2012

RSMG Working Paper Series

TITLE: Bounded Awareness, Heuristics and the Precautionary Principle

Authors:

Simon Grant and John Quiggin

Working Paper: R12_4

Schools of Economics and Political Science

The University of Queensland

St Lucia

Brisbane

Australia 4072

Web: www.uq.edu.au

15 April 2012

Bounded awareness, heuristics and the Precautionary Principle*

Abstract

Abstract: The precautionary principle has been proposed as a basis for making decisions about environmental issues but remains controversial. Using a model of inductively justified propositions about awareness, this paper shows how the precautionary principle may be interpreted as a heuristic guide for boundedly rational decisionmakers faced with the possibility of unfavorable surprises.

JEL Classification: D80, D82

Key words: precautionary principle, unawareness, bounded rationality, induction

Simon Grant
School of Economics, University of Queensland, Brisbane, Australia

John Quiggin
School of Economics, University of Queensland, Brisbane, Australia

*We thank David Kelsey, Jeff Kline, Priscilla Man, Antoine Nebout, Zvi Safra and seminar participants at the University of Exeter for useful comments and suggestions.

1 Introduction

The precautionary principle, most commonly presented as a guide to environmental policy decisions, is widely used, controversial and hard to define precisely. While no exact formulation of the principle has achieved unanimous support, the precautionary principle has been widely advocated, the version adopted by the Wingspread Conference (1998) is fairly representative:

> Where an activity raises threats of harm to the environment or human health, precautionary measures should be taken even if some cause and effect relationships are not fully established scientifically. In this context the proponent of an activity, rather than the public, bears the burden of proof.

Considered as a generic rule, this formulation of the principle could be read as a simple requirement to take risks into account, even if they are unproven, a prescription that is embodied in all mainstream approaches to the theory of decision under uncertainty. This is the interpretation given by Gollier, Jullien, and Treich (2000), who focus on the option value associated with waiting for further information. Since a complete risk analysis should incorporate option values, this defence of the precautionary principle amounts to the claim that, in practice, risk analyses are usually incomplete, and that this error can be corrected by adoption of the precautionary principle.

On the other hand, as critics have pointed out, stronger formulations of the principle, appear to rule out all courses of action that may have adverse consequences, which would potentially give rise to situations where every possible action was rejected. For example, Sunstein (2003, p1003) argues that, since no course of action can be risk-free, strong versions of the precautionary principle are incoherent.

> I aim to challenge the precautionary principle here, not because it leads in bad directions, but because read for all that it is worth, it leads in no direction at all. The principle threatens to be paralyzing, forbidding regulation, inaction, and every step in between.

Sunstein goes on to claim that operationalizing the precautionary principle requires ignoring some risks and that the application of standard principles of risk analysis will yield better outcomes on average.

As these examples show, most discussion of the precuationary principle, both favorable and unfavorable, has taken place on the basis of an assumption that decisionmakers can and should take all possible contingencies into account. That is, although decisions may be subject to uncertainty, they are taken on the basis of an evaluation of all possible contingencies, and the associated outcomes of particular courses of action.

In this paper we consider a more fundamental objection to a policy of exclusive reliance on benefit–cost analysis, leading to an alternative interpretation of the precautionary principle. Given the bounded rationality of human agents, it is impossible to enumerate and consider all relevant possibilities. This point is sometimes expressed with reference to 'unknown unknowns', that is, relevant possibilities of which we are unaware (Rumsfeld, 2002). It follows that benefit–cost analysis can never encompass all relevant possibilities. For some decisions, involving courses of action similar to those undertaken on many previous occasions, and for which there exists a well-developed theoretical framework, the set of possibilities considered may be sufficiently detailed and accurate to yield an analysis that may be regarded for practical purposes as complete. For other decisions, the role of 'unknown unknowns' may be critical.

It is one thing to observe that human decisions are commonly subject to unforeseen and unconsidered contingencies. A much more difficult problem is to reason about decision problems involving such contingencies and to prescribe appropriate actions.

