When is Non-Probabilistic Robustness a Good Probabilistic Bet?

Yakov Ben-Haim
Yitzhak Moda'i Chair in Technology and Economics
Technion — Israel Institute of Technology
Haifa 32000 Israel
yakov@technion.ac.il, +972-4-829-3262

Contents

Part I
Main Body of the Paper

Abstract

Concepts of robustness are often employed when decisions under uncertainty are made without probabilistic information. We present a theorem which establishes necessary and sufficient conditions for non-probabilistic robustness to be equivalent to probability of success. When this "proxy property" holds, probability of success is enhanced (or maximized) by enhancing (or maximizing) robustness.

[0]\papers\Proxy Thms\prx27a.tex, extended version of prx28.tex 27.4.2011.

[0]The author is pleased to acknowledge valuable comments and suggestions by Q. Farooq Akram, Lior Davidovitch, Maria Demertzis, L. Joe Moffitt, Rygnar Nymoen, Ross Pinsky, John K. Stranlund and Miriam Zacksenhouse.

Two further theorems establish important special cases. The proxy property has implications for survival advantage, even when the agent is unaware of the proxy property as in the case of animals. Applications to foraging, finance, forecasting, monetary policy formulation, principal-agent contracts, Bayesian model mixing, and Ellsberg's paradox of behavior under ambiguity are discussed.

Keywords. Satisficing, robustness, info-gap theory, probability of survival, bounded rationality, Knightian uncertainty, Ellsberg's paradox, equity premium puzzle.

1 Introduction

Robustness to severe uncertainty is often evaluated without probabilistic models, for instance when uncertainty is specified by a set of possibilities with no measure function on that set. When is a robust decision likely to succeed? What can we say about the probability of success of a decision when we have no probabilistic information about the uncertainties? If robustness is used as a criterion to select a decision, when is this criterion equivalent to selection according to the probability of successful outcome? In short, when is (non-probabilistic) robustness a good (probabilistic) bet?

We present three propositions, based on info-gap decision theory, which identify conditions in which the probability of success is maximized by an agent who robustly satisfices the outcome without using probabilistic information. We show that this strategy may differ from the outcome-optimizing strategy indicated by the best available data and models. We will refer to these propositions as "proxy theorems" since they establish conditions in which robustness is a proxy for probability. An info-gap robust-satisficing strategy attempts to attain an adequate or necessary (but not necessarily extremal) outcome while maximizing the agent's immunity to deficient information. The robust-satisficing approach requires no knowledge of probability distributions. These propositions provide insight into the prevalence of decision-making behavior which is inconsistent with outcome-optimization based on the best-available (but faulty) models and data. Best-model strategies are vulnerable to error, and other—robust-satisficing—strategies will be shown to have higher probability for survival by the agent in commonly occurring situations. We will consider a number of examples, including risky investment, foraging, forecasting, monetary policy formulation, principal-agent contracts, Bayesian model mixing, and the Ellsberg paradox of decisions under ambiguity.

Our analysis is based on two fundamental concepts—satisficing and robustness—which we now discuss.

1.1 Satisficing

Models based on outcome-optimization do not always satisfactorily resolve or explain economic and ethological puzzles and paradoxes such as the equity premium puzzle (Mehra and Prescott, 1985; Kocherlakota, 1996), the home bias paradox (French and Poterba, 1990, 1991; Jeske, 2001), and foraging behavior of animals (Nonacs, 2001; Ward, 1992). Simon's concepts of satisficing—doing good enough but not optimizing—and bounded rationality underlie many attempts to understand recalcitrant decision-making behavior (Simon, 1955, p.101; 1956, p.128). Conlisk (1996, p.670) lists a jeremiad of human decision-making weaknesses which support the vast empirical evidence for bounded, rather than global, rationality.

Simon's concepts of satisficing and bounded rationality are motivated by the psychological and epistemic limitations of an agent in its interaction with a complex, variable and uncertain environment. If agents *could* optimize globally, then they would; but they can't (for epistemic and cognitive reasons), so they don't. The occurrence of satisficing strategies also has strong psychological roots (Kahneman, 2003). For example, satisficing has been attributed to the impact of emotional states, specifically, insufficient or excessive emotional arousal (Kaufman, 1999). Considerations of the context, process and necessity of deciding can motivate satisficing and can shift behavior away from optimizing directly on outcomes. This is important in many economic, political, social and managerial processes such as labor relations, voting, business ethics, and network design, and has a distinctive axiomatic foundation (Sen, 1997). Furthermore, there is considerable evidence that humans (and other organisms, such as firms (Simon, 1979)) actually use boundedly rational (rather than globally optimal) strategies. Crain *et al.* (1984) show that satisficing managers attain competitive profits for their firms and seem to command competitive personal compensation in the market for managers. Thaler (1994) argues that savings behavior of households depends on the psychology of bounded rationality. Gabaix *et al.* (2006) show that sub-optimal strategies are implemented both in simple decision problems, where fully optimal search is feasible, and in complex problems. Gutierrez *et al.* (1996, p.362) "search for network configurations that are 'good' for a variety of likely future scenarios" but not necessarily optimal for a specific scenario. Numerous scholars have demonstrated the efficacy of simple, satisficing heuristics in human and animal decision making (Gigerenzer and Selten, 2001).

Why is satisficing so prevalent? When is robust-satisficing a good bet?

It has sometimes been thought that evolution under competition would lead to optimal strategies and optimizing behavior. This is the interpretation sometimes given to the concept of "survival of the fittest", either in biological, social, or economic competition. However, there is much evidence that agents who have been highly honed by lengthy competitive selection display satisficing rather than outcome-optimizing behavior. Competitive selection is a strong mechanism for removing agents who fail to "survive" in some relevant sense, suggesting that satisficing has survival value. We address the questions: why, and when, and in what form, is a satisficing strategy a better bet for survival of the agent, than a strategy which uses the best available information in attempting to optimize the outcome.

1.2 Robustness

'Robustness' has many meanings. As we will use it, the concept of robustness derives from a prior concept of non-probabilistic uncertainty. Knight (1921) distinguished between 'risk' based on known probability distributions and 'true uncertainty' for which probability distributions are not known. Similarly, Ben-Tal and Nemirovski (1999) are concerned with uncertain data within a prescribed uncertainty set, without any probabilistic information. Likewise Hites *et al.* (2006, p.323) view "robustness as an aptitude to resist to 'approximations' or 'zones of ignorance' ", an attitude adopted also by Roy (2010). We also are concerned with robustness against Knightian uncertainty. We consider uncertainty in probability distributions but we do not pursue an explicitly statistical approach to robustness as studied by Huber (1981) and many others. The concepts developed here are related to the idea of probability bounds, and to the concept of coherent lower previsions, as discussed elsewhere (Ben-Haim *et al.* 2009).

Wald (1945) studied the problem of statistical hypothesis testing based on a random sample whose probability distribution is not known, but whose distribution is known to belong to a given class of distribution functions. Wald states that "in most of the applications not even the existence of ... an a priori probability distribution [on the class of distribution functions] ... can be postulated, and in those few cases where the existence of an a priori probability distribution ... may be assumed this distribution is usually unknown." (p.267). Wald introduced a loss function expressing the "relative importance of the error committed by accepting" one hypothesized subset of distributions when a specific distribution in fact is true. (p.266). He notes that "the determination of the [loss function] is not a statistical question and is considered here as given." (p.266). Wald developed a decision procedure which "minimizes the maximum ... of the risk function." (p.267).

Many engineering researchers, beginning in the 1960s, developed estimation and control algorithms for linear dynamic systems based on sets of inputs. Schweppe (1973) for instance develops inference and decision rules based on assuming that the uncertain phenomenon can be quantified in such a way as to be bounded by an ellipsoid, with no probability function involved.

Hansen and Sargent have pioneered the introduction of robustness tools in economics. In their recent book (2008) they quantify model misspecification by taking "a given approximating model and surrounding it with a set of unknown possible data generating processes, one unknown element of which is the true process .... Our decision maker confronts model misspecification by seeking a decision rule that will work well across a set of models for which" the relative entropy is bounded. "The decision maker wants a single decision rule that is reliable for *all* [emphasis in original] models ... in the set" (p.11). They explain that "'Reliable' means good enough, but not necessary optimal, for each member of a set of models." (footnote 21, p.11). They then "maximize [an] intertemporal objective over decision rules when a hypothetical malevolent nature minimizes that same objective .... That is, we use a max-min decision rule." (p.12).

The concept of robustness in this paper is in the tradition of ideas which we have described.

The proxy theorems presented later establish conditions under which robustness is a proxy for probability of success or survival. That is, by enlarging or maximizing the robustness one also enlarges or maximizes the probability of satisfying survival requirements. This is important when probability distributions are not known. One is not able to *evaluate* the probability of success, but one is able to *enlarge* or *maximize* it by using a robustness function. The behavioral implication is that agents who robust-satisfice will tend to succeed (or survive) more than agents who optimize with respect to their best models. Robust-satisficers (and their strategies) will tend to dominate in uncertain competitive evolution. This has implications for learning and adaptation (of which the agent may be unaware) under uncertain competition.

Section 2 describes the proxy theorems intuitively. Section 3 is a précis of info-gap theory. Section 4 presents proposition 1 and three examples. Proposition 1 establishes two conditions which are necessary and sufficient for the proxy property to hold. The examples show that these requirements are satisfied in diverse and important situations. Sections 5 and 6 present two special cases of proposition 1. All proofs and derivations are presented in the Appendices.

2 Preliminary Discussion of the Proxy Theorems

Before embarking on the technical details, we describe the essence of the proxy theorems and their significance.

The agent chooses an action r which results in outcome $G(r, q)$ where q is an uncertain parameter, vector, function or a set of such entities. The discussion in this section assumes that $G(r, q)$ is a loss, but, with minor modifications, it applies to rewards as well. We will refer to $G(r, q)$ as the 'performance function'. For instance, q might be an uncertain estimate of a critical parameter such as a rate of return, or q could be a vector of uncertain returns, or q could be a probability density function (pdf) for uncertain returns, or q could be a set of such pdf's, or q could be uncertain constitutive relations such as supply and demand curves.

The agent's knowledge and beliefs about q are represented by a family of sets, $\mathcal{Q}(h)$, called an info-gap model of uncertainty, where h is a non-negative real number. $\mathcal{Q}(h)$ is a set of values of q, so if q is a vector, function, or set, then $\mathcal{Q}(h)$ is a set of vectors, functions or sets. As h increases, the range of possibilities grows, so $h' < h$ implies $\mathcal{Q}(h') \subseteq \mathcal{Q}(h)$. This is the property of *nesting,* and it endows h with its meaning as an *horizon of uncertainty.* All info-gap models have the property of nesting. Sometimes the agent may have a specific estimate of q, denoted $\widetilde{q}$. In this case, in the absence of uncertainty (that is, $h = 0$), $\widetilde{q}$ is the only possibility so $\mathcal{Q}(0) = \{\widetilde{q}\}$. This is the property of *contraction,* which is common among info-gap models (Ben-Haim, 2006), though our proxy theorems will not depend on the contraction property. An info-gap model is a quantification of Knightian uncertainty (Knight, 1921; Ben-Haim, 2006, sections 11.5.6 and 13.5) and is consistent with all of the set-models of uncertainty discussed in section 1.2.

The agent "survives" if the loss does not exceed a critical value G_c. The intention here is the same as Hansen and Sargent's concept of a reliable decision (2008) mentioned in section 1.2: good enough but not necessarily optimal. The robustness of action r is the greatest horizon of uncertainty, h, up to which $G(r, q) \leq G_c$ for all $q \in \mathcal{Q}(h, \widetilde{q})$. Denote the robustness by $\widehat{h}(r, G_c)$. More robustness is preferable to less robustness, so the robustness generates a preference ranking of the actions (denoted $\succ_r$), namely, $r \succ_r r'$ if $\widehat{h}(r, G_c) > \widehat{h}(r', G_c)$.

Given a probability distribution for q, called $P(q)$, we could compute the probability that the agent survives. Let $\Lambda(r, G_c)$ denote the set of all q's for which $G(r, q) \le G_c$. The probability of survival is $P[\Lambda(r, G_c)]$. This probability generates a ranking of preferences on the actions (denoted $\succ_p$), namely, $r \succ_p r'$ if $P[\Lambda(r, G_c)] > P[\Lambda(r', G_c)]$. Note that the set $\Lambda(r, G_c)$ differs from the class of sets $\mathcal{Q}(h)$ of the info-gap model.

Our proxy theorems establish conditions in which $\succ_r$ and $\succ_p$ are equivalent. Let us identify three central results.

1. A change in action which enhances the robustness need not enhance the probability of survival. That is, $\succ_r$ and $\succ_p$ are not necessarily equivalent. This is demonstrated in Appendix A. Davidovitch (2009) has shown that very strict conditions are needed in order to prove a proxy theorem.
2. A proxy theorem asserts that $\succ_r$ and $\succ_p$ *are* equivalent under certain conditions, as specified by the propositions to follow. The main contribution of this paper is to establish conditions which are strict enough to enable a proxy theorem and loose enough to encompass a wide range of important decision problems.
3. Satisficing is more robust than optimizing. That is, it is more robust to try to guarantee a larger loss than a smaller loss, and minimizing the loss has zero robustness. This is a rigorous and well known trade-off theorem (Ben-Haim, 2006), presented as eqs.(4) and (5) in section 3.

To understand the importance of items 2 and 3 let us consider a prototypical example: foraging. The net loss in energy by a foraging animal must not exceed some critical level, or the animal will perish before the next foraging session. (Analogies to net loss of resources by a business firm are obvious.) The long-term probability of survival of the animal depends on the probability that the critical loss is not exceeded. Item 3, the trade-off theorem, shows that satisficing the loss at the critical value is never less robust (and usually more robust) than minimizing the loss. Item 2, a proxy theorem, establishes conditions in which enhancing the robustness also enhances the probability of survival, and maximizing the robustness maximizes the probability of survival. In other words, when a proxy theorem holds, robust-satisficing strategies may lead to evolutionary success (though this paper does not deal with evolutionary processes per se).

We will show in this paper that proxy theorems hold for a very wide range of economic decisions. This contributes to an understanding of the success and prevalence of robust decision strategies as discussed by many authors in the economic literature.

This example and many others illustrate the importance of simple heuristics in bounded rationality, as stressed by Gigerenzer and Selten (2001). Simple heuristics can be extraordinarily robust to uncertainty since they depend on only very limited information. A proxy theorem, which links robustness to probability of survival, shows why simple heuristics have high survival value.

3 Info-Gap Robust-Satisficing: A Précis

This paper employs info-gap decision theory (Ben-Haim, 2006), which has been applied in a large array of decision problems under severe uncertainty in engineering (Ben-Haim, 2005), biological conservation (Burgman, 2005), economics (Ben-Haim, 2010) and other areas (see http://info-gap.com).

The agent must make a decision by choosing a value for r, which may be a scalar, vector, function, or linguistic variable such as "go" or "no-go". The outcome of the decision is expressed as a loss (or reward), quantified by a scalar performance function $G(r, q)$, which depends on the decision r and on an uncertain quantity q. q is an uncertain parameter, vector, function, or a set of such entities. The uncertainty in q is represented by an info-gap model, whose two properties—contraction and nesting—were defined in section 2.

We now define the robustness function and the robust-satisficing decision strategy, discuss three basic properties, define the probability of survival, and discuss the relation between min-max and robust-satisficing.

Robustness function: definition. By "survival" we mean that the loss or penalty, $G(r,q)$, from decision r is acceptably small, less than a critical value G_c. We will consider numerous examples in subsequent sections. The loss $G(r,q)$ may itself be a probabilistic entity such as a mean or a quantile, and q may be an uncertain probability distribution. Since we don't know the true value of q we cannot evaluate the loss. However, we can evaluate a decision, r, in terms of the range of q-values for which the loss is acceptable.

To quantify this, define the robustness function (Ben-Haim, 2006):[1]

$$\widehat{h}(r, G_c) \equiv \max \left\{ h : \left(\max_{q \in \mathcal{Q}(h)} G(r,q) \right) \leq G_c \right\} \tag{1}$$

We define $\widehat{h}(r, G_c) \equiv 0$ if the set of h's in eq.(1) is empty. We can "read" this equation from left to right as follows. The robustness, $\widehat{h}$, of decision r with aspiration for loss no greater than G_c, is the maximal horizon of uncertainty h up to which all realizations of $q \in \mathcal{Q}(h)$ result in loss $G(r,q)$ no greater than G_c.

If $G(r,q)$ is a reward rather than a loss then the inner 'max' in eq.(1) becomes 'min' and the '$\leq$' becomes '$\geq$'. The formulation of some examples is more natural by defining the performance function as a reward rather than a loss, and we will do this on occasion. However, without loss of generality, all of our propositions are formulated for the definition of robustness in eq.(1). Gains can always be treated as losses by considering the negation of the performance-reward function.

The robustness function—whether for reward or loss—generates preferences on the decision, $\succ_r$, defined in section 2 as:

$$r \succ_r r' \quad \text{if} \quad \widehat{h}(r, G_c) > \widehat{h}(r', G_c) \tag{2}$$

The robust-satisficing decision, at aspiration G_c, maximizes the robustness:

$$\widehat{r}(G_c) \equiv \arg\max_r \widehat{h}(r, G_c) \tag{3}$$

Robustness function: three properties. We now briefly discuss three properties of the info-gap robustness function (Ben-Haim, 2006) which will illuminate the significance of the proxy theorems proven later.

Robustness trades-off against performance. Satisficing the loss at a lower (better) loss entails lower (worse) robustness:[2]

$$G_c < G'_c \quad \Longrightarrow \quad \widehat{h}(r, G_c) \leq \widehat{h}(r, G'_c) \tag{4}$$

This relation is an immediate consequence of the nesting of the sets of an info-gap model (see section 2).

