alternative representation needed to do.

Similarly, the approaches discussed above give variations on the standard model, providing the analog to Savage. However, there is not yet an 'Ellsbergian' example for unforeseen contingencies, a thought experiment which would clearly show what is wrong with the (modified) Savage approach to modeling unforeseen contingencies. We suspect that this lacunae makes it difficult to

[19] One assumption of Maskin and Tirole's does seem inappropriate for unforeseen contingencies. They assume *dynamic programming*, roughly a 'rational expectations' assumption which says that agents correctly anticipate the utility consequences of what happens in the future. This assumption is also made in the GHM approach, but does not seem reasonable in the Krepsian reinterpretation. On the other hand, dropping this assumption seems unlikely to have interesting consequences.

identify a clear direction for improving on simple adaptations of Savage and hence a useful alternative approach.

One stumbling block to getting such an example is that we want to distinguish between unforeseen contingencies and 'standard' uncertainty aversion. By the latter, we mean those models, such as nonadditive probability, which are intended to represent an agent who knows the state space but not the appropriate probabilities and behaves 'conservatively' because of this lack of knowledge. Conceptually, at least, there is a difference between this and not knowing the state space and behaving conservatively as a result. However, this distinction is difficult to make precise. Intuitively, we want a problem in which it is 'difficult' for the agent to translate the options into utilities and where the agent exhibits some aversion to this difficulty. But how can we distinguish this from translating the options easily into utility space but then being uncertainty averse on the utility space? Perhaps the lesson here is that there is not a formalizable distinction. For example, one natural alternative to the Krepsian representation would have a menu $m$ evaluated by $\min_{\theta\in\Theta} \max_{b\in m} U(b, \theta)$, a functional form which clearly embodies a stronger notion of 'unforeseen contingency aversion'. This, of course, looks very similar to the well-known sets of measures approach to uncertainty.

An alternative is to take a 'procedural' approach. Rather than show that preferences are represented by a subjective state space, perhaps we should try to understand how the agent 'constructs' a state space and the probabilities. This might give us a better understanding of the connection between the subjective state space and the real world, facilitating extensions to multiple agents and obtaining meaningful probabilities. This would also address an issue which arises even without unforeseen contingencies; namely, where do subjective probabilities come from?

Overall, we feel that the implications of unforeseen contingencies on an otherwise rational decision maker are now (mostly) as well understood as the Savage model. Some troubling questions in the Savage model – such as how the agent determines probabilities – are even more disturbing here, though. Also, it is not clear that the resulting model can yield interesting applications in contract theory. A procedural approach might be more useful and might enhance our understanding of decision making even without unforeseen contingencies. Another approach is to find an 'Ellsbergian paradox' that indicates behaviors that are due to unforeseen contingencies and that conflict with standard subjective (non)expected utility models, thus helping focus the search for alternatives.

## Acknowledgements

We thank Drew Fudenberg and Paolo Ghirardato for helpful comments. Dekel and Rustichini thank the NSF and Lipman thanks SSHRCC for financial support for this research. Lipman also thanks Carnegie Mellon for their hospitality while this work was in progress.

## References


Dekel, E., Gul, F., 1997. Rationality and knowledge in game theory. In: Kreps, D.M., Wallis, K.F. (Eds.), Advances in Economic Theory. 7th World Congress, Cambridge University Press, Cambridge.

Dekel, E., Lipman, B.L., Rustichini, A., 1997a. Standard state-space models preclude unawareness. Econometrica, forthcoming.

Dekel, E., Lipman, B.L., Rustichini, A., 1997b. A unique subjective state space for unforeseen contingencies. Working paper.

Ellsberg, D., 1961. Risk, ambiguity, and the savage axioms. Quarterly Journal of Economics 75, 643–669.

Fagin, R., Halpern, J., 1988. Belief, awareness and limited reasoning. Artificial Intelligence 34, 39–76.

Fagin, R., Halpern, J., Moses, Y., Vardi, M., 1995. Reasoning About Knowledge. MIT Press, Cambridge.

Fishburn, P.C., 1970. Utility Theory for Decision Making; Publications in Operations Research, No. 18, Wiley, New York.

Geanakoplos, J., 1989. Game theory without partitions, and applications to speculation and consensus. Working paper, Yale University.

Ghirardato, P., 1996. Coping with ignorance: unforeseen contingencies and non-additive uncertainty. Mimeo, Cal Tech.

Gilboa, I., Schmeidler, D., 1994. Additive representations of non-additive measures and the choquet integral. Annals of Operations Research 52, 43–65.

Grossman, S., Hart, O., 1986. The costs and benefits of ownership: a theory of vertical and lateral integration. Journal of Political Economy 94, 691–719.

Gul, F., 1992. Savage's theorem with a finite number of states. Journal of Economic Theory 57, 99–110.

Hart, O., 1995. Firms, Contracts, and Financial Structure. Clarendon Press, Oxford.

Hart, O., Moore, J., 1988. Incomplete contracts and renegotiation. Econometrica 56, 755–786.

Hayek, F., 1960. The Constitution of Liberty. University of Chicago Press, Chicago.

Koopmans, T., 1964. On the flexibility of future preferences. In: Shelly, M.W., Bryan, G.L. (Eds.), Human Judgements and Optimality. Wiley, New York.

Kreps, D., 1979. A representation theorem for preference for flexibility. Econometrica 47, 565–576.

Kreps, D., 1992. Static choice and unforeseen contingencies. In: Dasgupta, P., Gale, D., Hart, O., Maskin, E. (Eds.), Economic Analysis of Markets and Games: Essays in Honor of Frank Hahn. MIT Press, Cambridge, pp. 259–281.

Maskin, E., Tirole, J., 1997. Dynamic programming, unforeseen contingencies, and incomplete contracts. Mimeo, Harvard University.

Modica, S., Rustichini, A., 1994. Awareness and partitional information structures. Theory and Decision 37, 107–124.

Modica, S., Rustichini, A., 1993. Unawareness: a formal theory of unforeseen contingencies: Part II. CORE Discussion Paper.

Modica, S., Rustichini, A., Tallon, J.-M., 1997. A model of general equilibrium with unforeseen contingencies. Economic Theory, forthcoming.

Mukerji, S., 1997. Understanding the nonadditive probability decision model. Economic Theory 9, 23–46.

Nehring, K., 1996. Preference for flexibility and freedom of choice in a Savage framework. Mimeo, Cal Davis.
Pires, C., 1994. Three essays in economic theory. Ph.D. Dissertation, MIT Press, Cambridge.
Samet, D., 1990. Ignoring ignorance and agreeing to disagree. Journal of Economic Theory 52, 190–207.
Savage, L.J., 1954. The Foundations of Statistics. Wiley, New York.
Shin, H.S., 1993. Logical structure of common knowledge. Journal of Economic Theory 60, 1–13.
Skiadis, C., 1997. Subjective probability under additive aggregation of conditional preferences, Journal of Economic Theory, forthcoming.