The first difficulty is to formulate a logical framework in which statements such as 'there exist possibilities of which I am unaware' makes sense. A large literature has now developed on this question. A variety of different approaches to the problem have been explored (Heifetz et al [2006, 2009], Halpern & Rêgo [2006, 2008, 2009, 2011]). In this paper, we will focus on the idea that induction from past experience permits decisionmakers to make judgements about whether they are, or are not, aware of all relevant contingencies in particular problems (Grant and Quiggin [2012], Walker [2012]).

An even more difficult problem is to offer guidance to decisionmakers in such situations. The central problem of decision-making with bounded awareness is evident in the terminology of 'unknown unknowns'. How can a decision-maker take account of her own bounded awareness, given that, by definition, the contingencies of which she is unaware are unknown to her? Grant and Quiggin (2012) propose the adoption of 'awareness-based heuristics', that is, decision rules that can be implemented based on the decision-makers' perception' of the game or decision problem they are facing, combined with an inductively-derived judgement as to whether there exist relevant contingencies of which they are unaware.

The central problem with such heuristics is that, unless they judge themselves to be aware of all relevant contingencies, associated with a given decision problem, decisionmakers cannot, *ex ante*, evaluate the expected payoff to the adoption of an awareness-based heuristic in a given situation. Indeed, as evaluated on the basis of a formal model of the contingencies under consideration the adoption of a non-trivial heuristic (that is, one that differs from the prescription 'choose the course of action that maximizes expected utility') can never increase welfare and may reduce it. Thus, the adoption of an awareness-based heuristic by an individual makes sense only on the basis of an inductively justified conclusion that there exist unconsidered possibilities relevant to the outcome.

The expected return to the adoption of a heuristic can be evaluated from the perspective an outside observer who is aware of all relevant contingencies. One interpretation of an awareness-based heuristic, therefore, is that it might represent advice that could be given by a fully aware external adviser to a boundedly rational decisionmaker who is capable of adopting and implementing the heuristic, but not of understanding and solving the full game (or decision problem).

An alternative approach to the evaluation of heuristics is based on the idea of 'ecological rationality', proposed by Goldstein and Gigerenzer (2002). Goldstein and Gigerenzer suggest a concept of ecological rationality for heuristics, explained as 'the capacity of the heuristic to exploit the structure of the information in natural environments'. In this paper, we will show how this concept can be applied to awareness-based heuristics.

The heuristic interpretation of the precautionary principle proposed in this paper arises when a decisionmaker is faced with a choice between alternatives, one of which leads to consequences for which the relevant elements of the state space are well-understood and the other which leads to consequences that depend to a significant extent on 'unknown unknowns'.

If most surprises are unpleasant, a risk analysis based only on known risks will underestimate the costs of choices of the second kind. That is, standard risk analysis leads to a bias in favour of taking chances on poorly-understood risks. The precautionary principle may be seen as a rule designed to offset such biases. Our primary aim in this paper is to show how the precautionary principle may be understood as an ecologically rationally heuristic constraint on decisions and strategies for individuals who are unaware of relevant contingencies and understand this to be the case.

The paper is organized as follows. In Section 2, we develop a formal model of reasoning and decision-making for an individual who is, in general, unaware of some relevant possibilities, but understands that they may become aware of new possibilities over time. This model is derived from that of Grant and Quiggin (2012), simplified to apply to the special case of individual decisions. Readers interested in a more detailed exposition of the modal-logical and game-theoretic aspects of the model are referred to Grant and Quiggin (2012). On the other hand, readers who are primarily interested in the application of the model to the precautionary principle may prefer to go directly to Section 3, which presents an extended example of an environmental decision problem involving unawareness of relevant contingencies.