Best-model outcomes have no robustness to uncertainty.[3] Satisficing the loss at the anticipated value,[4] $G(r,\widetilde{q})$, entails zero robustness:

$$G_c = G(r, \widetilde{q}) \quad \Longrightarrow \quad \widehat{h}(r, G_c) = 0 \tag{5}$$

This is true for any decision, r, so it is true for the best-model outcome-optimal decision which minimizes $G(r,\widetilde{q})$.

Robustness curves can cross one another, as illustrated in fig. 1. The anticipated loss, $G(r_1,\widetilde{q})$, from decision r_1 is lower than the anticipated loss, $G(r_2,\widetilde{q})$, from r_2. Likewise, at low loss and low

[1]Throughout the paper we will use 'min' and 'max' operators since, in practice, the sets in question are almost invariably closed. When open sets are involved our intention is to 'inf' and 'sup' operators.

[2]The first inequality is reversed if we consider reward rather than loss. The meaning is retained: robustness decreases if greater reward is required.

[3]This depends on a 'non-satiation' property: that the loss can always get worse as the uncertainty increases. See Ben-Haim, 2005, section 6.1.

[4]Note that $G(r,\widetilde{q})$ is quite general. If q is an uncertain vector or function, then $G(r,\widetilde{q})$ is the loss based on the best estimate of q. Or, if q is a pdf, then $G(r,\widetilde{q})$ can be a best estimate of a mean, or a quantile, of the loss. If q is a set of pdf's, then $G(r,\widetilde{q})$ can be the best estimate of the worst-case mean or quantile of the loss.

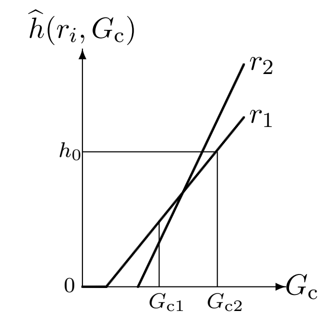

Figure 1: Crossing robustness curves, and illustration of modeller's equivalence and decision maker's preference between min-max and robust-satisficing.

robustness, r_1 is more robust and thus preferred over r_2 (according to $\succ_{\rm r}$ in eq.(2)). However, at higher loss and higher robustness, r_2 is more robust and thus preferred over r_1. Crossing of robustness curves entails the reversal of preference. The preference relation $\succ_{\rm r}$ in eq.(2) depends on the acceptable loss, G_c, if the robustness curves cross one another.

Crossing of robustness curves also implies that best-model outcome-optimization may differ from robust-satisficing. The decision which minimizes the loss based on the best available information is:

$$r^\star \equiv \arg\min_r G(r, \widetilde{q}) \tag{6}$$

$r^\star$ may differ from the robust-satisficing decision, $\widehat{r}(G_c)$ in eq.(3), if their robustness curves cross, depending on the value of G_c. We will encounter many examples of crossing robustness curves in this paper.

Probability of survival. Now consider the probability of survival for decision r, namely, the probability that q will take a value so that $G(r, q) \le G_c$. Let $p(q|r)$ denote the pdf for q, noting that it may depend on the decision r. We do not know this pdf, and q itself may be a probability density or a set of functions, so $p(q|r)$ could be quite complicated. Nonetheless, we can define the probability of survival as:

$$P_{\rm s}(r, G_c) \equiv \text{Prob}[G(r, q) \le G_c] = \int_{G(r,q)\le G_c} p(q|r)\,{\rm d}q \tag{7}$$

We cannot evaluate $P_{\rm s}(r, G_c)$ because the pdf $p(q|r)$ is unknown, but if we did know it, then it would generate preferences $\succ_{\rm p}$ over decisions, defined in section 2 as:

$$r \succ_{\rm p} r' \quad \text{if} \quad P_{\rm s}(r, G_c) > P_{\rm s}(r', G_c) \tag{8}$$

Min-max and robust-satisficing. The robust-satisficing decision strategy is very closely related to min-max decision making as studied by many authors, including Wald (1945) and Hansen and Sargent (2008) discussed in section 1.2, and many others. In fact, we will now explain how a robust-satisficing decision can be *represented* as a min-max decision, and how min-maxing can be represented as robust-satisficing. We will call this the *modeller's equivalence* between min-max and robust-satisficing. However, we will also show that min-max and robust-satisficing are not necessarily equivalent from the decision maker's point of view. We will call this the *decision maker's preference* between min-max and robust-satisficing. Our discussion will be brief and intuitive.

A min-max decision is one which ameliorates a worst case at a specified horizon of uncertainty, h_0. Using our notation, a min-max decision is:

$$r_{\rm m}(h_0) = \arg\min_r \max_{q\in\mathcal{Q}(h_0)} G(r, q) \tag{9}$$

The Hansen-Sargent "evil agent" maximizes the loss, and the min-maxer minimizes this worst loss. (If the info-gap model $\mathcal{Q}(h_0)$ is the family of sets of pdf's with bounded relative entropy then the horizon of uncertainty h_0 is equivalent to the parameter η_0 in Hansen and Sargent (2008, p.11) .)

We now consider the min-max and robust-satisficing choices between two options, r_1 and r_2, whose robustness curves are shown in fig. 1. We will suppose that the decision maker's uncertainty is h_0.

Modeller's equivalence. Suppose that the robust-satisficing agent requires loss no greater than G_{c2} in order to survive. The robust-satisficing agent will choose r_2 since it is more robust than r_1 at critical loss G_{c2} as seen in fig. 1. The modeller can represent this choice as a min-max decision by supposing or discovering that the agent's horizon of uncertainty equals h_0. Clearly, min-max can always represent a robust-satisficing agent's choice by allowing the modeller to deduce or discover an appropriate horizon of uncertainty (which need not be unique).

The reverse equivalence also holds: agents who use the min-max strategy can always be represented as robust-satisficing decision makers by suitable choice (by the modeller) of a survival requirement G_c.

Decision-maker's preference. Let us suppose that the agent still identifies h_0 as the horizon of uncertainty, but also suppose that the agent requires loss no greater than G_{c1} in fig. 1 in order to survive. The agent recognizes that the min-max choice is r_2 because of the value of h_0. However, r_1 is more robust to uncertainty than r_2 at requirement G_{c1}, so the robust-satisficing agent would choose r_1 rather than the min-max choice r_2. This robust-satisficing choice is re-enforced when a proxy theorem holds, since then the probability that the agent will survive (probability that the loss will not exceed G_{c1}) is greater with r_1 than with r_2. When a proxy theorem holds, robust-satisficing agents will choose r_1 and will tend to survive under competition more than agents who choose the min-max choice r_2. When a proxy theorem holds, robust-satisficing is never a worse bet than min-maxing, and will sometimes be a better bet. Even though robust-satisficing and min-maxing agents agree about the horizon of uncertainty, they will disagree about the action to choose when G_{c1} and h_0 are positioned as in fig. 1. This is the basis of the decision maker's preference.

The decision maker's preference can also be understood in terms of the information which is available to the decision maker. The min-maxer relies on defining the horizon of uncertainty that is felt to be relevant (h_0 in fig. 1); the robust satisficer on the other hand defines the maximum loss (or minimum aspiration) that is acceptable. Even though there are occasions when the two lead to the same decision, as described above, the information used by the decision maker to come to this decision is very different. In what information does the decision maker have more confidence, the horizon of uncertainty (i.e. what is possible), or the satisficing requirement (what the agent needs, likes, etc), when facing severe Knightian uncertainty? An agent who answers the latter will be a robust satisficer rather than a min-maxer. This is the basis of the decision maker's preference.

4 Proxy Theorem: Monotonicity and Coherence

We can now proceed to the first result of this paper.

A critical question is: when do the preference rankings in eqs.(2) and (8) agree? We can evaluate the robustness function, $\widehat{h}(r, G_c)$, while we cannot evaluate $P_s(r, G_c)$, so if $\succ_r$ and $\succ_p$ agree then robustness is a proxy for the probability of survival. By choosing r to enlarge or maximize robustness we would also enlarge or maximize the probability of survival. We are not able to *evaluate* the probability of survival, but we would be able to *enlarge* or *maximize* it. The behavioral implication is that agents who robust-satisfice will tend to survive more than agents who use any other strategy, such as optimizing with respect to their best models. Robust-satisficers will tend to dominate in competitive evolution. If we can identify general conditions for the selective advantage of satisficing, then we can understand the prevalence of satisficing behavior under competition.

In section 4.1 we define a concept of coherence between an info-gap model and a probability distribution and in appendix B we present two simple examples. This concept of coherence is not to be confused with the one in de Finetti's theory of subjective probability. The concept of coherence underlies our central proxy theorem in section 4.2. We apply this proxy theorem to the allocation of resources among risky assets in section 4.3. We demonstrate the coherence between a specific info-gap model and the normal distribution of uncertain payoffs in the risky-asset example, thus demonstrating the relevance of the proxy theorem to this class of problems. In section 4.4 we demonstrate coherence and the proxy property for an example from monetary policy. In section 4.5 we demonstrate coherence and the proxy property for the principal-agent problem.

4.1 Coherence: Definition

We are considering performance functions $G(r, q)$ which are scalar and depend on the decision r and on q which is an uncertain parameter, vector, function or set. Without loss of generality we may consider $G(r, q)$ itself to be the uncertain entity, whose info-gap model is generated by the info-gap model for a more complex underlying uncertainty q. It is, however, more convenient to retain the distinction between $G(r, q)$ (the performance function) and q (the uncertainty) and to assume that $G(r, q)$ is monotonic in q which is a scalar. This includes the case that q is itself the performance function which depends on more complex underlying uncertainties. Numerous examples are discussed in sections 4.3–4.5, 5.2–5.6 and 6.3 which will illustrate the aggregation of complex multi-dimensional uncertainties.

In summary, q is an uncertain scalar variable, r is a decision variable, and $G(r, q)$ is a scalar performance function. An info-gap model for uncertainty in q is $\mathcal{Q}_r(h)$, which may depend on the decision, r. The corresponding robustness function, eq.(1), is $\widehat{h}(r, G_c)$. The cumulative probability distribution (cpd) of q is $P(q|r)$.

For any $h \geq 0$, define $q^\star(h, r)$ and $q_\star(h, r)$, respectively, as the least upper bound and greatest lower bound of q-values in the set $\mathcal{Q}_r(h)$. Define $\mu(h)$ as the inner maximum in the definition of the robustness in eq.(1):

$$q^\star(h, r) \equiv \max_{q \in \mathcal{Q}_r(h)} q, \quad q_\star(h, r) \equiv \min_{q \in \mathcal{Q}_r(h)} q, \quad \mu(h) \equiv \max_{q \in \mathcal{Q}_r(h)} G(r, q) \tag{10}$$

We will consider performance functions $G(r, q)$ which are monotonic (though not necessarily strictly monotonic) in q at fixed r. We define the inverse of such functions, at fixed r, as follows. If $G(r, q)$ *increases* as q increases then its inverse is defined as:

$$G^{-1}(r, G_c) \equiv \max \{q : \ G(r, q) \leq G_c\} \tag{11}$$

If $G(r, q)$ *decreases* as q increases then its inverse is defined as:

$$G^{-1}(r, G_c) \equiv \min \{q : \ G(r, q) \leq G_c\} \tag{12}$$

$G(r, q)$ is assumed to be monotonic but we do not assume that $G(r, q)$ is continuous in q, which is why we need the inequalities rather than equality.

Definition 1 . *$\mathcal{Q}_r(h)$ and $P(q|r)$ are **upper coherent** at decisions r_1 and r_2 and critical value G_c, with performance function $G(r, q)$, if the following two relations hold for $i = 1$ or $i = 2$, and $j = 3 - i$:*

$$P[G^{-1}(r_i, G_c)|r_i] \quad > \quad P[G^{-1}(r_j, G_c)|r_j] \tag{13}$$

$$G^{-1}(r_i, G_c) - q^\star(h, r_i) \quad > \quad G^{-1}(r_j, G_c) - q^\star(h, r_j)$$
$$\text{for } h = \widehat{h}(r_j, G_c) \text{ and } h = \widehat{h}(r_i, G_c) \tag{14}$$

*$\mathcal{Q}_r(h)$ and $P(q|r)$ are **lower coherent** if eqs.(13) and (14) hold when $q^\star(h, r)$ is replaced by $q_\star(h, r)$.*

Roughly speaking, coherence implies some "information overlap" between the info-gap model, $\mathcal{Q}_r(h)$, and the probability distribution, $P(q|r)$. Eq.(13) depends on $P(q|r)$ but not on h or $\mathcal{Q}_r(h)$, while eq.(14) depends on h and $\mathcal{Q}_r(h)$ but not on $P(q|r)$. Both relations depend on G_c, r_i, r_j and the performance function $G(r, q)$. $\mathcal{Q}_r(h)$ and $P(q|r)$ are coherent if each of these relations holds.

Coherence does not imply that either function, $\mathcal{Q}_r(h)$ or $P(q|r)$, can be deduced from the other. Coherence does imply that knowledge of one function reveals something about the other.

If the cpd $P(q|r)$ does not depend on r then eq.(13) is equivalent to:

$$G^{-1}(r_i, G_c) > G^{-1}(r_j, G_c) \tag{15}$$

Likewise, if the info-gap model $\mathcal{Q}_r(h)$ does not depend on r then $q^\star(h,r)$ and $q_\star(h,r)$ also do not depend on r and eq.(14) is identical to eq.(15). In other words $P(q|r)$ and $\mathcal{Q}_r(h)$ are always upper and lower coherent if neither of them depends on the decision, r. The implications of this are explored in section 5.

Upper coherence becomes interesting if the uncertainty models, $P(q|r)$ and $\mathcal{Q}_r(h)$, *do* depend on the decision. Now eq.(13) does not imply eq.(15) because the cpd may change as r changes. However, if the info-gap model is coherent with the probability distribution then $q^\star(h,r)$ "compensates" for the change in the cpd and eq.(14) is the resulting "correction" of eq.(15).

Some further insight into the meaning of coherence, and two simple examples, are discussed in section B of the Appendix.

Proposition 1, to be presented shortly, asserts, roughly, that coherence is necessary and sufficient for the proxy property to hold. But how does an agent choose a coherent info-gap model without knowing the pdf? The answer derives from the adaptive survival implications of the proxy property. An agent who chooses an info-gap model which is coherent with the pdf has a survival advantage over an agent who chooses a non-coherent info-gap model because of the proxy property. This is true even if the agent was unaware of the coherence when choosing. The learning or adaptation which takes place—even if it is non-volitional as in animals—leads to the identification of coherent info-gap models.

4.2 Proposition 1

Definition 2 *$\mathcal{Q}_r(h)$ and $P(q|r)$ have the* ***proxy property*** *at decisions r_1 and r_2 and critical value G_c, with performance function $G(r,q)$, when:*

$$\widehat{h}(r_1, G_c) > \widehat{h}(r_2, G_c) \quad \textit{if and only if} \quad P_s(r_1, G_c) > P_s(r_2, G_c) \tag{16}$$

The proxy property is symmetric between robustness and probability of success. However, we are particularly interested in the implication from robustness to probability. Thus, when the proxy property holds we will sometimes say that robustness is a proxy for probability of success.

Nesting of the sets $\mathcal{Q}_r(h)$ implies that $q^\star(h,r)$ and $q_\star(h,r)$, defined in eq.(10), are monotonic increasing and decreasing functions, respectively. They are continuous if the following additional properties hold.

Definition 3 *An info-gap model, $\mathcal{Q}_r(h)$,* ***expands upward continuously*** *at h if, for any $\varepsilon > 0$, there is a $\delta > 0$ such that:*

$$|q^\star(h', r) - q^\star(h, r)| < \varepsilon \quad \textit{if} \quad |h' - h| < \delta \tag{17}$$

Continuous downward expansion *is defined similarly with $q_\star(\cdot)$ instead of $q^\star(\cdot)$.*

We can now state our first proposition, whose proof appears in section C of the Appendix.

Proposition 1 *Info-gap robustness to an uncertain scalar variable, with a loss function which is monotonic in the uncertain variable, is a proxy for probability of survival if and only if the info-gap model $\mathcal{Q}_r(h)$ and the probability distribution $P(q|r)$ are coherent.*

Given:

• *At any fixed decision r, the performance function, $G(r,q)$, is monotonic (though not necessarily strictly monotonic) in the scalar q.*

• $\mathcal{Q}_r(h)$ *is an info-gap model with the property of nesting.*
• r_1 *and* r_2 *are decisions with positive, finite robustnesses at critical value* G_c.
• $\mathcal{Q}_r(h)$ *is continuously upward (downward) expanding at* $\widehat{h}(r_1, G_c)$ *and at* $\widehat{h}(r_2, G_c)$ *if* $G(r, q)$ *increases (decreases) with increasing* q.

Then: *The* ***proxy property*** *holds for* $\mathcal{Q}_r(h)$ *and* $P(q|r)$ *at* r_1, r_2 *and* G_c *with performance function* $G(r, q)$.

If and only if: $\mathcal{Q}_r(h)$ *and* $P(q|r)$ *are* ***upper (lower) coherent*** *at* r_1, r_2 *and* G_c *with performance function* $G(r, q)$ *which increases (decreases) in* q.

This proposition establishes that coherence is both necessary and sufficient (together with some other conditions) for the proxy property to hold. The most important additional condition is that the performance function is monotonic in a single scalar uncertainty. In sections 4.3–4.5 we illustrate applications of this proposition to investment in risky assets, monetary policy, and the principal-agent problem, demonstrating both coherence and monotonicity. Coherence with a range of probability distributions is explored. We illustrate how the monotonicity requirement is satisfied by aggregating complex multi-dimensional uncertainties. Two special types of coherence are developed and illustrated in sections 5 and 6.

4.3 Example: Risky Assets

Formulation. Consider N risky assets in a 2-period investment. We will indicate the generalization to more than two periods later.

The investor purchases amount r_i of asset i in the first period, at price p_i; no purchases are made in the second period. In the second period, the payoff of asset i is $q_i = p_i + d_i$ where d_i is the uncertain dividend. The initial wealth is w and the consumptions in the two periods, c_1 and c_2, are:

$$c_1 = w - p^T r, \quad c_2 = q^T r \tag{18}$$

where superscript T implies matrix transposition.