The central contribution of the paper is Section 4, where we present two versions of the Precautionary Principle, developed with reference to the example of Section 3, and formalized using the model of Section 2. The Strong Form of the principle prescribes avoidance of heuristics leading to behavioral rules that are subject to unfavorable surprises. We show that the Strong Form of the Precautionary Principle may lead to excessive conservatism in some cases and proposed a Modified Form which allows for behavioral rules subject to unfavorable surprises, provided they incorporate a fallback option that limits the associated damage. We characterize conditions for the two forms of the Precautionary Principle to be ecologically rational relative to each other and relative to the behavioral rule that requires decisionmakers to maximize their best estimate of expected utility, given their awareness at any point in the decision process.

In Section 5, we provide an extension of the example to the case of interactive decisions, such as those involving the proponent of an activity and a regulator who must approve or reject it.[1] We argue that, in the regulatory context, the precautionary principle may be understood as a burden-of-proof requirement on project proponents, consistent with the Wingspread statement cited above.

Finally, we offer some concluding comments.

2 Decision-making and reasoning with awareness

The central idea of the present paper is that, if players in a game, or a decision-maker in a (single-player) game against nature, understand that they are not fully aware of the possibilities relevant to their choice of actions, it may be ecologically rational for them to adopt heuristics such as the precautionary principle. This idea may be represented formally using models in which awareness changes over time. Grant and Quiggin (2012) model an extensive-form game with awareness, using notation derived from Osborne and Rubinstein (1994). In this paper we present a simplified single-player version of their model.

The possible positions for a decision-maker (DM) in a (single-player) game against nature, Γ, are represented by the set of histories, denoted by H and defined to be sequences of actions $\langle \alpha_0, \alpha_1, \alpha_2, ..., \alpha_k \rangle$, where (without essential loss of generality) α_0 is taken to be 'chosen' by nature and $\langle \alpha_1, \alpha_2, ..., \alpha_k \rangle$ are actions chosen by the DM. The initial (null) history is denoted by $<>$. The extensive-form structure of the game Γ is embodied in the sub-history relationship $\langle \alpha_0, \alpha_1, ..., \alpha_k \rangle \preceq \langle \alpha_0, \alpha_1, \alpha_2, ..., \alpha_k, \alpha_{k+1}, .., \alpha_l \rangle$. If $h = \langle \alpha_1, \alpha_2, ..., \alpha_k \rangle$, then $h \bullet \alpha_{k+1} = \langle \alpha_1, \alpha_2, ..., \alpha_k, \alpha_{k+1} \rangle$ and we observe that $h \preceq h \bullet \alpha_{k+1}$. The set of terminal histories is denoted Z, and associated with Z is the payoff function $v : Z \to \mathbb{R}$. For the initial history $<>$, $A(<>)$ denotes the set of actions available to nature and for all other non-terminal histories h, $A(h)$ is the set of actions available to the DM at h. Formally, for each $h \in H \backslash Z$, $A(h) = \{\alpha : h \bullet \alpha \in H\}$.

The information structure of the game is determined by the information set assignment function, $\mathcal{I} : H \backslash \{<>\} \to 2^H \backslash \varnothing$ where $\mathcal{I}(h)$ is the set of histories considered possible by the DM at h.[2] The initial move by nature is governed by the chance assignment f^c.

[1] We do not develop a fully-specified model of heuristics for interactive awareness (for this, see Grant and Quiggin [2012]), since this requires, in general, modelling of the awareness imputed by one player to others. In the regulatory context of the example, such issues are of secondary importance.

[2] Throughout we shall assume the DM satisfies perfect recall. In this single-player setting, this corresponds to

A *behavioral rule* r for the DM is a collection of actions $(r(h) : h \in H\backslash(\{<>\} \cup Z))$ where each $r(h)$ is an element of $A(h)$, the actions available to the DM at her information set $\mathcal{I}(h)$.[3]

Given a rule r, we denote by $\tau(r)$ the probability distribution over terminal histories induced by r and the chance assignment f^c of nature. Hence the value associated with the rule r is given by:

$$V(r) = \sum_{h \in Z} \tau(r)[h]\, v(h). \tag{1}$$

Following Grant and Quiggin (2012) we shall represent unawareness by imputing to the DM at h awareness of a game $\tilde{\Gamma}$ that is a *restriction* of the full awareness game Γ. The restriction is obtained by deleting some of the terminal histories of the original game and then constructing the restricted game to be consistent with the original game in terms of the subhistory ordering and information structure.

More formally, we introduce the idea of a perception mapping.

Definition 1 *Fix a game against nature Γ. For each history h in H, $\tilde{Z}(h)$ is the set of terminal histories considered by the DM and so we impute to the DM awareness of the game against nature $\Gamma_{\tilde{Z}(h)}$, the unique game that has terminal histories $\tilde{Z}(h)$ and is consistent with the original game in terms of subhistory ordering and information structure.*

The set of histories in $\Gamma_{\tilde{Z}(h)}$ is defined by $H_{\tilde{Z}(h)} = \{h' \in H : \exists z \in \tilde{Z}(h) \text{ s.t. } h' \preceq z\}$ and the valuation $v_{\tilde{Z}(h)}$ is the restriction of v to $\tilde{Z}(h)$. Moreover, the chance assignment of nature $f^c_{\tilde{Z}(h)}$ in the restricted game is taken to be the conditional update of f^c. Finally, at any non-initial history h in the full game, $\mathcal{I}_{\tilde{Z}(h)}(h) (= \mathcal{I}_{\tilde{Z}(h)}(h) \cap H_{\tilde{Z}(h)})$ shall denote the information set of the game $\Gamma_{\tilde{Z}(h)}$ the DM perceives herself to be in, with the set of actions she perceives to be available given by $A_{\tilde{Z}(h)} = \{\alpha : \exists h' \in \mathcal{I}_{\tilde{Z}(h)}(h), \text{ s.t. } h' \bullet \alpha \in H\}$.

Grant and Quiggin (2012) impose the following requirements on the perception mapping.

IN: (information neutrality): For all h, h' in $H\backslash\{<>\}$, $h' \in \mathcal{I}(h) \Rightarrow \tilde{Z}(h') = \tilde{Z}(h)$.

NI: (No impossibility): For all h, $\mathcal{I}_{\tilde{Z}(h)}(h) \neq \varnothing$.

IA: (Increasing awareness): If $h \preceq h'$ then $\tilde{Z}(h) \subseteq \tilde{Z}(h')$.

The first property requires the DM's level of awareness be congruent with the information structure of the full awareness game. This is in line with the usual notion that histories that reside in the same information set are indistinguishable to the player who is to select an action

the requirement that if any pair of histories h and h' are in the same information set then they can only differ in their initial action (that is, in the choice made by nature at the initial history $<>$) .

[3] If the DM is fully aware of the game Γ that she is playing against nature, then the behavioral rule corresponds to what is normally referred to as the DM's 'strategy' or 'plan-of-action'. In the setting we consider below, however, the DM may not be fully aware of the game Γ that she is actually playing. Although she may adopt a heuristic that will determine her choice at ever information set she may encounter in the game Γ, in general she will not have access to the behavioral rule generated by the heuristic. Hence we feel it inappropriate to refer to such a behavioral rule as a strategy or plan-of-action.

in that information set. The second requires that the DM always considers some history possible. The third, which is the most important for the purposes of this paper, means that at any point in the game, the DM will, in general, be aware of possibilities she had not considered earlier in the game.

The idea of a perception mapping allows for the following definition of a *(single-player) game with awareness.*

Definition 2 *A game with awareness $\mathcal{G}$ is characterized by the tuple $\left(\Gamma, \tilde{Z}(\cdot)\right)$ where Γ is the full (maximally aware) extensive form game against nature the DM is actually playing and $\tilde{Z}(\cdot)$ is the associated perception mapping.*