The utility from consumption c_j is $u(c_j)$ which we assume to be strictly increasing in c_j: positive marginal utility. The discounted utility for the two periods is $u(c_1) + \beta u(c_2)$ where β is a positive discount factor. This is the "natural" reward function for this problem, but it is not consistent with our formal definitions and results which assume the performance function is a loss. We define the performance function as:

$$G(r, q) = -u(c_1) - \beta u(c_2) \tag{19}$$

Uncertainty and Robustness. The uncertainty derives from the unknown payoff vector of the risky assets in the 2nd period, q. We do not know a probability distribution for q and we cannot reliably evaluate moments. There are many types of info-gap models which could be used (Ben-Haim, 2006). We will consider a specific example subsequently.

Now note from eqs.(18)–(19) that the performance function, $G(r, q)$, depends on the uncertain payoffs only through the consumption in the second period, c_2, which is a scalar uncertainty. To emphasize that the performance function depends on the uncertain payoff vector only through c_2 we write $G(r, c_2)$. Note that $G(r, c_2)$ decreases monotonically in the scalar uncertainty c_2, thus satisfying the monotonicity requirement of proposition 1. In this way the N-dimensional uncertain vector, q, is "aggregated" into a single scalar uncertainty, c_2.

Whatever info-gap model is adopted for q, denoted $\mathcal{Q}(h)$, an info-gap model for c_2 is:

$$\mathcal{C}_r(h) = \left\{ c_2 : \ c_2 = r^T q, \ q \in \mathcal{Q}(h) \right\}, \quad h \ge 0 \tag{20}$$

The investor prefers less negative utility $G(r, c_2)$ rather than more, and G_c is the greatest value of discounted 2-period negative utility which is acceptable. If G_c cannot be attained (or reasonably anticipated) then the investment is rejected. G_c is a "reservation price" on the negative utility.

For given investments r, the robustness to uncertainty in the consumption in the second period, c_2, is the greatest horizon of uncertainty h up to which all realizations of c_2 result in discounted negative utility no more than G_c:

$$\widehat{h}(r, G_c) = \max \left\{ h : \left(\max_{c_2 \in \mathcal{C}_r(h)} G(r, c_2) \right) \leq G_c \right\} \tag{21}$$

More robustness is preferable to less, at the same level G_c at which the negative utility is satisficed.

The conditions of proposition 1 hold if the info-gap model, $\mathcal{C}_r(h)$, and the pdf of c_2 are coherent. (An example of coherence is developed in section D of the Appendix, showing coherence between an infinity of info-gap models and the normal and other similarly standardizable distributions.) When coherence holds, any change in the investment, r, which augments the robustness also augments (or at least does not reduce) the probability that the performance requirement, $G \leq G_c$, will be satisfied. The probability of success can be maximized by maximizing the robustness, without knowing the probability distribution of the vector of returns on the risky asset.

Because of the proxy property, coherence of an agent's info-gap model is a re-enforcing attribute: the survival value is greater for coherent than for non-coherent models. This suggests the possibility of an evolutionary process by which coherent info-gap models are selected (though the agent may be unaware of this selection process). This process could work because very simple info-gap models can be coherent with the corresponding pdf even though their information-content is much less than the pdf itself. For example, Gigerenzer and Selten (2001) have demonstrated the efficacy of simple, satisficing heuristics in human and animal decision making.

Many periods. If there are more than two periods then uncertain payoffs occur in intermediate periods as well as in the last period. Consequently the above "aggregation" of the uncertain payoff vector q into the scalar consumption of the second (that is, last) period does not work if there are more than two periods. In that case, however, we can aggregate the utilities of all periods after the first. Define:

$$g = \sum_{i=2}^{K} \beta^{i-1} u(c_i) \tag{22}$$

where K is the number of periods. Let q denote the concatenation of the uncertain payoff vectors in all periods, with info-gap model $Q(h)$. We then replace $\mathcal{C}_r(h)$ in eq.(21) by:

$$\mathcal{G}_r(h) = \left\{ g : \; g = \sum_{i=2}^{K} \beta^{i-1} u(c_i), \; q \in \mathcal{Q}(h) \right\}, \quad h \geq 0 \tag{23}$$

The performance function is $G(r, q) = -u(c_1) - g$, which is monotonic in the scalar uncertainty g. The robustness in eq.(21) is now defined as:

$$\widehat{h}(r, G_c) = \max \left\{ h : \left(\max_{g \in \mathcal{G}_r(h)} G(r, q) \right) \leq G_c \right\} \tag{24}$$

4.4 Example: Monetary Policy with Uncertain Expectations

In this section we consider a simple monetary policy analysis in which public expectations about inflation and output are uncertain to the central bank which must choose an interest rate to keep the inflation from rising excessively. We will show that a natural info-gap model for the uncertain expectations is upper coherent with a wide range of probability distributions. Furthermore the various uncertainties can be aggregated so that the performance function is monotonic in a single scalar uncertainty. This means that the conditions of proposition 1 are satisfied, so that a robust-satisficing strategy for choosing the interest rate is a proxy for the probability of success of the outcome.

Macro model. We use a simple model based on Clarida, Galí and Gertler (1999) to represent the bank's approximate understanding of the economy:

$$\pi_{t+1} = \lambda y_t + \beta \mathrm{E}_t \pi_{t+1}, \quad y_{t+1} = -(r_t - \mathrm{E}_t \pi_{t+1})\phi + \mathrm{E}_t y_{t+1} \tag{25}$$

ϕ, λ and β are positive parameters. π_t is the inflation in period t defined as the percent change in the price level from $t-1$ to t. y_t is the output gap, defined as 100 times the difference between the actual and potential output, both expressed in logs, after removal of the long-run trend. The decision variable for the central bank is r_t, the nominal interest rate. Both π_t and r_t are evaluated after removal of the long-run trend. E_t is the expectation operator for the representative agent based on information available at time t: $\pi_t, \pi_{t-1}, \ldots$ and $y_t, y_{t-1}, \ldots$. We concentrate on the average behavior under uncertainty in the expectations, and ignore zero-mean shocks. (For inclusion of model uncertainty and shocks whose distribution is uncertain see Ben-Haim (2010).)

The central bank announces the credible intention to target inflation and output gap at the values π_{m} and y_{m}, respectively. The public's expectations are formed to converge on these targets:

$$\mathrm{E}_t \pi_{t+1} = \pi_t - \psi_\pi(\pi_t - \pi_{\mathrm{m}}), \quad \mathrm{E}_t y_{t+1} = y_t - \psi_y(y_t - y_{\mathrm{m}}) \tag{26}$$

Uncertain expectations. The central bank is highly uncertain about the values of the feedback coefficients, ψ_π and ψ_y. The bank has estimates, $\widetilde{\psi}_\pi$ and $\widetilde{\psi}_y$, with approximate errors s_π and s_y, but these are not thought to be accurate. A fractional-error info-gap model for the uncertain coefficients is:

$$\mathcal{U}(h) = \left\{ (\psi_\pi, \psi_y) : \left| \frac{\psi_\pi - \widetilde{\psi}_\pi}{s_\pi} \right| \le h, \ \left| \frac{\psi_y - \widetilde{\psi}_y}{s_y} \right| \le h \right\}, \quad h \ge 0 \tag{27}$$

The horizon of uncertainty, h, is not known, so this is an unbounded family of nested sets of ψ_π and ψ_y values.

Performance requirement. The nominal interest rate chosen at time t, r_t, influences the inflation only at time $t+2$. However, π_{t+2} depends on π_{t+1} which is unknown at time t so we estimate π_{t+1} with $\mathrm{E}_t\pi_{t+1}$. Likewise y_{t+1} is estimated by its expectation, $\mathrm{E}_t y_{t+1}$. After some algebra one obtains the following expression for π_{t+2} which we adopt as the performance function:

$$G(r_t, q) = \pi_{t+2} \tag{28}$$

$$= -\lambda\phi r_t + (\lambda\phi + \beta)\pi_t + \lambda y_t \underbrace{-(y_t - y_{\mathrm{m}})\lambda\psi_y + (\lambda\phi - 2\beta)(\pi_t - \pi_{\mathrm{m}})\psi_\pi + (\pi_t - \pi_{\mathrm{m}})\beta\psi_\pi^2}_{q} \tag{29}$$

which defines the uncertain scalar q. The performance function increases as q increases, thus satisfying the monotonicity requirement of proposition 1. Note that q "aggregates" the two underlying uncertainties, ψ_π and ψ_y.

The info-gap model for ψ_π and ψ_y, eq.(27), induces the following info-gap model for q:

$$\mathcal{Q}(h) = \left\{ q : \ q = -(y_t - y_{\mathrm{m}})\lambda\psi_y + (\lambda\phi - 2\beta)(\pi_t - \pi_{\mathrm{m}})\psi_\pi + (\pi_t - \pi_{\mathrm{m}})\beta\psi_\pi^2, \ (\psi_\pi, \psi_y) \in \mathcal{U}(h) \right\}, \quad h \ge 0 \tag{30}$$

We note that neither q nor its info-gap model depend on the central bank's decision, r_t.

If inflation and output gap were on target ($\pi_t = \pi_{\mathrm{m}}$ and $y_t = y_{\mathrm{m}}$) then expectations would be stable (eqs.(26) become $\mathrm{E}_t\pi_{t+1} = \pi_{\mathrm{m}}$ and $\mathrm{E}_t y_{t+1} = y_{\mathrm{m}}$). The inflation would increase as:

$$\pi_{t+2} = -\lambda\phi r_t + (\lambda\phi + \beta)\pi_t + \lambda y_t \tag{31}$$

The central bank would choose the interest rate r_t to maintain this happy situation.

However, suppose that the current inflation and output gap, π_t and y_t, are below their target values. $\widetilde{\psi}_\pi$ and $\widetilde{\psi}_y$ are positive so public expectations are thought to indicate rising inflation and output gap, recalling that the values of $\widetilde{\psi}_\pi$ and $\widetilde{\psi}_y$ are highly uncertain. The central bank wishes to choose the current interest to keep the inflation from rising excessively. That is, the *performance requirement* is to keep the inflation below a critical value:

$$\pi_{t+2} \le \pi_{\rm c} \tag{32}$$

It is desirable that inflation will increase to some extent in order to prevent undue output disturbance, but the bank would like to limit this to values no greater than the stable no-intervention situation. Thus, from eq.(31), $\pi_{\rm c}$ will be chosen in the range:

$$\pi_{\rm c} < (\lambda\phi + \beta)\pi_t + \lambda y_t \tag{33}$$

The robustness analysis of this and related problems is studied in Ben-Haim (2010).

In section E of the Appendix we establish that the two conditions for upper coherence, eqs.(13) and (14) in definition 1, hold for a wide range of probability distributions. Hence proposition 1 holds and robustness of the choice of interest rate is a proxy for the probability of satisfying the requirement on the inflation.

4.5 Example: Principal-Agent Problem

The essential challenge of the principal-agent relation derives from the different information which is available to the two parties. The principal wants to design a contract which the agent will accept, but which will satisfy the principal's needs as well. This depends on the agent's attributes, such as skill or effort as well as the agent's utility function, all of which are better known to the agent than to the principal.

There are two separate, though inter-related, issues here: will the agent accept the contract, and will the outcome of the agent's subsequent actions satisfy the principal. We will demonstrate an info-gap analysis of the first issue, showing that the proxy theorem, proposition 1, holds for a wide range of circumstances. A similar analysis can be developed for the second issue. Our formulation of the principal-agent problem is motivated by Stiglitz (1975, 1998).

Notation and formulation. The states of the world are denoted by $i = 1, \ldots, N$. The probability of the ith state is $p_i(e)$ which is influenced by the agent's attributes e which we will refer to as 'effort'. The contract offered by the principal grants reward r_i to the agent when state i prevails, so the vector r of rewards is the decision to be made by the principal. The utility to the agent from state i is $u_i(r_i, e)$, and the utility to the principal from state i is $v_i(r_i)$. The agent's and principal's expected utilities from contract r are:

$$U(r, u, p) = \sum_{i=1}^{N} u_i(r_i, e) p_i(e), \quad V(r, p) = \sum_{i=1}^{N} v_i(r_i) p_i(e) \tag{34}$$

The classical problem statement has the principal choose r to maximize V and satisfy the agent's reservation constraint $U \ge U_{\rm c}$.

Uncertainty. The principal has an estimate $\widetilde{p}$ of the vector of probabilities based on the principal's estimate of the agent's effort. $\widetilde{p}$ is a normalized probability distribution. However, the actual probabilities are uncertain due to uncertainty in the effort and perhaps other unknown factors influencing the state of the world. The following info-gap model expresses the fact that the principal simply does not know the magnitude of error of $\widetilde{p}$:

$$\mathcal{P}(h) = \left\{ p: \ |p_i - \widetilde{p}_i| \le h, \ p_i \ge 0, \ \forall i, \ \sum_{i=1}^{N} p_i = 1 \right\}, \quad h \ge 0 \tag{35}$$

Each element of $\widetilde{p}$ errs up to an unknown amount, h, subject to the constraints of non-negativity and normalization. The horizon of uncertainty, h, is unbounded.

The principal has an estimate of the agent's vector of utility functions, $\widetilde{u}(r)$, based on the principal's estimate of the agent's effort. However, the principal does not know the extent to which this estimate is accurate. This is expressed by the following info-gap model:

$$\mathcal{A}_r(h) = \{u(r) = \widetilde{u}(r) + \eta_i : \ |\eta_i| \le h \ \forall i\}, \quad h \ge 0 \tag{36}$$

Robustness. We now formulate the info-gap robustness to uncertainty regarding acceptance of the contract by the agent. To be consistent with the loss-formulation of robustness, eq.(1), we define the performance function as the negative utility to the agent:

$$G(r, u, p) = -U(r, u, p) \tag{37}$$

The performance requirement is that $G \le G_c$ where $G_c = -U_c$. The robustness of contract r can now be formulated, as in eq.(1), as:

$$\widehat{h}(r, G_c) \equiv \max \left\{ h : \ \left(\max_{\substack{u \in \mathcal{A}_r(h) \\ p \in \mathcal{P}(h)}} G(r, u, p) \right) \le G_c \right\} \tag{38}$$

An explicit expression for the robustness is derived in appendix F.3.

Proposition 1 requires that the performance function be monotonic in a single scalar uncertainty. In section F.1 of the Appendix we demonstrate how this is achieved by aggregating the uncertain vectors.

In section F.2 of the Appendix we establish that the two conditions for upper coherence, eqs.(13) and (14), hold for a wide range of probability distributions. Our discussion is similar to that in section E of the Appendix.

The other conditions of proposition 1 also hold, in particular the condition of monotonicity. Hence robustness is a proxy for the probability of satisfying the agent's requirement. A robust-satisficing choice of the contract, by the principal, will maximize the probability that the agent's reservation condition will be satisfied. The principal can maximize the probability that the contract will be accepted by the agent, without explicitly knowing the probability distribution involved. The principal will not know the value of the probability of acceptance, but it will be known that no other contract has greater acceptance probability.

We have demonstrated, in sections 4.3–4.5, info-gap models which are coherent with a range of diverse probability distributions for disparate and important economic situations. However, it is clear in these examples that there are many probability distributions with which the info-gap models are *not* coherent. This motivates the issue of learning and adaptation which will be discussed briefly in section 7.

5 Proxy Theorem: Monotonicity and Independence

5.1 Proposition 2

A particularly important and commonly occurring situation is that the info-gap model, $\mathcal{Q}(h)$, and the probability distribution, $P(q)$, are both independent of the decision, r. We will examine a number of examples later. In discussing eq.(15) following definition 1 in section 4 we noted that the property of coherence holds if $\mathcal{Q}(h)$ and $P(q)$ are both independent of r. Using proposition 1, this allows us to immediately assert the following proposition.

Proposition 2 *Info-gap robustness to an uncertain scalar variable, with a loss function which is monotonic in the uncertain variable, is a proxy for probability of survival if the info-gap model $\mathcal{Q}(h)$ and the probability distribution $P(q)$ are both independent of the decision r.*

Given:

• *At any fixed decision r, the performance function, $G(r,q)$, is monotonic (though not necessarily strictly monotonic) in the scalar q.*

• *$\mathcal{Q}(h)$ is an info-gap model with the property of nesting.*

• *r_1 and r_2 are decisions with positive, finite robustnesses at critical value G_c.*

• *$\mathcal{Q}(h)$ is continuously upward (downward) expanding at $\widehat{h}(r_1, G_c)$ and at $\widehat{h}(r_2, G_c)$ if $G(r,q)$ increases (decreases) with increasing q.*

• *$\mathcal{Q}(h)$ and $P(q)$ are both independent of the decision r.*

Then: *The proxy property holds for $\mathcal{Q}(h)$ and $P(q)$ at r_1, r_2 and G_c with performance function $G(r,q)$.*

We now consider a series of examples which illustrate the application of proposition 2.

5.2 Example: Foraging

Foraging is an essential activity for all animals, and has been the focus of extensive theoretical and field study. Attempts to explain foraging decisions such as allocation of time between foraging sites have generally had limited success (Nonacs, 2001). In particular, theoretical explanations based on the assumption that animals attempt to maximize the intake of energy have often been less than satisfactory. In this section we illustrate a very simple foraging model which obeys the conditions of proposition 2. This is a somewhat simpler model than studied elsewhere (Carmel and Ben-Haim, 2005).

The basic idea is that an animal needs a specific minimal quantity of energy in order to survive. Garnering more energy might be nice, but it is not necessary for survival. A robust-satisficing foraging strategy will attempt to maximize the robustness to uncertainty in attaining this critical quantity. When a proxy theorem holds—as it does in the example developed here—this strategy will be more likely than any other to achieve the survival requirement. This means that robust-satisficing strategies will tend to have an evolutionary advantage, and will tend to persist and prevail in competition against other strategies. It must be remembered, however, that the present example is much simpler than real-life foraging.