For each history h' in $\mathcal{I}_{\tilde{Z}(h)}(h)$, let $\mu_h[h']$ denote the probability the DM assigns to being at history h' conditional on being at information set $\mathcal{I}_{\tilde{Z}(h)}(h)$ in the game $\Gamma_{\tilde{Z}(h)}$. Let β_h denote a *continuation (pure) strategy* in the game $\Gamma_{\tilde{Z}(h)}$ starting at the information set $\mathcal{I}_{\tilde{Z}(h)}(h)$ and for each $h' \in \mathcal{I}_{\tilde{Z}(h)}(h)$, let $z(\beta_h|h')$ be the terminal history in $\tilde{Z}(h)$ reached by the play of the continuation strategy β_h starting from history h'. Hence the DM who perceives herself to be at the information set $\mathcal{I}_{\tilde{Z}(h)}(h)$ in the game $\Gamma_{\tilde{Z}(h)}$, evaluates the expected payoff from playing β_h to be:

$$V_h(\beta_h) = \sum_{h' \in \mathcal{I}_{\tilde{Z}(h)}(h)} \mu_h[h'] \, v^i(z(\beta_h|h')) \,. \tag{2}$$

Grant and Quiggin (2012) argue that, although players cannot foresee possibilities of which they are unaware, they can reason inductively about the existence of such possibilities. To illustrate how, consider a game with awareness in which at a particular history, the DM only has access to some restriction of the full game. From the perspective of the full game, we say the DM believes a proposition to be true at that history, if it is true for all histories she considers possible in the restricted game she has access to at that history. Similarly, she believes a proposition to be false at that history, if it is false for all histories she considers possible in the restricted game. Building on those definitions we can also define what it means for the DM to believe a proposition was true (respectively, was false). A little more formally, we have:

Definition 3 *Fix a game with awareness $\mathcal{G} = \left(\Gamma, \tilde{Z}(\cdot)\right)$. At h, the DM:*

(*i*) 'believes *the proposition p is true (respectively, false)' if p is true (false) for all histories in $\mathcal{I}_{\tilde{Z}(h)}(h)$;*

(*ii*) 'believes *the proposition p* was *true (false) in the past' if p was true (false) at some previous information set in the play of the game.*

Remark 1 *Since it is possible that the DM is not aware of the history h in which it just happens p is false, the belief of the DM that a proposition is true does not necessarily entail it actually is* *(unlike the 'knowledge' operator in standard modal-logical frameworks). That is, in the restricted game a DM may hold 'false' beliefs.*

To avoid awkward contortions in our syntax when we discuss inductive reasoning, at a history h in the full game, we shall take it that the DM '*does not believe* p *is true*' if p is not true for all histories in $\mathcal{I}_{\tilde{Z}(h)}(h)$. Similarly, at the history h in the full game, we shall take it that the DM '*does not believe* p *was false in the past*,' if p was never false at some previous information set in the play of the game.'

To consider how a DM might employ his beliefs about propositions to reason inductively, recall the principle of historical (or temporal) induction states that if a proposition has been found to be true in (many) past instances and has never found to be false in any past instance, this provides support for the belief that it will hold true in the future. For example, the fact that the proposition 'the sun will rise tomorrow' was true yesterday, the day before and the day before that and so on, provides inductive support for the belief that the same proposition is true today. From the perspective of the current situation, however, the DM may not be able to ascertain the truth or falsity of some propositions at each past instance. But if it is the case whenever the truth or falsity of a proposition in a past instance can be gleaned from the perspective of the DM's awareness today, the proposition was found to be true, we contend that also provides inductive support for the belief that it will hold true in the future. That is, we simply require the proposition was true at some instance in the past and was never false at any instance in the past to allow the DM to regard the propositon as supported by historical induction. Again, a little more formally, we have:

Definition 4 *Fix a game with awareness* $\mathcal{G}$*. At* h*, the DM* 'regards *the proposition* p *as* supported by historical induction*' if she* 'believes p was *true in the past' and she 'does not believe* p *was false in the past.'*

Since the perception mapping exhibits **IA** (increasing awareness), it follows that if the DM has become aware of histories she had previously not considered then she believes the proposition 'there exist possibilities of which I am unaware' was true in the past and she does not believe it was ever false in the past, thus making it supported by historical induction. On the other hand, the DM may reason, on the basis of past experience, that the consequences of some actions are well enough understood that the perceived game $\Gamma_{\tilde{Z}(h)}$ encompasses all possibilities relevant to the outcome of these actions. As a result, a DM may consider two classes of actions. The first class are those for which the outcome may be affected by possibilities of which she is unaware, while the second are those for which the possible outcomes are fully described by the game of which she is aware.

Grant and Quiggin (2012) suggest that such a DM's choices need not induce the behavioral rule that maximizes her expected payoff in the game $\Gamma_{\tilde{Z}(h)}$ that she has access to at h. Instead, they propose the adoption of 'awareness-based heuristics'. Such heuristics are defined as admissibility rules for actions, that can be implemented based on the DM's perception of the game or decision problem she is facing, combined with an inductively-derived judgement as to whether there exist relevant contingencies of which she is unaware. And again, a little more formally, we say:

Definition 5 *Fix a game with awareness $\mathcal{G}$. An awareness-based heuristic $\mathcal{H}$ for the game $\mathcal{G}$ is an admissibilty rule such that, for each h in $H\backslash(\{<>\}\cup Z)$, $\mathcal{H}(h)$ is a subset of the continuation strategies available in the game $\Gamma_{\tilde{Z}(h)}$ from the information set $\mathcal{I}_{\tilde{Z}(h)}(h)$. The restriction $\mathcal{H}(h)$ depends only on the perceived game $\Gamma_{\tilde{Z}(h)}$, and on the judgement of the DM at h as to whether she is aware of all relevant possibilities. Moreover, $\mathcal{H}$ induces the behavioral rule r, given by: for each h*

$$r(h) = \beta_h\left(\mathcal{I}_{\tilde{Z}(h)}(h)\right), \text{ for some } \beta_h \in \underset{\hat{\beta}\in\mathcal{H}(h)}{\operatorname{argmax}} V_h^i\left(\hat{\beta}\right).$$

In a game with awareness $\mathcal{G}$, an awareness-based heuristic $\mathcal{H}$ constrains the choice by the DM over actions at every non-initial and non-terminal history in the full game Γ. As a baseline for comparison we consider the case of the (naive) Bayesian heuristic, which we will denote by $\mathcal{H}_0$, where $\mathcal{H}_0(h)$ is the set of all continuation strategies in the game $\Gamma_{\tilde{Z}(h)}$ from the information set $\mathcal{I}_{\tilde{Z}(h)}(h)$. The Bayesian heuristic prescribes at each h, a choice from the set $A_{\tilde{Z}(h)}(h)$ that forms part of a strategy for the game $\Gamma_{Z(h)}$ that yields the maximum expected payoff in the continuation of $\Gamma_{Z(h)}$ from the information set $\mathcal{I}_{\tilde{Z}(h)}(h)$. Other heuristics (including the precautionary principles discussed in this paper) may preclude the adoption of particular strategies or require that strategies satisfy some admissibility condition.

Crucially, provided individuals can make judgements about their own awareness, they can implement such heuristics *without awareness* of the full game Γ. On the other hand, unless they judge themselves to be aware of all relevant possibilities, they cannot completely describe the choices implied by the rule r or evaluate $V(r)$, the expected payoff to r in the full game Γ. Nor can they rely solely on the evaluation available to them in the restricted game they perceive. Indeed, as evaluated in the game $\Gamma_{\tilde{Z}(h)}$, the adoption of a heuristic other than the Bayesian heuristic cannot increase the value of the continuation of the game, and will, in general, reduce it. Thus, the adoption of a heuristic by an individual makes sense only on the basis of an inductively justified conclusion that there exist unconsidered possibilities relevant to the outcome.

The expected payoff resulting from the adoption of a heuristic can be evaluated from the perspective of the full game Γ, even though this perspective is not available to the DM. One interpretation of a heuristic, therefore, is that it might represent advice that could be given by a fully aware external adviser to a boundedly rational DM who is capable of adopting and implementing the heuristic, but not of understanding and solving the full game (or decision problem).