The foraging model. Consider an animal who has duration T remaining in which to forage, say until nightfall, and must acquire at least G_c calories in that period in order to survive the night. The animal is currently foraging at location 0 at which the rate of energy acquisition is g_0 calories per hour. g_0 is known precisely and will remain constant for the duration of the foraging session. Another site, 1, is available at which the acquisition rate is estimated to be $\widetilde{g}_1$, which exceeds g_0. However, $\widetilde{g}_1$ is highly uncertain, with an error estimated at about s but the true value, g_1, is not known. How much longer, t, should the animal remain at site 0 before moving to site 1?

The performance function—total calories acquired—for remaining at site 0 for time t and then moving to site 1 for the remainder of the time is: $G(t) = tg_0 + (T-t)g_1$. The performance requirement is that $G(t)$ be no less than G_c: $G(t) \geq G_c$.

The uncertainty in the rate of energy collection at site 1 is represented by a fractional-error info-gap model:

$$\mathcal{Q}(h) = \left\{ g_1 : \left| \frac{g_1 - \widetilde{g}_1}{s} \right| \leq h \right\}, \quad h \geq 0 \tag{39}$$

That is, the fractional deviation of the true energy-accumulation rate g_1, from the estimated value $\widetilde{g}_1$, in units of the estimated error s, is bounded by the horizon of uncertainty h, but the value of h is not known. We assume that g_0, $\widetilde{g}_1$ and s are positive.

The robustness of remaining for duration t at site 0 is the greatest horizon of uncertainty, h, up to which the performance requirement is guaranteed:

$$\widehat{h}(t, G_c) = \max \left\{ h : \left(\min_{g_1 \in \mathcal{Q}(h)} G(t) \right) \geq G_c \right\} \tag{40}$$

One readily finds the following expression for the robustness of duration t:

$$\widehat{h}(t, G_{\rm c}) = \begin{cases} \dfrac{\widetilde{G}(t) - G_{\rm c}}{(T-t)s} & \text{if } \widetilde{G}(t) \ge G_{\rm c} \\ 0 & \text{else} \end{cases} \tag{41}$$

where $\widetilde{G}(t) = tg_0 + (T-t)\widetilde{g}_1$ is the estimated value of the performance function.

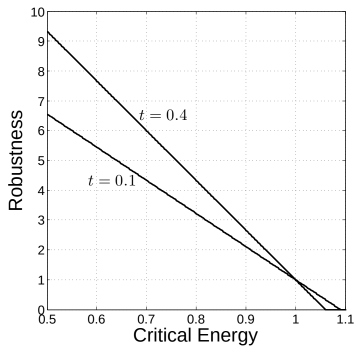

Figure 2: Foraging. Robustness curves, $\widehat{h}(t, G_{\rm c})$ vs. $G_{\rm c}$, for two t values. $T = 1$, $g_0 = 1$, $\widetilde{g}_1 = 1.1$, $s = 0.1$.

Discussion. It is readily shown that the robustness curves for different choices of t cross one another if and only if: $g_0 < \widetilde{g}_1$. This is illustrated in fig. 2. Furthermore, this relation implies that the estimated reward, $\widetilde{G}(t)$, increases as t decreases.

The energy-acquisition rate of site 0, g_0, is known with certainty, while the acquisition rate of site 1 is highly uncertain and estimated to equal $\widetilde{g}_1$. The relation '$g_0 < \widetilde{g}_1$' embodies a dilemma facing the forager (and many other decision makers). When this relation holds, site 0 is thought to be less productive but is known to be more reliable than site 1.

This dilemma is expressed by the intersection between robustness curves for different choices of t, as seen in fig. 2. If very high energy is required—large $G_{\rm c}$ is essential for survival—then small t is more robust than large t, meaning that more time should be spent at the risky but potentially more productive site. On the other hand, when low $G_{\rm c}$ is adequate then the less-productive but completely reliable site is allocated more foraging time. Specifically, when choosing between $t = 0.1$ and $t = 0.4$ (see fig. 2) the robust-satisficing forager will choose $t = 0.1$ if and only if the critical energy requirement, $G_{\rm c}$, exceeds the value at which the robustness curves cross one another, which is $G_{\rm c} = 1$. Incidentally, the value of $G_{\rm c}$ at which the curves cross is $G_\times = Tg_0$, independent of the value of t.

When a proxy theorem holds—as it does in this case, proposition 2—the robust-satisficing choice of t is the one with greater probability of achieving the critical energy requirement. If the animal's survival requirement for energy is less than $G_\times$ then the larger value, $t = 0.4$, is more robust and thus more likely to achieve the survival requirement than the smaller value, $t = 0.1$. This is true even though the best-model estimate of the accumulated energy, $\widetilde{G}(t)$, favors $t = 0.1$. Conversely, if $G_{\rm c} > G_\times$ then the smaller value of t is more robust and more likely to result in survival. Only in the latter case does the robust-satisficing solution agree with the putative best-model optimum ($t = 0.1$).

5.3 Example: Bayesian Model Mixing

A decision, r, must be made, whose outcome depends on the state of the world, which is either A or B. The outcome is a loss, which is $g(r, A)$ or $g(r, B)$, depending on the state of the world. That is, the decision maker has two models for describing the outcome. Uncertain contextual understanding suggests that the probability of state A is $\widetilde{q}$, and hence the probability of state B is $1 - \widetilde{q}$. However, $\widetilde{q}$ is highly uncertain; it is hunch, like "2 to 1 for model A".

The expected loss of decision r, if the true probability of state A is q, is:

$$G(r, q) = g(r, A)q + g(r, B)(1 - q) \tag{42}$$

It is required that the expected loss be no greater than the critical value G_c: $G(r, q) \le G_c$.

All we know about the probability of the state of the world is the estimated probability, $\widetilde{q}$, of state A and that this estimate is highly uncertain. The true probability, q, is thought to equal $\widetilde{q}$ but could be any value between zero and one. We will consider an asymmetric info-gap model in which the interval of q values expands around $\widetilde{q}$ and reaches the boundary values, 0 and 1, when the horizon of uncertainty equals unity:

$$\mathcal{Q}(h) = \{q : \ q \in [0, 1], \ (1-h)\widetilde{q} \le q \le \widetilde{q} + (1-\widetilde{q})h\}, \quad h \ge 0 \tag{43}$$

The robustness of decision r is the greatest horizon of uncertainty, h, up to which the performance requirement is guaranteed, eq.(1).

Define $\Delta(r) = g(r, A) - g(r, B)$. One can show that, if $G_c \ge G(r, \widetilde{q})$, the robustness function is:

$$\widehat{h}(r, G_c) = \begin{cases} \dfrac{G(r,\widetilde{q}) - G_c}{\widetilde{q}\Delta(r)} & \text{if } \Delta(r) < 0 \\ \dfrac{G_c - G(r,\widetilde{q})}{(1-\widetilde{q})\Delta(r)} & \text{else} \end{cases} \tag{44}$$

The robustness is zero if $G_c < G(r, \widetilde{q})$.

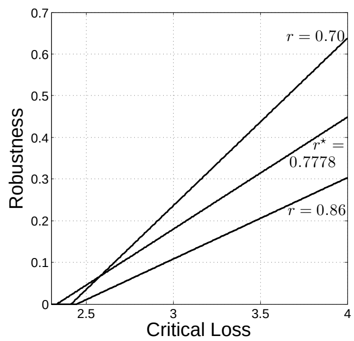

Figure 3: Model-mixing. Robustness curves, $\widehat{h}(r, G_c)$ vs. G_c. $\widetilde{q} = 0.3$, $a = 10$, $b = 15$.

We illustrate robustness curves in fig. 3 with the following penalty functions, $g(r, A) = ar^2$ and $g(r, B) = b(r-1)^2$, where a and b are both positive.

The decision, $r^\star$, which minimizes the estimated loss, $G(r, \widetilde{q})$, is: $r^\star = \frac{(1-\widetilde{q})b}{\widetilde{q}a+(1-\widetilde{q})b}$.

We note in fig. 3 that the robustness curve for $r^\star$ sprouts off the horizontal axis to the left—at lower critical loss—than the other curves. This is necessary since $G(r^\star, \widetilde{q}) \le G(r, \widetilde{q})$ for all r. The robustness curve for any r hits the G_c axis at $G(r, \widetilde{q})$. This is the zeroing property discussed in eq.(5). However, the robustness curve for $r = 0.7$ crosses the robustness curve for $r^\star$ at a fairly low value of G_c.

The conditions of proposition 2 hold in this example. Consequently, if very low loss is required then $r^\star$ is more robust than the other options, and thus will have greater probability of actually achieving the required outcome. On the other hand, if greater loss is tolerable then $r = 0.70$ is more robust than $r^\star$ and thus $r = 0.70$ is more likely to keep the loss below the critical limit.

5.4 Example: One-Sided Forecasting

In this example we consider one-sided forecasting, in which forecast error in one direction (either under- or over-estimate) must not be too large. See also Ben-Haim (2009). Here are some examples. (1) You must catch a plane at the airport on the other side of the metropolis. Being too early is inconvenient but being late is terrible. How long will it take to get to the airport? (2) You must allocate funding for a new project. Under-allocation might mean some problems later on, but over-allocation means other important projects will not be funded at all. How much is needed for the project? (3) You must estimate enemy fire-power and under-estimation can have severe consequences for your forces in the field. (4) Major fiscal programs will increase the rate of inflation unless monetary counter measures are implemented. It is necessary to forecast the amount by which inflation could rise.

One-sided objectives like these are quite common and can reflect contextual understanding of the dominant type of failure. They can also arise due to asymmetric utility: what is perceived as a loss is subjectively costlier than what is perceived as a reward. This asymmetry is a central idea in prospect theory developed by Kahnemann and Tversky (1979).

A forecaster's prediction of the scalar quantity of interest is r, while the true future value, q, is unknown. That is, r is a forecast model developed by the analyst while q is reality. Thus r is the decision and q is the uncertainty, consistent with our notation throughout. The performance function is the error, $G(r,q) = r - q$. If over-prediction must be no larger than $G_{\rm c}$ then the performance requirement is: $G(r,q) \le G_{\rm c}$, where $G_{\rm c}$ will usually be positive. A constraint on under-prediction is represented by the reverse inequality.

Uncertainty in the actual outcome, q, is represented by an info-gap model $\mathcal{Q}(h)$, which does not depend on the prediction, r. The two central conditions of proposition 2—monotonicity of the performance function and independence of the info-gap model—are satisfied and the robustness is a proxy for success in the one-sided forecast requirement. Any change in the forecasting model, r, which enhances the robustness also increases the probability of one-sided forecast success. The forecasting model may be very different from a statistically estimated or scientifically realistic model. Nonetheless, if r's robustness exceeds the robustness of the statistically estimated model (due to crossing of their robustness curves) then r has higher probability of successful one-sided forecasting. Since "success" means "acceptable one-sided error", a model whose robustness at acceptable forecast error is large (or maximal) will be preferred, even if that model is "sub-optimal" as a representation of reality.

Let us note that this example is actually more general than it looks. For instance, suppose that q is the mean of an uncertain probability distribution function (pdf) $p(x)$:

$$\mathrm{E}(x|p) = \int x p(x)\, \mathrm{d}x \tag{45}$$

The info-gap model for uncertainty in q actually embodies uncertainty in the pdf, for example:

$$\mathcal{Q}(h) = \left\{ q = \mathrm{E}(x|p) : \ p(x) \ge 0, \ \int p(x)\, \mathrm{d}x = 1, \ |p(x) - \widetilde{p}(x)| \le h \widetilde{p}(x) \right\}, \quad h \ge 0 \tag{46}$$

In this way the infinite-dimensional uncertainty in the shape of the pdf is "aggregated" into a single scalar uncertainty, q, as required by propositions 1 and 2.

The robustness is defined precisely as in eq.(1). The performance function $G(r,q) = r - q$ is monotonic in a scalar uncertainty, q, and the info-gap model and the pdf are independent of the decision, so proposition 2 holds. The underlying uncertainty in the pdf, however, is far richer than simply the uncertain parameter q.

5.5 Example: Equity Premium Puzzle

Consider a special case of the example in section 4.3 with two assets: one risky ($i = 1$) and one risk-free ($i = 2$). We will illustrate an explanation of the equity premium puzzle based on the proxy property implied by proposition 2.

The uncertainty derives from the unknown payoff of the risky asset in the 2nd period, q_1. The payoff of the risk-free asset, q_2, is known. We do not know a probability distribution for q_1 and we cannot reliably evaluate moments. We have an estimate of the payoff, $\widetilde{q}_1$, which is positive, but the fractional error of this estimate is unknown. Meaningful bounds on the error are unavailable. A simple info-gap model for uncertainty in the payoff is the following unbounded family of nested sets of payoffs:

$$\mathcal{Q}(h) = \{ q_1 : \ |q_1 - \widetilde{q}_1| \le h \widetilde{q}_1 \}, \quad h \ge 0 \tag{47}$$

Other info-gap models are also available.

This info-gap model is independent of the investment, r, if the anticipated payoff $\widetilde{q}_1$ is unaffected by the agent's investment. If the pdf of q_1 is also independent of r then the info-gap model and the pdf are coherent. If the other conditions of proposition 2 hold then robustness is a proxy for probability.

Denote the discounted utility by $U(r, q_1) = u(c_1) + \beta u(c_2)$. For given investments r, the robustness to uncertainty in the payoff q_1 is the greatest horizon of uncertainty h up to which all realizations of the uncertain payoff result in discounted utility no less than U_c:

$$\widehat{h}(r, U_\mathrm{c}) = \max\left\{h : \left(\min_{q_1 \in \mathcal{Q}(h)} U(r, q_1)\right) \geq U_\mathrm{c}\right\} \tag{48}$$

Let U_c be a critical utility which is no larger than the utility anticipated from the estimated return, $U(r, \widetilde{q}_1)$. Assuming positive investment in the risky asset, $r_1 > 0$, and positive marginal utility of $u(c)$, the inner minimum in eq.(48) occurs when $q_1 = (1-h)\widetilde{q}_1$. One can now readily show that the robustness is the solution, for $\widehat{h}$, of:

$$U_\mathrm{c} = u(\underbrace{w - p_1 r_1 - p_2 r_2}_{c_1}) + \beta u[\underbrace{(1-\widehat{h})\widetilde{q}_1 r_1 + q_2 r_2}_{c_2}] \tag{49}$$

Let us assume that $\widehat{r}(U_\mathrm{c})$ is an investment vector which maximizes the robustness: $\partial \widehat{h}(r, U_\mathrm{c})/\partial r_i = 0$, $i = 1, 2$ (conditions for satisfying this assumption are explored in Ben-Haim, 2006). Using this assumption, we can differentiate eq.(49) with respect to r_1 and r_2 to obtain the following relations for the maximal robustness:

$$p_1 \frac{\partial u(c_1)}{\partial c_1} = \beta \frac{\partial u(c_2)}{\partial c_2}(1-\widehat{h})\widetilde{q}_1 \tag{50}$$

$$p_2 \frac{\partial u(c_1)}{\partial c_1} = \beta \frac{\partial u(c_2)}{\partial c_2} q_2 \tag{51}$$

where $\widehat{h} = \widehat{h}(\widehat{r}, U_\mathrm{c})$ is the maximal robustness, at aspiration U_c, obtained with investment $\widehat{r}(U_\mathrm{c})$. Eqs.(50) and (51) are the info-gap generalizations of the first-order conditions in the Lucas asset-pricing model, (Blanchard and Fischer, 1989, p.511, eq.(11)).

The basic trade-off relation, eq.(4), asserts that robustness decreases as aspiration increases: $\widehat{h}(r, U_\mathrm{c})$ decreases as U_c increases. Furthermore, eq.(5) asserts that the robustness vanishes at the anticipated utility: $\widehat{h}(r, U_\mathrm{c}) = 0$ if $U_\mathrm{c} = U(r, \widetilde{q}_1)$. (Both relations hold for arbitrary investment r, as well as for the robust-satisficing investment $\widehat{r}(U_\mathrm{c})$.) An investor who chooses r to maximize the best-estimate of the discounted utility, $U(r, \widetilde{q}_1)$, will have zero robustness for attaining this outcome. Only lower aspirations, $U_\mathrm{c} < U(r, \widetilde{q}_1)$, will have positive robustness. The ordinary Lucas relations—for utility maximization—result when the robustness is zero, $\widehat{h} = 0$, which is a result of the trade-off relations, eqs.(4) and (5).

We can now express the equity premium puzzle and propose a resolution.

Define $\rho_1 = \widetilde{q}_1/p_1$ and $\rho_2 = q_2/p_2$, which are the estimated rates of return for the two assets (there is no uncertainty in ρ_2). Assume that $\beta u'(c_2) \neq 0$, subtract eq.(51) from eq.(50) and re-arrange to obtain a relation asserting that the equity premium is proportional to the robustness:

$$\rho_1 - \rho_2 = \widehat{h}(\widehat{r}, U_\mathrm{c})\rho_1 \tag{52}$$

The lefthand side of eq.(52) is the equity premium: the excess rate of return to the risky asset. Utility-maximizers (as opposed to robust-satisficers) have zero robustness as explained in the previous paragraph, so eq.(52) asserts that optimizers do not require a premium to attract them to the risky asset. The equity premium puzzle can be stated by noting that positive equity premia are universally observed, and yet inconsistent with utility-maximization based on highly uncertain best-estimated return to the risky asset. One possible resolution of the puzzle[5] is to suppose that investors satisfice, rather than optimize, their utility aspirations. The investor does not need to maximize utility in order to justify the investment. It is only necessary to attain an acceptably high reward, greater than anticipated from alternative uses of the resource. Our proxy theorem now explains that satisficers—whose robustness is positive—are more likely to attain acceptable reward than optimizers—whose robustness is zero. Hence the prevalence of positive equity premia, as predicted by eq.(52) for satisficers.

The simplified 2-period example suggests that investors need not forego much utility-aspiration in order to explain ordinary equity premia. If $\rho_1 = 1.07$ and $\rho_2 = 1.01$, then the robustness in eq.(52) which accounts for this 6% premium is $\widehat{h} = 0.06/1.07 \approx 0.056$. This is fairly low robustness (compared to volatility of risky returns on the order of 10 or 20%). Together with the trade-off of robustness against utility, this suggests that investors satisfice only slightly below the nominal maximum.