As we noted in the introduction, a closely related approach to the evaluation of heuristics is based on the idea of 'ecological rationality', proposed by Goldstein and Gigerenzer (2002). In a game with awareness, the concept of ecological rationality may be seen as the way a heuristic adopted by a DM with a model characterized by bounded awareness would be evaluated from the perspective of a more aware player, or an external modeller aware of the full game. Moreover, if the game with awareness exhibits IA, the DM may evaluate the ecological rationality of her own heuristics, applied early in the game, from the more aware perspective available to her later in the game.

Adapting the definition of ecological rationality adopted by Grant and Quiggin (2012) yields in the simplified version of their model developed above, the following.

Definition 6 *In a game with awareness $\mathcal{G}$, a heuristic $\mathcal{H}$ that induces the behavioral rule r is ecologically rational if the associated expected payoff $V(r)$ is greater than the expected payoff $V(r_0)$ of the behavioral rule r_0 associated with the Bayesian heuristic $\mathcal{H}_0$. More generally, a heuristic $\mathcal{H}$ that induces the behavioral rule r is ecologically rational (in $\mathcal{G}$) relative to an alternative heuristic $\mathcal{H}'$ that induces the behavioral rule r' if $V(r) \geq V(r')$.*

Our definition above of ecological rationality is cast from the perspective of an unboundedly rational external modeller with access to the full game. Perhaps more relevantly from the perspective of actual actors, the belief that a particular heuristic is ecologically rational in a given setting may be justified inductively on the basis of its past performance in similar settings.

In the next section, we illustrate and expand on the ideas and concepts introduced above by setting up and analyzing a game with awareness which suggests a way to operationalize the precautionary principle as a decision heuristic.

3 An extended example

To illustrate the type of reasoning we introduced in the preceding section we present the following stylized decision problem. A boundedly rational DM must choose whether to undertake a project or to maintain the status quo. Given the state of her current knowledge about and understanding of the situation, the project is expected to yield net economic benefits compared to maintaining the status quo. In the absence of more detailed consideration, the DM is not aware of any environmental hazards that may be associated with the project that could inflict high damage. Before making the decision, however, there is the option of undertaking a detailed, and costly, study which will reveal such hazards if they exist.

From the perspective of an unboundedly rational external observer, it is possible to assign a probability to the event that a high-damage hazard will be discovered, and an expected (dis)utility to the associated loss. However, until and unless the study is undertaken, the awareness of the DM is much more limited. We will assume that she is conscious of the possibility that a high-damage hazard exists. Furthermore, once she is made aware of a previously unknown high-damage hazard, she has available to her a 'fall-back' action that bounds from below the utility associated with the resulting adverse event.

The structure of the problem, while stylized, is rich enough to capture essential aspects of the typical settings in which the precautionary principle may be applied.

In figure 1 we represent the problem as a single-player extensive-form game with Nature, denoted Γ. Nature moves first, and randomly chooses one of four actions: a_H or 'high damage' with probability p_H; a_L or 'low damage' with probability p_L; and two 'no damage' actions a_1 and a_2, with probabilities p_1 and p_2, respectively.

The second action is taken by the DM who chooses whether or not to undertake a study of the problem. The action 'study' is denoted a_S and 'no study' is denoted a_N. The study yields a signal regarding Nature's action, either 'high-damage hazard' $\{a_H, a_1\}$ or 'low-damage hazard' $\{a_L, a_2\}$. The third action, also undertaken by the DM is to approve (a_A) or reject (a_R) the project. Nature's action is then revealed and the game ends.

Following the notation introduced in section 2, we represent a history h as a sequence of actions. So, for example, the sequence of actions 'nature chooses high damage and the DM chooses study' is denoted by $\langle a_H, a_S \rangle$. Associated with any terminal history, such as $z = \langle a_H, a_S, a_R \rangle$ is a payoff $v(z)$ for the DM. The payoff to the DM is the sum of three terms: a benefit B, received if the project is approved, a cost C incurred if the study is undertaken, and damage H, L or 0, determined by nature's choice and incurred if the project is approved.