5.6 Example: Ellsberg Paradox

Mas-Colell *et al.* (1995, p.207) explain the Ellsberg paradox (Ellsberg, 1961) in terms of perceptions of uncertainty which have a natural formulation with info-gap decision theory.

Ellsberg's observation, as adapted by Mas-Colell *et al.*, begins with two urns, R and H, where R contains a well shaken mixture of 49 white and 51 black balls, while H contains an unknown mixture of 100 white and black balls. Two balls are chosen randomly, one from each urn, and their colors are not revealed. The agent must choose one of these balls in each of two experiments. In the first experiment the agent wins \$1000 only if the selected ball is black. Most participants choose the ball from R, suggesting that their subjective probability for a white ball from H is greater than 0.49. In the second experiment the agent wins \$1000 only if the selected ball is white. Again most respondents choose the ball from R even though they are aware that the probability of a white ball from R is precisely 0.49. Ellsberg's paradox is the agent's anomalous disregard for the fact that the subjective probability for white balls is presumably greater for urn H than for urn R.

The experiment is usually performed without requiring a financial investment by the participant. Nonetheless, the participant has a psychological commitment: an aspiration for reward or a desire not to appear foolish. The info-gap explanation of the Ellsberg observation supposes that the agent's commitment (whether psychological or financial) establishes a reservation value $G_{\rm c}$ for expected utility. Expected utility less than $G_{\rm c}$ entails psychological discomfort (or losing money). The dominant uncertainty, q, is the unknown fraction of white balls in urn H, represented by an info-gap model $\mathcal{Q}(h)$.

We now formulate an info-gap explanation of the Ellsberg paradox which is similar to, though not exactly the same as, the explanation in Ben-Haim, 2006, section 11.1.[6] See also Davidovitch (2009, section 4.2).

The known probability of a white ball from the R urn is $p = 0.49$. Let $r = +1$ denote the choice of the ball from the R urn, and let $r = -1$ denote the choice of the ball from the H urn. Let u denote the utility of winning and assume the utility of not winning is zero.

In the 1st experiment—win on black—the expected utility of choice r is:

$$G(r, q) = \frac{1+r}{2}(1-p)u + \frac{1-r}{2}(1-q)u \tag{53}$$

In the second experiment—win on white—the expected utility of choice r is:

$$G(r, q) = \frac{1+r}{2}pu + \frac{1-r}{2}qu \tag{54}$$

[5]There are many possible resolutions. For instance, the observed positive equity premia may not reflect market equilibrium, which has been assumed in deriving the pricing model.

[6]There are many explanations of the Ellsberg paradox, and our explanation does not invalidate the others. Models in behavioral science are under-determined: many different explanations are consistent with the same observations. The point of this example is to illustrate the psychological motivation for robust-satisficing, and to show its consistency with the Ellsberg observation.

In both cases the performance function, $G(r, q)$, depends monotonically on the scalar uncertainty, q. Let $\mathcal{Q}(h)$ denote an info-gap model for the decision maker's uncertainty about q, the probability of white with the H urn. Let $P(q)$ be the probability distribution of q. $\mathcal{Q}(h)$ and $P(q)$ do not depend on the decision, r, so the conditions of proposition 2 hold, and the robustness is a proxy for the probability of an adequate outcome.

As a specific example, consider the following fractional error info-gap model for uncertainty in q:

$$\mathcal{Q}(h) = \left\{ q : \ q \in [0, 1], \ \left| \frac{q - \widetilde{q}}{\widetilde{q}} \right| \le h \right\}, \quad h \ge 0 \tag{55}$$

where $\widetilde{q}$ is the decision maker's guess of the probability of white in urn H.

The robustness of decision r is the greatest horizon of uncertainty, h, up to which the expected utility does not fall short of $G_{\rm c}$:

$$\widehat{h}(r, G_{\rm c}) = \max \left\{ h : \ \left(\min_{q \in \mathcal{Q}(h)} G(r, q) \right) \ge G_{\rm c} \right\} \tag{56}$$

In the first experiment—win on black—one can readily derive the following robustness functions[7] for choosing the R and H urns:

$$\widehat{h}(R, G_{\rm c}) = \left\{ \begin{array}{ll} \dfrac{(1-\widetilde{q})u - G_{\rm c}}{\widetilde{q}u} & \text{if } G_{\rm c} \le (1-\widetilde{q})u \\ 0 & \text{else} \end{array} \right. , \quad \widehat{h}(H, G_{\rm c}) = \left\{ \begin{array}{ll} \infty & \text{if } G_{\rm c} \le (1-p)u \\ 0 & \text{else} \end{array} \right. \tag{57}$$

In the second experiment—win on white—one can readily derive the following robustness functions for choosing the R and H urns:

$$\widehat{h}(R, G_{\rm c}) = \left\{ \begin{array}{ll} \dfrac{\widetilde{q}u - G_{\rm c}}{\widetilde{q}u} & \text{if } G_{\rm c} \le \widetilde{q}u \\ 0 & \text{else} \end{array} \right. , \quad \widehat{h}(H, G_{\rm c}) = \left\{ \begin{array}{ll} \infty & \text{if } G_{\rm c} \le pu \\ 0 & \text{else} \end{array} \right. \tag{58}$$

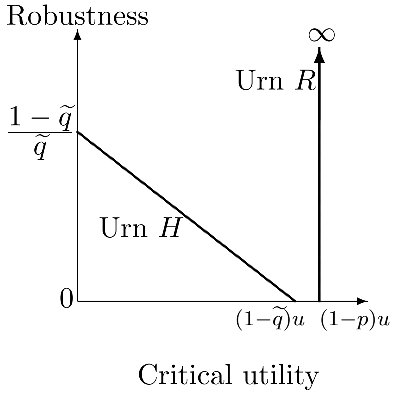

Figure 4: Robustness curves for Ellberg's first experiment.

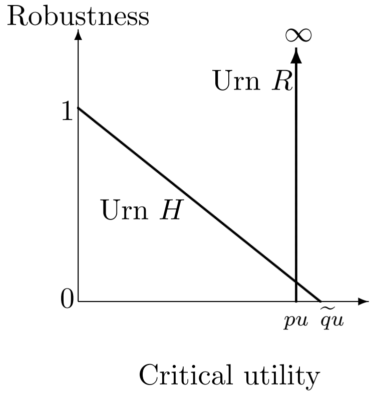

Figure 5: Robustness curves for Ellberg's second experiment.

In Ellsberg's first experiment—win on black—the typical participant chooses the R urn, which presumably reveals that the participant's guess of the probability of black in the H urn, $1 - \widetilde{q}$, is less than $1 - p$. Hence the robustness curves in eqs.(57) will appear as in fig. 4. The R urn is more robust than the H urn at all levels of expected utility. But since $\widetilde{q} > p$, the robustness curves for the second experiment, eqs.(58), will cross one another as seen in fig. 5. If $\widetilde{q}$ is near p this crossing occurs very near the anticipated expected utility from the H urn, meaning that the R urn is robust-dominant

[7] For clarity we are corrupting the notation and denoting the decision by the name of the urn, R or H, rather than by the value of r, $+1$ or -1.

over almost all the range of utility. The robust-satisficing decision maker will again choose R, as Ellsberg's experiments tended to show.

Most of the agents in both of Ellsberg's experiments choose robustness-maximizing urns, which, according to our proxy theorem, is equivalent to maximizing the probability of satisficing the expected utility. There is nothing anomalous about Ellsberg's observation if decision makers are robust-satisficers. And there is nothing anomalous about the robust-satisficing strategy since it is a proxy for the probability of success in Ellsberg's experiments. The robust-satisficer will tend to achieve required goals more frequently than the best-model optimizer.

6 Proxy Theorem: Monotonicity and Standardization

Proposition 1 depends on two properties: monotonicity of the performance function, $G(r, q)$, in a single scalar uncertainty, q, and coherence of the info-gap model $\mathcal{Q}_r(h)$ and the probability distribution $P(q|r)$. We have shown through several examples (sections 4.3–4.5, 5.4) how monotonicity in a scalar variable can be obtained by using an aggregate uncertain function when the underlying uncertainty is not a scalar. We now discuss the idea of standardization, employed in section 6.3, and show that it implies coherence for a particular info-gap model. This then shows that standardization implies the proxy property when this info-gap model is used.

6.1 Standardization

Definition 4 *Let q be a scalar random variable with a pdf which depends on parameters r. The pdf is **standardizable** and $\theta(q, r)$ is a **standardization function** if $\theta(q, r)$ is a scalar function which is strictly increasing and continuous in q at any fixed r and whose pdf is the same for all r.*

This concept of standardization is somewhat different from the usual probabilistic concept of standardization, which is the transformation of a variate to a form having zero mean and unit variance. Most, though not all, of our examples of standardization have this 0–1 property. However, the gist of definition 4 is that the distribution which is obtained by the transformation has no information about the distribution from which the transformation arises. Note that the standardized random variable, $\theta(q, r)$, need not belong to the family to which q belongs.

For example, suppose that the mean and standard deviation of q, $\mu(r)$ and $\sigma(r)$, depend on the decision parameters r. Define the function $\theta(q, r) = [q - \mu(r)]/\sigma(r)$. This function standardizes many families of distributions, both in the sense of definition 4 and in the usual probabilistic sense. The normal distribution is standardized to the normal distribution with zero mean and unit variance. The uniform distribution on the interval $[a, b]$ is standardized to the uniform distribution on $[-1/\sqrt{3},\ 1/\sqrt{3}]$. The exponential distribution is standardized to the density $\mathrm{e}^{1-\theta}$ on $\theta > -1$.

As a different example consider the Cauchy distribution on $(-\infty, \infty)$, $p(q|r) = 1/[r\pi(1 + (q/r)^2)]$, whose mean and the variance do not exist. A standardization function is $\theta = q/r$ whose pdf is $1/[\pi(1 + \theta^2)]$.

A 1-sided distribution whose mean and variance do not exist is: $f(q|r) = \frac{r}{q^2}$, $q \geq r$, where r is positive. A standardization function is again $\theta = q/r$ whose pdf is $1/\theta^2$ for $\theta \geq 1$.

Any standardization function $\theta(q, r)$ generates this info-gap model for uncertain q:

$$\mathcal{Q}_r(h) = \{q : \ |\theta(q, r)| \leq h\}, \quad h \geq 0 \tag{59}$$

This info-gap model will play a role in proposition 3.

A final comment on standardization functions. Definition 4 allows $\theta = P(q|r)$ as a standardization function, since the pdf of this θ is uniform on the interval [0, 1] in all cases. While this θ is a standardization function, we are interested in standardization functions whose specification requires less information that the full cpd of q.

6.2 Standardization and Coherence

The following lemma, whose proof appears in the Appendix, explains the importance of the concept of standardizability.

Lemma 1 *Standardizability implies coherence.*

Given:

• *The probability distribution of the scalar variable* q*,* $P(q|r)$*, is standardizable with a standardization function* $\theta(q, r)$ *which is strictly increasing and continuous in* q *at any fixed* r*.*

• $G(r, q)$ *is a performance function which is monotonic (though not necessarily strictly monotonic) in* q *at any fixed decision* r*.*

• r_1 *and* r_2 *are decisions with positive robustness at critical value* $G_{\rm c}$*, using the info-gap model* $\mathcal{Q}_r(h)$ *in eq.(59).*

Then: $\mathcal{Q}_r(h)$ *and* $P(q|r)$ *are upper (lower) coherent at* r_1*,* r_2 *and* $G_{\rm c}$ *if the performance function* $G(r, q)$ *is increasing (decreasing) in* q*.*

Lemma 1 and proposition 1 enable us to assert the following proposition. A special case of this proposition appears in Ben-Haim (2006, section 11.4.2).

Proposition 3 *Info-gap robustness to an uncertain scalar variable, with a loss function which is monotonic in the uncertain variable, is a proxy for probability if the probability distribution is standardizable.*

Given:

• *The probability distribution of the scalar variable* q*,* $P(q|r)$*, is standardizable with a standardization function* $\theta(q, r)$ *which is strictly increasing and continuous in* q *at any fixed* r*.*

• $G(r, q)$ *is a performance function which is monotonic (though not necessarily strictly monotonic) in* q *at any fixed decision* r*.*

• r_1 *and* r_2 *are decisions with positive robustness at critical value* $G_{\rm c}$*, using the info-gap model* $\mathcal{Q}_r(h)$ *in eq.(59).*

Then: *The proxy property holds at* r_1*,* r_2 *and* $G_{\rm c}$ *with the performance function* $G(r, q)$*.*

The standardization property is rather specific but nonetheless relevant to an important class of problems. Our examples in section 6.1 showed that the same transformation can standardize pdf's from totally different families, e.g. the normal and uniform families. An adaptive search for a standardization function is re-enforced by the proxy property: probability of survival is maximized by a robust-satisficing agent who is able to standardize. This means that the probability of survival can be maximized without knowing the pdf or even its family, provided that all the pdf's belong to the same family and a standardization function is found.

6.3 Example: Risky Assets Revisited

We now illustrate how proposition 3 provides an additional method to handle the risky-asset example in section 4.3.

We consider two risky assets in a 2-period investment where both payoffs, q_1 and q_2, are uncertain. The consumption in the second period, $c_2 = q^T r$ in eq.(18), is a scalar uncertainty. The performance function, $G(r, q)$ eq.(19), is monotonic in c_2 which is the only uncertainty in $G(r, q)$.

c_2 depends on the investment vector r so it is plausible that the pdf of c_2 depends on r as well. But it can happen that the *family* of pdf's does not change as r changes. For instance, the pdf's may all be normal, or they may all be uniform, etc. Recall that the same transformation can standardize more than one family. If the agent can find a standardizing transformation for the family of pdf's—whatever it is—then proposition 3 holds and robustness is a proxy for probability of success.

The search for a standardizing transformation is re-enforcing because of the proxy property which endows standardized transformations with survival advantage.

7 Summary and Discussion

This paper presents an approach to economic rationality, linking Knightian uncertainty, robustness and satisficing behavior in a coherent quantitative theory. The paper identifies general conditions for the competitive advantage of robust-satisficing, facilitating an understanding of satisficing behavior under uncertain competition. We have used a concept of robustness which is consistent with current economic literature (e.g. Hansen and Sargent, 2008), and which has a long tradition in engineering literature (e.g. Schweppe, 1973). We have shown that, in many circumstances, robust-satisficing behavior is more likely to meet the requirements for survival, than any other strategy including behavior based on optimization with the best available (but faulty) models and data. This has been illustrated for a range of economic and related situations, including investment in risky assets, the equity premium puzzle, Ellsberg's paradox, monetary policy formulation, principal-agent contracts, Bayesian model mixing, foraging, and forecasting.

The results are based on the properties of info-gap models of uncertainty. An info-gap model is a stark non-probabilistic quantification of the disparity between the best available information and full knowledge. An info-gap model is a family of nested sets whose elements are scalars, vectors, functions or sets, $\mathcal{Q}(h)$, $h \geq 0$, characterized by the contraction and nesting axioms (see section 2). An info-gap model is a quantification of Knightian uncertainty and does not entail identification of a worst case.

The info-gap robustness to uncertainty in a scalar, vector, function, or set, q, with decision r, is the greatest horizon of uncertainty h up to which the loss, $G(r, q)$, cannot exceed G_c, eq.(1). (A similar definition applies when considering reward rather than loss.) Info-gap robustness is consistent with other definitions of robustness. The robustness function, $\widehat{h}(r, G_c)$, generates preferences on the decision, eq.(2). Robustness curves—$\widehat{h}(r_i, G_c)$ vs. G_c—for different decisions r_i may cross one another, implying reversal of preferences as discussed in connection with figs. 1, 2 and 3. This curve-crossing may occur for the best-model outcome-optimizing decision, $r^\star$, and the robust-satisficing decision $\widehat{r}(G_c)$, eqs.(6) and (3), implying that the maximum-robustness decision can differ from the putative outcome-optimizing decision.

In a competitive environment, agents may be removed if their losses exceed some relevant "survival" level (or if their rewards are too low). Less productive firms leave the market, less accurate forecasters are not consulted, less successful foragers (human or animal) may die. Survival does not require absolute optimality. Survival requires being good enough, meeting environmental challenges, or beating the competition. 'Survival of the fittest' means 'survival of the more fit over the less fit', not necessarily of the global-optimally fit.

In a competitive environment, the probability of survival, and other factors, determine the course of long-term evolution. However, in complex, variable and uncertain environments, the bounded rationality of the agents may preclude the selection of an action directly in terms of its survival probability. This is most obviously the case when the relevant probability distributions are unknown to the agents. An example is the evolution of an industry's technology in which successful firms innovate, imitate and grow based on using organizational routines which satisfice a goal rather than using global optimization (Iwai, 2000). Another example is the successfulness of simple heuristic decision rules (Gigerenzer and Selten, 2001).

The proxy theorems in this paper suggest an explanation of why agents robust-satisfice in order to survive in an uncertain competitive environment. Actions which are sub-optimal when evaluated with the best available models may, in fact, have greater survival probability than the putatively optimal actions. Optimization with faulty models and data is not necessarily the best bet for survival, which brings us closer to an understanding of why outcome-optimization has often failed to explain economic and ethological puzzles and paradoxes such as the equity premium paradox, the home bias puzzle, and foraging by animals and economic agents.

The proxy theorems depend on several structural assumptions. Foremost, all three propositions assume that the performance function, $G(r, q)$, depends monotonically on a single scalar uncertainty, q. This is much less restrictive than it may at first appear. As we have shown in numerous examples, the underlying uncertainty may be a vector (e.g. an uncertain vector of returns) or a function (e.g. an uncertain pdf). If the agent's survival requirement is that the scalar performance must satisfy an inequality, then the condition of monotonicity will hold by adopting a high-level or aggregate term in the performance function as the scalar uncertainty. This was illustrated in the forecasting example in section 5.4 in which the scalar uncertainty is actually the mean of an uncertain pdf. Or, in the examples in sections 4.3 and 6.3 the underlying uncertainty is a vector of returns over multiple time steps and the scalar uncertainty is either the consumption in the last period or a partial sum in the discounted utility. Similar aggregation is demonstrated in sections 4.4 and 4.5 dealing with monetary policy under uncertain expectations, and the principal-agent problem. In short, complicated multi-dimensional underlying uncertainties can be aggregated in a scalar performance function to satisfy the monotonicity requirement of the proxy theorems.

Our most general result—proposition 1—assumes that the probability distribution and the info-gap model are coherent as specified in definition 1. Coherence entails weak informational overlap between the probability distribution and the info-gap model. We have shown that coherence holds in many situations, including simple examples (section B), risky-assets (section 4.3), monetary policy (section 4.4) and the principal-agent problem (section 4.5), for a wide range of probability distributions including normal, Cauchy and gamma distributions, and all the examples in sections 5.2–5.6.

But of course the proxy property—which implies that the probability of survival can be maximized by maximizing a non-probabilistic robustness—will not always hold. The three propositions establish conditions under which the probability of survival *can* be maximized without knowing the probability distribution. However, these conditions need not obtain in practice. This has an important implication for learning and adaptation under uncertain competition. In light of the proxy theorems, learning can focus on weakly characterizing the uncertainty. Proposition 1 implies that the agent just needs to learn enough to formulate an info-gap model which is coherent with the probability distribution. As illustrated by the examples in appendix B, coherence is obtained with very limited informational overlap between the pdf and the info-gap model. Proposition 3 implies that the proxy property holds if the agent learns enough to standardize the pdf. This does not necessarily require knowledge of the family to which the pdf belongs, as seen in the examples in section 6.1.

These considerations suggest that learning need not entail developing models with high fidelity to reality. Rather, learning can focus on characterizing the info-gap between what the agent *does* know, and *does not* know. Once this info-gap is adequately characterized, as defined by coherence, the agent can maximize the probability of survival. And of course the learning need not be explicit or intentional, but simply a process of trial and error about which the agent may have no awareness at all. The proxy property imbues robust-satisficing strategies with the competitive advantage of being more likely to satisfy critical requirements than any other strategy. This re-enforces the use of these strategies.

8 References

Ben-Haim, Y., 2005. Info-gap Decision Theory For Engineering Design. Or: Why 'Good' is Preferable to 'Best', chapter 11 in *Engineering Design Reliability Handbook,* E. Nikolaides, D. Ghiocel and Surendra Singhal, eds., CRC Press.

Ben-Haim, Y., 2006. *Info-Gap Decision Theory: Decisions Under Severe Uncertainty,* 2nd edition, Academic Press, London.

Ben-Haim, Y., 2009. Info-gap forecasting and the advantage of sub-optimal models, *European Journal of Operational Research,* 197: 203–213.

Ben-Haim, Y., 2010. *Info-Gap Economics: An Operational Introduction,* Palgrave.

Ben-Haim, Y. Dacso, C.C., Carrasco J. and Rajan, N., 2009. Heterogeneous Uncertainties in Cholesterol Management, *International Journal of Approximate Reasoning,* 50: 1046–1065.

Ben-Tal, A., Nemirovski, A., 1999. Robust solutions of uncertain linear programs, *Oper. Res. Lett.* 25, 1–13.

Blanchard, O.J., Fischer, S., 1989. *Lectures on Macroeconomics,* MIT Press.

Burgman, M., 2005. *Risks and Decisions for Conservation and Environmental Management,* Cambridge University Press, Cambridge.

Carmel, Y., Ben-Haim, Y., 2005. Info-gap robust-satisficing model of foraging behavior: Do foragers optimize or satisfice?, *American Naturalist,* 166, 633–641.

Clarida, R., Galí, J., Gertler, M., 1999. The Science of Monetary Policy: A New Keynesian Perspective, *Journal of Economic Literature,* XXXVII(4), 1661–1707.

Conlisk, J., 1996. Why bounded rationality? *Journal of Economic Literature,* 34, 669–700.

Crain, W.M., Shugart, W.F. II, Tollison, R.D., 1984. The convergence of satisficing to marginalism: An empirical test, *Journal of Economic Behavior & Organization,* 5, 375–385.

Davidovitch, L., 2009, *Strategic Interactions Under Severe Uncertainty,* PhD Thesis, Technion—Israel Institute of Technology.

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

French, K., Poterba, J., 1990. Japanese and U.S. Cross-Border Common Stock Investments, *Journal of the Japanese and International Economics,* 4, 476–493.

French, K., Poterba, J., 1991. Investor Diversification and International Equity Markets (in Behavioral Finance), *American Economic Review,* Papers and Proceedings of the Hundred and Third Annual Meeting of the American Economic Association. (May, 1991), 81(2), 222–226.

Gabaix, X., Laibson, D., Moloche G., Weinberg, S., 2006. Costly information acquistion: Experimental analysis of a bounded rational model, *American Economic Review,* 96(4), 1043–1068.

Gigerenzer, G., Selten, R., eds., 2001. *Bounded Rationality: The Adaptive Toolbox,* MIT Press.

Gutierrez, G.J., Kouvelis, P., Kurawarwala, A.A., 1996. A robustness approach to uncapacitated network design problems, *Eur. J. Oper. Res.* 94, 362–376.

Hansen, L.P., and Sargent, T.J., 2008. *Robustness,* Princeton University Press, Princeton and Oxford.

Hites, R., De Smet, Y., Risse, N., Salazar-Neumann, M., Vincke, P., 2006. About the applicability of MCDA to some robustness problems, *Eur. J. Oper. Res.* 174, 322–332.

Huber, P.J., 1981. *Robust Statistics,* John Wiley, New York.

Iwai, K., 2000. A contribution to the evolutionary theory of innovation, imitation and growth, *Journal of Economic Behavior & Organization,* 43, 167–198.

Jeske, K., 2001. Equity Home Bias: Can Information Cost Explain the Puzzle?, *Economic Review,* Federal Reserve Bank of Atlanta, Third Quarter, available at http://www.frbatlanta.org/filelegacydocs/ACF62A.pdf.

Kahneman, D., 2003, Maps of bounded rationality: Psychology for behavioral economics, *American Economic Review,* 93(5), 1449–1475.

Kahneman, D., Tversky, A., 1979. Prospect theory: An analysis of decision under risk, *Econometrica,* XLVII, 263–291.

Kaufman, B.E., 1999. Emotional arousal as a source of bounded rationality, *Journal of Economic Behavior & Organization,* 38, 135–144.

Knight, F.H., 1921. *Risk, Uncertainty and Profit,* Hart, Schaffner and Marx. Re-issued by Harper Torchbooks, New York, 1965.

Kocherlakota, N.R., 1996. The equity premium: It's still a puzzle, *Journal of Economic Literature,* 34, 42–71.

Mas-Colell, A., Whinston, M.D., Green, J.R., 1995. *Microeconomic Theory,* Oxford University Press.

Mehra, R. Prescott, E.C., 1985. The equity premium: A puzzle, *Journal of Monetary Economics,* 15, 145–161.

Nonacs, P., 2001, State dependent behavior and the Marginal Value Theorem, *Behavioral Ecology,* 12(1), 71–83.

Roy, B., 2010. Robustness in operational research and decision aiding: A multi-faceted issue, *Eur. J. Oper. Res.* 200, 629–638.

Schweppe, F.C., 1973. *Uncertain Dynamic Systems,* Prentice-Hall, Englewood Cliffs.

Sen, A., 1997. Maximization and the act of choice, *Econometrica,* 65(4), 745–779.

Simon, H.A., 1955. A behavioral model of rational choice, *Quarterly Journal of Economics,* 69(1), 99–118.

Simon, H. A., 1956. Rational choice and the structure of the environment, *Psychological Review,* 63(2), 129–138.

Simon, H.A., 1979. Rational decision making in business organizations, *American Economic Review,* 69(4), 493–513.

Stiglitz, J.E., 1975. Incentives, risk , and information: Notes towards a theory of hierarchy, *The Bell Journal of Economics,* 6(2), 552–579.

Stiglitz, J.E., 1998. Article on 'principal and agent' in J.Eatwell, M.Milgate and P.Newman, eds., *The New Palgrave: A Dictionary of Economics,* 3, 966–972, Palgrave Publishers, New York.

Thaler, R.H., 1994. Psychology and savings policies, *American Economic Review,* 84(2), 186–192.

Wald, A., 1945. Statistical decision functions which minimize the maximum risk, *Annals of Mathematics,* 46(2), 265–280.

Ward, D., 1992. The role of satisficing in foraging theory, *Oikos,* 63(2), 312–317.

Part II

Appendices

A Why $\succ_{\rm r}$ and $\succ_{\rm p}$ Are Not Necessarily Equivalent

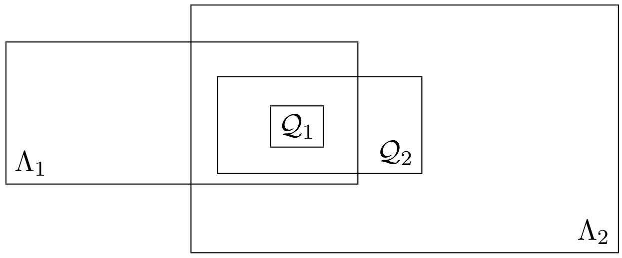

Figure 6: Uncertainty sets.

Proposition 1 establishes conditions for the equivalence of $\succ_{\rm r}$ and $\succ_{\rm p}$, defined in eqs.(2) and (8). However, $\succ_{\rm r}$ and $\succ_{\rm p}$ are not *necessarily* equivalent, as we now explain with the aid of fig. 6. See also Davidovitch (2009).

Consider two actions, r_1 and r_2, with robustnesses $\widehat{h}(r_1, G_{\rm c}) < \widehat{h}(r_2, G_{\rm c})$ based on an info-gap model $\mathcal{Q}_r(h)$. Denote $\widehat{h}_i = \widehat{h}(r_i, G_{\rm c})$ and $\mathcal{Q}_i = \mathcal{Q}(\widehat{h}_i, \widetilde{q})$. As in section 2, define $\Lambda(r, G_{\rm c})$ as the set of all q's for which $G(r, q) \le G_{\rm c}$. Denote $\Lambda_i = \Lambda(r_i, G_{\rm c})$, for $i = 1$, 2. The sets $\mathcal{Q}_i$ belong to an info-gap model and represent the agent's beliefs, while the sets Λ_i differ from the sets $\mathcal{Q}_i$ and do not reflect the agent's beliefs.

The sets $\mathcal{Q}_i$ are nested as shown in the figure, $\mathcal{Q}_1 \subseteq \mathcal{Q}_2$, because the robustnesses are ranked, $\widehat{h}_1 < \widehat{h}_2$. Furthermore, $\mathcal{Q}_i$ must belong to Λ_i. However, the sets Λ_i need not be nested; each may contain a region not belonging to the other, as shown. Consequently, there is no constraint, in general, on the relation between $P(\Lambda_1)$ and $P(\Lambda_2)$; either may exceed the other, depending on the structure of the probability distribution. In general, ranked robustness does not imply ranked probability of survival, and ranked probability of survival does not imply ranked robustness.

B Coherence: Further Insight and Simple Examples

Some further insight into the meaning of coherence is obtained by considering a special case. Let the decision, r, be a scalar variable, and choose $r_i = r_j + \epsilon$ where $0 < \epsilon \ll 1$. Then, assuming differentiability, eqs.(13) and (14) become:

$$- \left. \frac{\partial P(q|r_j)}{\partial r_j} \right|_{q=G^{-1}(r_j, G_{\rm c})} < \frac{\partial G^{-1}(r_j, G_{\rm c})}{\partial r_j} \left. \frac{\partial P(q|r_j)}{\partial q} \right|_{q=G^{-1}(r_j, G_{\rm c})} \qquad (60)$$

$$\left. \frac{\partial q^\star(h, r_j)}{\partial r_j} \right|_{h=\widehat{h}(r_j, G_{\rm c})} < \frac{\partial G^{-1}(r_j, G_{\rm c})}{\partial r_j} \qquad (61)$$

The second derivative on the righthand side of eq.(60) is a pdf, which is non-negative. If this pdf is positive then we can re-write these relations as:

$$\frac{-1}{p[G^{-1}(r_j, G_{\rm c})|r_j]} \left. \frac{\partial P(q|r_j)}{\partial r_j} \right|_{q=G^{-1}(r_j, G_{\rm c})} < \frac{\partial G^{-1}(r_j, G_{\rm c})}{\partial r_j} \qquad (62)$$

$$\left. \frac{\partial q^\star(h, r_j)}{\partial r_j} \right|_{h=\widehat{h}(r_j, G_{\rm c})} < \frac{\partial G^{-1}(r_j, G_{\rm c})}{\partial r_j} \qquad (63)$$

We can understand the coherence between $\mathcal{Q}_r(h)$ and $P(q|r)$ which is implied by these relations as follows. Recall that $G(r,q)$ is the performance function, and $G^{-1}(r,G_{\rm c})$ is the q-value which produces the critical system-response $G_{\rm c}$. Suppose that $\partial G^{-1}/\partial r < 0$. Thus, if $\mathcal{Q}_r(h)$ and $P(q|r)$ are coherent, then $q^\star$ decreases and P increases as r goes down. If we are able to evaluate the response of $q^\star$ to a change in the decision, r, then we know something about the response of the cpd. The reverse is also true, from P to $q^\star$. In other words, coherence implies some weak informational overlap between the probability distribution and the non-probabilistic info-gap model of uncertainty.

We now examine two simple examples of coherence between an info-gap model and a probability distribution. In sections 4.3–4.5 we will consider more realistic examples for risky assets, monetary policy, and the principal-agent problem.

Example 1 Let the performance function be $G(r,q) = q/r$ with positive r and q, so $G^{-1}(r,G_{\rm c}) = rG_{\rm c}$ and $\partial G^{-1}/\partial r = G_{\rm c}$ for any positive critical value $G_{\rm c}$. Consider an exponential distribution, so $P(q|r) = 1 - {\rm e}^{-rq}$ for $q \ge 0$ and $\partial P/\partial r = q{\rm e}^{-qr}$. Use the following asymmetric info-gap model:

$$\mathcal{Q}_r(h) = \left\{ q:\ 0 \le q \le \frac{h}{r} \right\}, \quad h \ge 0 \tag{64}$$

One finds $q^\star(h,r) = h/r$ so $\partial q^\star/\partial r = -h/r^2$. The robustness of performance requirement $G(r,q) \le G_{\rm c}$ is $\widehat{h} = r^2 G_{\rm c}$. Eqs.(62) and (63) each reduce to $-1 < 1$, so they both hold. The two uncertainty models, $\mathcal{Q}_r(h)$ and $P(q|r)$, are coherent. Looking at the specific forms of $P(q|r)$ and $\mathcal{Q}_r(h)$ we see that, as r increases, $P(q|r)$ and $\mathcal{Q}_r(h)$ both become more highly concentrated. One would *not* say that $\mathcal{Q}_r(h)$ is a good representation of $P(q|r)$, or the reverse. On the contrary: these two uncertainty models are utterly different from each other; one is probabilistic and one is not. Nonetheless, each reveals something about the other. There is some "coherence" between them. ▮

Example 2 Use the probability distribution of example 1, let the performance function be $G(r,q) = qr^{-\alpha}$ with positive r and q, and use the following info-gap model rather than eq.(64):

$$\mathcal{Q}_r(h) = \{ q:\ 0 \le q \le rh \}, \quad h \ge 0 \tag{65}$$

Eq.(62) reduces to $-1 < \alpha$ and eq.(63) becomes $1 < \alpha$. These two uncertainty models, $\mathcal{Q}_r(h)$ and $P(q|r)$, are incoherent when $\alpha \le 1$. This seems reasonable since $\mathcal{Q}_r(h)$ becomes more dispersed as r increases while $P(q|r)$ becomes more concentrated as r increases. However, one must be cautious in interpreting coherence, since $\mathcal{Q}_r(h)$ and $P(q|r)$ *are* coherent with this performance function if $\alpha > 1$. When $\alpha > 1$ the system model $G(r,q)$ decreases "strongly enough" as r increases to make the info-gap model and cpd coherent. We see that coherence is a property of the uncertainty models together with the performance function. ▮

C Proofs

C.1 Proposition 1

We need a lemma before proving proposition 1. The gist of this lemma is to establish conditions under which the inverse function, $G^{-1}(r,G_{\rm c})$, equals either $q^\star$ or $q_\star$.

Lemma 2 ***Given:***

• *At any fixed decision r, the performance function, $G(r,q)$, is monotonic (though not necessarily strictly monotonic) in the scalar q.*

• *$\mathcal{Q}_r(h)$ is an info-gap model with the property of nesting.*

• *$\mathcal{Q}_r(h)$ is continuously upward (downward) expanding at $\widehat{h}(r,G_{\rm c})$ if $G(r,q)$ increases (decreases) with increasing q.*

Then, *if $G(r,q)$ is **increasing** in q:*

$$q^\star[\widehat{h}(r,G_{\rm c}), r] = G^{-1}(r,G_{\rm c}) \tag{66}$$

*and if $G(r,q)$ is **decreasing** in q:*

$$q_{\star}[\widehat{h}(r,G_{\rm c}),r] = G^{-1}(r,G_{\rm c}) \tag{67}$$

Proof of lemma 2. We will prove eq.(66). Proof of eq.(67) is analogous and will not be elaborated.

Using the definition of robustness in eq.(1) and the monotonicity of $G(r,q)$ we can write the robustness as:

$$\begin{aligned} \widehat{h}(r,G_{\rm c}) &= \max\left\{h: \left(\max_{q\in\mathcal{Q}_r(h)} G(r,q)\right) \le G_{\rm c}\right\} & (68)\\ &= \max\left\{h: \left(\max_{q\in\mathcal{Q}_r(h)} q\right) \le G^{-1}(r,G_{\rm c})\right\} & (69)\\ &= \max\left\{h:\ q^{\star}(h,r) \le G^{-1}(r,G_{\rm c})\right\} & (70)\end{aligned}$$

recalling the definition of $q^{\star}(h,r)$ in eq.(10).

For notational convenience let us define the function:

$$\gamma(h,r) = G^{-1}(r,G_{\rm c}) - q^{\star}(h,r) \tag{71}$$

By eq.(70):

$$\gamma[\widehat{h}(r,G_{\rm c}),r] \ge 0 \tag{72}$$

Suppose:

$$\gamma[\widehat{h}(r,G_{\rm c}),r] > 0 \tag{73}$$

Then, since $\mathcal{Q}_r(h)$ is continuously upper expanding at $\widehat{h}(r,G_{\rm c})$, there is an $h' > \widehat{h}(r,G_{\rm c})$ such that:

$$\gamma(h',r) > 0 \tag{74}$$

which implies that the robustness is no less than h' and exceeds $\widehat{h}(r,G_{\rm c})$, which is a contradiction. Hence the supposition in eq.(73) is false and we have proven that:

$$\gamma[\widehat{h}(r,G_{\rm c}),r] = 0 \tag{75}$$

This completes the proof. ▮

The following related lemma will be useful later.

Lemma 3 ***Given:***

• *At any fixed decision r, the performance function, $G(r,q)$, is monotonic (though not necessarily strictly monotonic) in the scalar q.*

• *$\mathcal{Q}_r(h)$ is an info-gap model with the property of nesting.*

• *$\mathcal{Q}_r(h)$ is continuously upward (downward) expanding at $\widehat{h}(r,G_{\rm c})$ if $G(r,q)$ increases (decreases) with increasing q.*

• *r_1 and r_2 are two decisions with positive robustness at critical value $G_{\rm c}$.*

Then, *if $G(r,q)$ is **increasing** in q, and $j = 3-i$:*

$$\gamma[\widehat{h}(r_i,G_{\rm c}),r_j][\widehat{h}(r_i,G_{\rm c}) - \widehat{h}(r_j,G_{\rm c})] < 0 \tag{76}$$

*And if $G(r,q)$ is **decreasing** in q then eq.(76) holds when $\gamma(h,r)$ in eq.(71) is defined with $q_{\star}$ rather than $q^{\star}$.*

It is sometimes useful to write eq.(76) more explicitly as follows:

$$\gamma[\widehat{h}(r_i,G_{\rm c}),r_j] < 0 \quad \text{if and only if} \quad \widehat{h}(r_i,G_{\rm c}) > \widehat{h}(r_j,G_{\rm c}) \tag{77}$$

$$\gamma[\widehat{h}(r_i,G_{\rm c}),r_j] > 0 \quad \text{if and only if} \quad \widehat{h}(r_i,G_{\rm c}) < \widehat{h}(r_j,G_{\rm c}) \tag{78}$$

Proof of lemma 3. We will prove eq.(77). The proof of eq.(78) is analogous and will not be elaborated. Likewise we will only consider the case that $G(r,q)$ is increasing in q.

First we note that $q^\star(h,r)$ is a (not necessarily strictly) increasing function of h, at fixed r, because $\mathcal{Q}_r(h)$ is a nested info-gap model.

(1) Suppose that the lefthand inequality in eq.(77) holds. That is, $-\gamma[\widehat{h}(r_i,G_c),r_j] > 0$ which more explicitly is:

$$q^\star[\widehat{h}(r_i,G_c),r_j] > G^{-1}(r_j,G_c) \tag{79}$$

From the expression for robustness in eq.(70) in the proof of lemma 2, and from the monotonic increase of $q^\star(h,r)$ in h, eq.(79) implies the righthand inequality in eq.(77).

(2) Suppose that the righthand inequality in eq.(77) holds. Monotonicity of $q^\star(h,r)$ implies:

$$q^\star[\widehat{h}(r_i,G_c),r_j] \geq q^\star[\widehat{h}(r_j,G_c),r_j] \tag{80}$$

Suppose equality in eq.(80):

$$q^\star[\widehat{h}(r_i,G_c),r_j] = q^\star[\widehat{h}(r_j,G_c),r_j] \tag{81}$$

By lemma 2:

$$q^\star[\widehat{h}(r_j,G_c),r_j] = G^{-1}(r_j,G_c) \tag{82}$$

Hence the supposition in eq.(81) implies:

$$q^\star[\widehat{h}(r_i,G_c),r_j] = G^{-1}(r_j,G_c) \tag{83}$$

But from the expression for robustness in eq.(70) in the proof of lemma 2, this implies:

$$\widehat{h}(r_j,G_c) \geq \widehat{h}(r_i,G_c) \tag{84}$$

This contradicts the righthand inequality in eq.(77), so the supposition in eq.(81) is false and eq.(80) is a strict inequality. Thus, from the definition of $\gamma(h,r)$ in eq.(71):

$$\gamma[\widehat{h}(r_i,G_c),r_j] < \gamma[\widehat{h}(r_j,G_c),r_j] = 0 \tag{85}$$

where the equality on the right results from lemma 2. This is the lefthand inequality in eq.(77).

This completes the proof. ∎

Proof of proposition 1. We prove the proposition for the case that $G(r,q)$ increases monotonically in q. The analogous proof for monotonic decrease will not be elaborated.

From eq.(7), and using the monotonicity of $G(r,q)$ and the definition of G^{-1} in eq.(11), we can write the probability of survival as:

$$\begin{aligned} P_s(r,G_c) &= \mathrm{Prob}[G(r,q) \leq G_c | r] && (86)\\ &= \mathrm{Prob}[q \leq G^{-1}(r,G_c)|r] && (87)\\ &= P[G^{-1}(r,G_c)|r] && (88) \end{aligned}$$

Using the definition of robustness in eq.(1) and the monotonicity of $G(r,q)$ we can write the robustness as in eq.(70) in the proof of lemma 2:

$$\widehat{h}(r,G_c) = \max\left\{h: \ q^\star(h,r) \leq G^{-1}(r,G_c)\right\} \tag{89}$$

recalling the definition of $q^\star(h,r)$ in eq.(10).

We first assume that $\mathcal{Q}_r(h)$ and $P(q|r)$ are coherent and prove eq.(16): items (1) and (2) below. We then prove the converse: items (3) and (4).

(1) We now prove that the righthand inequality in eq.(16) implies the lefthand inequality, assuming that $\mathcal{Q}_r(h)$ and $P(q|r)$ are coherent. $P(q|r)$ is a cpd so it is non-decreasing in q. Thus the righthand inequality in eq.(16), together with eq.(88), imply:

$$P[G^{-1}(r_1,G_c)|r_1] > P[G^{-1}(r_2,G_c)|r_2] \tag{90}$$

This, together with the supposition of coherence, imply:

$$G^{-1}(r_1, G_\mathrm{c}) - q^\star(h, r_1) > G^{-1}(r_2, G_\mathrm{c}) - q^\star(h, r_2) \tag{91}$$

where h equals either $\widehat{h}(r_1, G_\mathrm{c})$ or $\widehat{h}(r_2, G_\mathrm{c})$.

As before, let us denote $\gamma_i(h) = G^{-1}(r_i, G_\mathrm{c}) - q^\star(h, r_i)$ for $i = 1$ or 2. Note that $\gamma_i(h)$ does not increase as h increases since $Q_r(h)$ is a nested info-gap model.

By lemma 2 we know that:

$$\gamma_i[\widehat{h}(r_i, G_\mathrm{c})] = 0 \tag{92}$$

for $i = 1$ and 2. In particular:

$$\gamma_2[\widehat{h}(r_2, G_\mathrm{c})] = 0 \tag{93}$$

Hence, by eq.(91):

$$\gamma_1[\widehat{h}(r_2, G_\mathrm{c})] > 0 \tag{94}$$

Hence, by continuous upper expansion of the info-gap model:

$$\widehat{h}(r_1, G_\mathrm{c}) > \widehat{h}(r_2, G_\mathrm{c}) \tag{95}$$

which proves the lefthand inequality in eq.(16).

(2) We now prove that the lefthand inequality in eq.(16) implies the righthand inequality, assuming that $\mathcal{Q}_r(h)$ and $P(q|r)$ are coherent. The lefthand inequality of eq.(16) implies, by lemma 2 and continuous upper expansion:

$$0 = \gamma_1[\widehat{h}(r_1, G_\mathrm{c})] \geq \gamma_2[\widehat{h}(r_1, G_\mathrm{c})] \tag{96}$$

Suppose that $\gamma_2[\widehat{h}(r_1, G_\mathrm{c})] = 0$. This would imply, by continuous upper expansion, that $\widehat{h}(r_2, G_\mathrm{c}) = \widehat{h}(r_1, G_\mathrm{c})$, which contradicts lefthand inequality of eq.(16). Hence:

$$\gamma_1[\widehat{h}(r_1, G_\mathrm{c})] > \gamma_2[\widehat{h}(r_1, G_\mathrm{c})] \tag{97}$$

Thus, by coherence, $P[G^{-1}(r_1, G_\mathrm{c})|r_1] > P[G^{-1}(r_2, G_\mathrm{c})|r_2]$. Arguing as in eqs.(86)–(88) this implies the righthand side of eq.(16).

We have now completed the proof that coherence (together with the other suppositions) is sufficient for the proxy property to hold. We now prove that coherence is necessary.

Suppose that $\mathcal{Q}_r(h)$ and $P(q|r)$ are *not* coherent. We will show that eq.(16) cannot hold.

(3) Consider the implication from the righthand to the lefthand inequality in eq.(16). Arguing as in eqs.(88) and (90), the righthand inequality in eq.(16) implies:

$$P[G^{-1}(r_1, G_\mathrm{c})|r_1] > P[G^{-1}(r_2, G_\mathrm{c})|r_2] \tag{98}$$

Since $\mathcal{Q}_r(h)$ and $P(q|r)$ are not coherent (by supposition), we see from the definition of coherence that either or both of the following must hold:

$$\gamma_1[\widehat{h}(r_1, G_\mathrm{c})] \leq \gamma_2[\widehat{h}(r_1, G_\mathrm{c})] \tag{99}$$

$$\gamma_1[\widehat{h}(r_2, G_\mathrm{c})] \leq \gamma_2[\widehat{h}(r_2, G_\mathrm{c})] \tag{100}$$

Lemma 2 and eq.(99) imply that:

$$\gamma_2[\widehat{h}(r_1, G_\mathrm{c})] \geq 0 \tag{101}$$

$\gamma_i(h, G_\mathrm{c})$ decreases as h increases because $\mathcal{Q}_r(h)$ is nested. Hence eq.(101) implies that $\widehat{h}(r_1, G_\mathrm{c}) \leq \widehat{h}(r_2, G_\mathrm{c})$. This contradicts the lefthand side of eq.(16).

Similarly, lemma 2 and eq.(100) imply that $\gamma_1[\widehat{h}(r_2, G_\mathrm{c})] \leq 0$ which implies that $\widehat{h}(r_1, G_c) \leq \widehat{h}(r_2, G_c)$. This again contradicts the lefthand side of eq.(16).

In either case, eq.(99) or (100), we find that the righthand inequality in eq.(16) does not imply the lefthand inequality if $\mathcal{Q}_r(h)$ and $P(q|r)$ are not coherent.

(4) Now consider the implication from the lefthand to the righthand inequality in eq.(16). Suppose that the lefthand inequality holds. Arguing as in eq.(96) we reach eq.(97) as before. In a similar manner we conclude that:

$$\gamma_1[\widehat{h}(r_2, G_{\rm c})] > \gamma_2[\widehat{h}(r_2, G_{\rm c})] \tag{102}$$

Now, since $\mathcal{Q}_r(h)$ and $P(q|r)$ are not coherent at r_1 and r_2, we conclude from eqs.(97) and (102) that eq.(13) does not hold, so:

$$P[G^{-1}(r_1, G_{\rm c})|r_1] \le P[G^{-1}(r_2, G_{\rm c})|r_2] \tag{103}$$

which contradicts the righthand side of eq.(16).

We have now completed the proof that coherence is necessary for the proxy property to hold. ∎

C.2 Propositions 2 and 3

Proof of proposition 2. As noted earlier, coherence holds since $\mathcal{Q}(h)$ and $P(q)$ are both independent of the decision r. Hence, together with the other suppositions of the proposition, the conditions of proposition 1 prevail, and the proxy property holds. ∎

Proof of lemma 1. We will prove the lemma for upper coherence. An analogous proof applies for lower coherence.

We will prove that eq.(13) implies eq.(14) in item (2) below, and that eq.(14) implies eq.(13) in item (3). Before that we derive an explicit expression for the robustness in item (1).

(1) We derive the robustness as follows. At horizon of uncertainty h, $\theta(q, r)$ is constrained by the info-gap model of eq.(59) to the interval:

$$-h \le \theta(q, r) \le h \tag{104}$$

Hence, by monotonic increase of $\theta(q, r)$, q obeys the following constraint at horizon of uncertainty h:

$$\theta^{-1}(-h, r) \le q \le \theta^{-1}(h, r) \tag{105}$$

Using the definition of robustness in eq.(1), the monotonicity of $G(r, q)$ and $\theta(q, r)$, the definition of $G^{-1}(r, G_{\rm c})$ in eq.(11), and eq.(105) we can write the robustness as:

$$\begin{aligned}
\widehat{h}(r, G_{\rm c}) &= \max\left\{h:\ \left(\max_{q\in\mathcal{Q}_r(h)} G(r,q)\right) \le G_{\rm c}\right\} && (106)\\
&= \max\left\{h:\ \left(\max_{q\in\mathcal{Q}_r(h)} q\right) \le G^{-1}(r, G_{\rm c})\right\} && (107)\\
&= \max\left\{h:\ \theta^{-1}(h, r) \le G^{-1}(r, G_{\rm c})\right\} && (108)\\
&= \max\left\{h:\ h \le \theta[G^{-1}(r, G_{\rm c}), r]\right\} && (109)\\
&= \theta[G^{-1}(r, G_{\rm c}), r] && (110)
\end{aligned}$$

We note from eq.(105) that $q^\star(h, r) = \theta^{-1}(h, r)$. $\theta^{-1}(h, r)$ is continuous and increasing in h because $\theta(q, r)$ is continuous and increasing in q. Hence $\mathcal{Q}_r(h)$ in eq.(59) is continuously upper expanding.

(2) We now prove that eq.(13) implies eq.(14). Let $F(\theta)$ denote the probability distribution of $\theta(q, r)$, recalling that this distribution is independent of r. We can assert:

$$\begin{aligned}
P[G^{-1}(r, G_{\rm c})|r] &= {\rm Prob}[q \le G^{-1}(r, G_{\rm c})|r] && (111)\\
&= {\rm Prob}[\theta(q, r) \le \theta(G^{-1}(r, G_{\rm c}), r)|r] && (112)\\
&= F[\theta(G^{-1}(r, G_{\rm c}), r)] && (113)
\end{aligned}$$

Thus eq.(13) implies:

$$F[\theta(G^{-1}(r_i, G_{\rm c}), r_i)] > F[\theta(G^{-1}(r_j, G_{\rm c}), r_j)] \tag{114}$$

which, because $F(\cdot)$ is a probability distribution and thus monotonic, implies:

$$\theta(G^{-1}(r_i, G_c), r_i) > \theta(G^{-1}(r_j, G_c), r_j) \tag{115}$$

This, together with eq.(110), implies:

$$\widehat{h}(r_i, G_c) > \widehat{h}(r_j, G_c) \tag{116}$$

Define $\gamma(h, r) = G^{-1}(r, G_c) - q^\star(h, r)$ as in eq.(71). Since $\mathcal{Q}_r(h)$ is continuously upper expanding, we see that eq.(116), together with lemmas 2 and 3, imply both of the following:

$$\gamma[\widehat{h}(r_i, G_c), r_j] \quad < \quad \gamma[\widehat{h}(r_i, G_c), r_i] \tag{117}$$

$$\gamma[\widehat{h}(r_j, G_c), r_j] \quad < \quad \gamma[\widehat{h}(r_j, G_c), r_i] \tag{118}$$

These two relations are precisely eq.(14).

(3) We now prove that eq.(14) implies eq.(13). Eq.(14) is equivalent to eqs.(117) and (118). Hence, by continuous upper expansion of the info-gap model we conclude:

$$\widehat{h}(r_i, G_c) > \widehat{h}(r_j, G_c) \tag{119}$$

By eq.(110) this implies:

$$\theta(G^{-1}(r_i, G_c), r_i) > \theta(G^{-1}(r_j, G_c), r_j \tag{120}$$

Hence:

$$F[\theta(G^{-1}(r_i, G_c), r_i)] > F[\theta(G^{-1}(r_j, G_c), r_j)] \tag{121}$$

Therefore, from eq.(113):

$$P[G^{-1}(r_i, G_c)|r_i] > P[G^{-1}(r_j, G_c)|r_j] \tag{122}$$

which is precisely eq.(13). ▮

Proof of proposition 3. We prove the proposition for the case that $G(r, q)$ increases with q. An analogous proof applies for decreasing $G(r, q)$.

The conditions of lemma 1 hold, so $\mathcal{Q}_r(h)$ and $P(q|r)$ are upper coherent at r_1, r_2 and G_c with the system model $G(r, q)$.

It is evident that $\mathcal{Q}_r(h)$ has the property of nesting. Arguing as in the proof of lemma 1 following eq.(110), we conclude that $\mathcal{Q}_r(h)$ is continuously upper expanding. Thus the conditions of proposition 1 hold.

Hence we conclude from proposition 1 that the proxy property holds. ▮

D Coherence for the Risky Asset Example in Section 4.3

We now consider the 2-period formulation and specify a probability distribution and an info-gap model which are coherent according to definition 1, thus satisfying the conditions of proposition 1.

The performance function is $G(r, c_2)$ in eq.(19) and the robustness is defined in eq.(21) in accordance with eq.(1).

Let the payoff vector, q, be normal with mean μ and covariance matrix Σ. Thus the consumption in the second period, c_2 in eq.(18), is normal with mean $m(r) = r^T\mu$ and variance $s^2(r) = r^T\Sigma r$.

The investor does not know this probability model, and instead uses the following ellipsoidal-bound info-gap model to represent payoff uncertainty:

$$\mathcal{Q}(h) = \left\{ q : \ (q - \widetilde{q})^T W^{-1} (q - \widetilde{q}) \leq h^2 \right\}, \quad h \geq 0 \tag{123}$$

where $\widetilde{q}$ is an estimated (or guessed) payoff vector and W is a positive definite real symmetric matrix. We will characterize the infinity of choices of $\widetilde{q}$ and W which result in coherence.

One readily shows that the info-gap model for consumption in the 2nd period, $\mathcal{C}_r(h)$ in eq.(20), which is induced by $\mathcal{Q}(h)$ in eq.(123), is the following unbounded family of nested intervals:

$$\mathcal{C}_r(h) = \left\{ c_2 : \ r^T\widetilde{q} - h\sqrt{r^T W r} \le c_2 \le r^T\widetilde{q} + h\sqrt{r^T W r} \right\}, \quad h \ge 0 \tag{124}$$

In preparation for demonstrating coherence as defined in eqs.(13) and (14), we consider two investment vectors, r_i and r_j. The inverse of the performance function in eq.(19), for investment r_i, is the value of c_2 at which $G(r, c_2) = G_{\rm c}$, namely:

$$G^{-1}(r_i, G_{\rm c}) = u^{-1}(\eta_i) \tag{125}$$

where we have defined:

$$\eta_i = -\frac{G_{\rm c} + u[c_1(r_i)]}{\beta} \tag{126}$$

Analogous expressions hold for investment vector r_j.

Now, using the normality of c_2, the **first condition for coherence,** eq.(13), can be expressed as follows. Let $\Phi(\cdot)$ represent the cpd of the standard normal variable. Eq.(13) is equivalent to:

$$\Phi\left(\frac{G^{-1}(r_i, G_{\rm c}) - m(r_i)}{s(r_i)}\right) > \Phi\left(\frac{G^{-1}(r_j, G_{\rm c}) - m(r_j)}{s(r_j)}\right) \tag{127}$$

which implies:

$$\frac{G^{-1}(r_i, G_{\rm c}) - m(r_i)}{s(r_i)} > \frac{G^{-1}(r_j, G_{\rm c}) - m(r_j)}{s(r_j)} \tag{128}$$

After some algebraic manipulation one finds the following expression for the 1st condition for coherence, eq.(13):

$$u^{-1}(\eta_i) - \frac{s(r_i)}{s(r_j)} u^{-1}(\eta_j) > -\frac{s(r_i)}{s(r_j)} m(r_j) + m(r_i) \tag{129}$$

We note that this derivation applies to *any* family of probability distributions which, like the normal family, can be standardized to a parameter-free distribution ($\Phi(\cdot)$ in the normal case). Thus the current example, which demonstrates coherence between the ellipsoidal info-gap model of eq.(123) and the normal distribution, is in fact more general. We will explore the implications of standardization in section 6.

The robustness of investment vector r_i is readily found to be:

$$\widehat{h}(r_i, G_{\rm c}) = \frac{r_i^T\widetilde{q} - u^{-1}(\eta_i)}{\rho(r_i)} \tag{130}$$

where we have defined $\rho(r_i) = \sqrt{r_i^T W r_i}$. An analogous expression holds for the robustness of r_j.

The **second condition for coherence,** eq.(14), entails two relations, one for each robustness. One can show after some manipulations that these two relations are both equivalent to:

$$u^{-1}(\eta_i) - \frac{\rho(r_i)}{\rho(r_j)} u^{-1}(\eta_j) > -\frac{\rho(r_i)}{\rho(r_j)} r_j^T\widetilde{q} + r_i^T\widetilde{q} \tag{131}$$

The info-gap model and the probability distribution are coherent if eqs.(129) and (131) both hold. It is immediately evident that a *sufficient* condition for coherence is:

$$\widetilde{q} = \mu \quad \text{and} \quad W = \Sigma \tag{132}$$

since in this case $r^T\widetilde{q} = m(r)$ and $\rho(r) = s(r)$. That is, an ellipsoidal info-gap model is coherent with a normal probability distribution if the center-point and shape-matrix of the info-gap model equal the mean and covariance of the probability distribution. More importantly, it is clear that eq.(132) is *not necessary* for coherence. The center-point $\widetilde{q}$ and the shape matrix W can deviate somewhat from μ and Σ and the coherence still holds.

In short, we have demonstrated that an infinite neighborhood of info-gap models is coherent with the normal distribution. If the unknown pdf of the payoffs is normal (or any other distribution which is similarly standardized), then the proxy property will hold when the agent uses any of these ellipsoidal info-gap models to non-probabilistically represent uncertainty in the payoffs.

E Coherence for the Monetary Policy Example of Section 4.4

We now establish that the two conditions for upper coherence, eqs.(13) and (14) in definition 1, hold for a wide range of probability distributions for the monetary policy example of section 4.4.

Eq.(13), probabilistic condition. We will derive an explicit expression from eq.(13) for three classes of probability distributions of q: normal, Cauchy and gamma.

(1) *Normal distribution.* Suppose that q is normally distributed with mean μr and variance $\sigma^2 r^2$ where μ and σ are constant and where r is the interest rate, r_t, chosen by the central bank. Arguing as in eqs.(127) and (128) one finds that eq.(13) is equivalent to:

$$\frac{G^{-1}(r_i, \pi_c) - \mu r_i}{\sigma r_i} > \frac{G^{-1}(r_j, \pi_c) - \mu r_i}{\sigma r_j} \tag{133}$$

This is readily shown to be equivalent to:

$$\frac{\pi_c - \lambda y_t - (\lambda\phi + \beta)\pi_t}{r_i} > \frac{\pi_c - \lambda y_t - (\lambda\phi + \beta)\pi_t}{r_j} \tag{134}$$

If the critical inflation, π_c, satisfies the constraint in eq.(33), then eq.(134) becomes the following probabilistic condition for coherence:

$$r_i > r_j \tag{135}$$

Other families of distributions can similarly be "standardized" to lead to the same result. We will explore the concept of standardization more fundamentally in section 6.

(2) *Cauchy distribution.* As a different example, suppose that q is described by a Cauchy distribution:

$$p(q|r) = \frac{1}{\pi r[1 + (q/r)^2]}, \quad -\infty < q < \infty \tag{136}$$

This distribution is non-negative and normalized but its mean and variance are both unbounded. However, direct integration leads to:

$$P[G^{-1}(r, \pi_c)|r] = \frac{1}{2} + \frac{1}{\pi} \tan^{-1} \frac{G^{-1}(r, \pi_c)}{r} \tag{137}$$

This relation is readily used to show that the first condition for coherence, eq.(13), is equivalent to eq.(134). If π_c satisfies the constraint in eq.(33), then eq.(134) becomes eq.(135). Once again, eq.(135) is the probabilistic condition for coherence, eq.(13).

(3) *Gamma distribution.* Now suppose that q is distributed according to the gamma distribution:

$$p(q|r) = \frac{1}{n!(\beta r)^{n+1}} q^n \mathrm{e}^{-q/\beta r}, \quad 0 < q < \infty \tag{138}$$

where $\beta r > 0$ and n is a non-negative integer. This family is not standardizable in the way that the normal distribution is. However, one finds by direct integration:

$$P[G^{-1}(r, \pi_c)|r] = 1 - \mathrm{e}^{-\gamma} \left(\frac{\gamma^n}{n!} + \sum_{k=1}^{n} \frac{1}{(n-k)!} \gamma^{n-k} \right) \tag{139}$$

where $\gamma = G^{-1}(r, \pi_c)/\beta r$ which is positive since q is positive. Differentiating eq.(139) one finds $\mathrm{d}P/\mathrm{d}\gamma = \mathrm{e}^{-\gamma}\gamma^n/n!$ which is a strictly increasing function of γ. Thus the probabilistic condition for coherence, eq.(13), becomes:

$$\frac{G^{-1}(r_i, \pi_c)}{\beta r_i} > \frac{G^{-1}(r_j, \pi_c)}{\beta r_j} \tag{140}$$

If π_c is bounded as in eq.(33) then eq.(140) is equivalent to eq.(135).

In short, eq.(135) is the probabilistic condition for coherence for each of the three families of distributions we have considered.

Eq.(14), info-gap condition. We note that the info-gap model for q, eq.(30), does not depend on the chosen interest rate, r_t. Consequently the function $q^\star(h, r)$, defined in eq.(10), also does not depend on r. Hence eq.(14), for both choices of h, becomes:

$$G^{-1}(r_i, \pi_c) > G^{-1}(r_j, \pi_c) \tag{141}$$

which, after some manipulation, is seen to be identical to eq.(135).

In summary, we have shown that this info-gap model is upper coherent with a wide range of probability distributions. Hence proposition 1 holds and robustness of the choice of interest rate is a proxy for the probability of satisfying the requirement on the inflation.

F Derivations of the Principal-Agent Example in Section 4.5

F.1 Monotonicity and Scalar Uncertainty

Proposition 1 requires that the performance function be monotonic in a single scalar uncertainty. We now demonstrate how this is achieved by aggregating the uncertain vectors.

The info-gap model of eq.(35) can be more conveniently written as:

$$\mathcal{P}(h) = \left\{ p = \widetilde{p} + \pi : \ \max[-\widetilde{p}_i, \ -h] \le \pi_i \le h \ \forall i, \ \sum_{i=1}^{N} \pi_i = 0 \right\}, \quad h \ge 0 \tag{142}$$

Using the notation introduced in the info-gap models of eqs.(36) and (142), the performance function can be written:

$$\begin{aligned} G(r, u, p) &= -\sum_{i=1}^{N} (\widetilde{u}_i(r) + \eta_i)(\widetilde{p}_i + \pi_i) && (143) \\ &= \underbrace{-\sum_{i=1}^{N} \widetilde{u}_i(r)\widetilde{p}_i}_{\widetilde{G}(r)} \ \underbrace{-\sum_{i=1}^{N} \eta_i \widetilde{p}_i - \sum_{i=1}^{N} \widetilde{u}_i(r)\pi_i - \sum_{i=1}^{N} \eta_i \pi_i}_{q} && (144) \end{aligned}$$

which defines the known putative performance function $\widetilde{G}(r)$ and the uncertain scalar quantity q. We will henceforth denote the performance function as:

$$G(r, q) = \widetilde{G}(r) + q \tag{145}$$

which is monotonically increasing in the scalar uncertainty q.

An info-gap model for uncertainty in q is induced by the uncertainty in η and π:[8]

$$\mathcal{Q}_r(h) = \left\{ q = -\eta^T \widetilde{p} - \pi^T \widetilde{u}(r) - \eta^T \pi : \ \eta \in \mathcal{A}_r(h), \ \pi \in \mathcal{P}(h) \right\}, \quad h \ge 0 \tag{146}$$

The robustness in eq.(38) is now equivalently written as:

$$\widehat{h}(r, G_c) \equiv \max \left\{ h : \ \left(\max_{q \in \mathcal{Q}_r(h)} G(r, q) \right) \le G_c \right\} \tag{147}$$

[8]For the sake of clarity we are abusing our notation slightly in eq.(146) by writing $\eta \in \mathcal{A}(h)$ and $\pi \in \mathcal{P}(h)$. These are actually sets of u and p vectors, not η and π vectors.

F.2 Upper Coherence and the Proxy Property

We now establish that the two conditions for upper coherence, eqs.(13) and (14), hold for a wide range of probability distributions. Our discussion is similar to that in section E of the Appendices.

Eq.(13), probabilistic condition. We will derive an explicit expression from eq.(13) for three classes of probability distributions of q: normal, Cauchy and gamma.

(1) *Normal distribution.* Suppose that q is normally distributed with mean $\mu s(r)$ and variance $\sigma^2 s^2(r)$ where μ and σ are constant and $s(r)$ is defined in eq.(158). Arguing as in eqs.(127) and (128) one finds that eq.(13) is equivalent to:

$$\frac{G^{-1}(r_i, \pi_{\text{c}}) - \mu s(r_i)}{\sigma s(r_i)} > \frac{G^{-1}(r_j, \pi_{\text{c}}) - \mu s(r_i)}{\sigma s(r_j)} \tag{148}$$

This is readily shown to be equivalent to:

$$\frac{G_{\text{c}} - \widetilde{G}(r_i)}{s(r_i)} > \frac{G_{\text{c}} - \widetilde{G}(r_j)}{s(r_j)} \tag{149}$$

This is the probabilistic condition for coherence.

(2) *Cauchy distribution.* Suppose that q has a Cauchy distribution, eq.(136), with r replaced by $s(r)$. In the present case, direct integration leads to eq.(137) with $\tan^{-1}[G^{-1}(r, \pi_{\text{c}})/r]$ replaced by $\tan^{-1}[G^{-1}(r, G_{\text{c}})/s(r)]$. This is readily shown to be equivalent to eq.(149) which is, once again, the probabilistic condition for coherence, eq.(13).

(3) *Gamma distribution.* Now suppose that q is distributed according to the gamma distribution, eq.(138), with r replaced by $s(r)$. One finds by direct integration that $P[G^{-1}(r, G_{\text{c}})|r]$ is given by eq.(139) where now $\gamma = G^{-1}(r, G_{\text{c}})/\beta s(r)$. One readily shows that the probabilistic condition for coherence is again eq.(149).

Eq.(14), info-gap condition. From the expressions for $q^\star(h, r)$ following eq.(158) in appendix F.3 and for the robustness in eq.(159) for small G_{c}, we see that:

$$q^\star[\widehat{h}(r_i, G_{\text{c}}),\, r_j] = \frac{G_{\text{c}} - \widetilde{G}(r_i)}{s(r_i)} s(r_j) \tag{150}$$

Algebraic manipulation shows that both conditions in eq.(14) devolve to eq.(149). Thus the probabilistic and info-gap conditions for coherence are equivalent and the info-gap model is upper coherent with each of the classes of probability distributions, provided that G_{c} satisfies the conditions on eq.(159).

F.3 Deriving the Robustness

From eq.(147) we see that evaluation of the robustness requires maximizing $G(r, q)$ at horizon of uncertainty h. The uncertain performance function, eq.(144), can be re-written:

$$G(r, q) = \widetilde{G}(r) - \sum_{i=1}^{N} (\widetilde{p}_i + \pi_i)\eta_i - \sum_{i=1}^{N} \widetilde{u}_i(r)\pi_i \tag{151}$$

Since $\widetilde{p}_i + \pi_i \ge 0$, the choice of $\eta \in \mathcal{A}_r(h)$ which maximizes $G(r, q)$ is $\eta_i = -h$, regardless of how $\pi \in \mathcal{P}(h)$ is chosen. Since $\widetilde{p} + \pi$ is a normalized probability distribution, eq.(151) now becomes:

$$G(r, q) = \widetilde{G}(r) + h - \sum_{i=1}^{N} \widetilde{u}_i(r)\pi_i \tag{152}$$

We must now choose $\pi \in \mathcal{P}(h)$ to maximize $G(r, q)$ in eq.(152). We use a "Robin Hood" principle: make π_i small if $\widetilde{u}_i$ is large, and make π_i large if $\widetilde{u}_i$ is small. To do this, denote the order statistics of the $\widetilde{u}_i$'s with subscripts (i):

$$\widetilde{u}_{(1)} \ge \widetilde{u}_{(2)} \ge \cdots \ge \widetilde{u}_{(N)} \tag{153}$$

Using these same subscripts, define the following partial sum of minimal values of the $\pi_{(i)}$'s:

$$T_m = \sum_{i=1}^{m} \max[-\widetilde{p}_{(i)},\ -h] \tag{154}$$

This is the sum of the smallest (most negative) values of π_i's in $\mathcal{P}(h)$ of eq.(142) corresponding to the m largest $\widetilde{u}_i$'s. The remaining $(N-m)$ terms π_i can be chosen to be as large as h, and must be chosen to guarantee that $\sum_{i=1}^{N} \pi_i = 0$. This is achieved, and $G(r,q)$ is maximized, by choosing m so that:

$$-T_m \le (N-m)h \quad \text{and} \quad -T_{m+1} > (N-m-1)h \tag{155}$$

Using this value of m we find that $G(r,q)$ in eq.(152) is maximized by choosing π as:

$$\widehat{\pi}_{(i)}(h) = \begin{cases} \max[-\widetilde{p}_{(i)},\ -h], & i = 1,\ \ldots,\ m \\ -\dfrac{T_m}{N-m}, & i = m+1,\ \ldots,\ N \end{cases} \tag{156}$$

Using this choice of π in the performance function of eq.(152) yields the inner maximum in the definition of the robustness, eq.(147). Equating this maximum to G_{c} and solving for h yields the robustness.

An illuminating special case occurs for small G_{c}. Specifically, consider horizons of uncertainty h which satisfy:

$$h \le \underbrace{\min_{i=1,\ldots,m} \widetilde{p}_{(i)}}_{p_{\min}} \tag{157}$$

which defines $p_{\min}$. Now $\widehat{\pi}_{(i)}(h)$ in eq.(156) simplifies and we find:

$$\max_{q \in \mathcal{Q}_r(h)} G(r,q) = \widetilde{G}(r) + \underbrace{\left[1 + \sum_{i=1}^{m} \widetilde{u}_{(i)}(r) - \sum_{i=m+1}^{N} \widetilde{u}_{(i)}(r) \right]}_{s(r)} h \tag{158}$$

Define the quantity in square brackets as $s(r)$, which is necessarily positive. Note that $q^{\star}(h,r) = s(r)h$, as seen from eqs.(10) and eq.(145).

Equating the righthand side of eq.(158) to G_{c} and solving for h yields the robustness:

$$\widehat{h}(r, G_{\text{c}}) = \frac{G_{\text{c}} - \widetilde{G}(r)}{s(r)} \tag{159}$$

or zero if this is negative. This expression is valid for values of G_{c} small enough so that $\widehat{h}(r, G_{\text{c}}) \le p_{\min}$.