The information set $\mathcal{I}(h)$ available to the DM at history h is represented in the game-tree by the dotted oval surrounding the node at the end of history h. A behavioral rule r for the DM is a rule prescribing an element from the set of available actions at each of the non-initial and non-terminal information sets. Following the notation introduced in section 2, $V(r)$ equals the expected payoff to the DM from playing the behavioral rule r.

<INSERT FIGURE 1 AROUND HERE>

Incomplete awareness is modelled by the assumption that the DM initially represents the problem by the subjective game structure Γ', shown in figure 2, which does not include the possibility that nature will choose action a_H 'high damage'. We further assume unawareness of the 'no damage' action a_1, and, for analytical convenience, scale the probabilities p_L and p_2 so as to add to 1. That is, we set p_D so that:

$$\frac{p_D}{1 - p_D} = \frac{p_H + p_L}{p_1 + p_2}$$

Hence at information set $\{a_H, a_1, a_2, a_L\}$ in Γ, deciding whether to undertake the study, the perceived game Γ' represents the DM's unawareness of the possibility that nature may choose a_H or a_1. Thus, the DM is aware only of histories beginning with the actions a_2 or a_L.

<INSERT FIGURE 2 AROUND HERE>

If the DM does not undertake the study, she moves to information set $\{\langle a_H, a_{NS}\rangle, \langle a_1, a_{NS}\rangle, \langle a_2, a_{NS}\rangle, \langle a_L, a_{NS}\rangle\}$, and her subjective perspective remains the game Γ'. Only at the terminal history $\langle a_H, a_{NS}, a_A\rangle$ where high damage is realised, does the DM become aware of this possibility, and by then it is too late to take any further action.

If the DM undertakes the study, and receives the signal 'high-damage hazard' she becomes, and remains thereafter, aware of the full game Γ and that the cost of damage (if it occurs) will be H rather than, as previously assumed, L. Under the simplifying assumption

$$\frac{p_H}{p_1} = \frac{p_L}{p_2} = \frac{p_D}{1 - p_D}$$

the study yields no information about the probability of damage occurring.

Finally, consider the case when the DM undertakes the study and receives the signal 'low-damage hazard'. As far as the outcome of the game is concerned, it does not matter whether she becomes aware of the unrealised possible actions a_H or a_1. However, for subsequent exposition, we will assume that the DM becomes aware of the full game Γ whenever action a_S 'study' is chosen at her first information set.

4 Two Precautionary Principles as Guides for Choice.

Given this setup, the question that naturally arises is: how should the DM proceed?

Some insight may be gained by considering the position of a fully aware DM whose model of the game is Γ. Conditional on not undertaking the study, the expected payoff from approving the project is $B - p_H H - p_L L$ while the expected payoff from rejecting the project is zero, so the project will be approved if and only if $B - p_H H - p_L L > 0$. Conditional on undertaking the study, the expected payoff from approving the project is $B - p_D H - C$, if the study yields the signal 'high-damage hazard' leading to the information set $\{\langle a_H, a_S\rangle, \langle a_1, a_S\rangle\}$ and $B - p_D L - C$ if the study yields the signal 'low-damage hazard' leading to the information set $\{\langle a_L, a_S\rangle, \langle a_2, a_S\rangle\}$. To make the study instrumentally relevant we assume $B - p_H H - p_L L < 0 < B - p_D L$. Since $B - p_D H < B - p_H H - p_L L$, it thus follows that a DM whose model of the game is Γ will reject the project if no study is undertaken or if the study is undertaken and yields the signal 'high-damage hazard' and will only approve the project if the study is undertaken and yields the signal 'low-damage hazard'.

With these assumptions, the information value of the study may be seen to be $(p_2 + p_L)(B - p_D L)$, so the DM will undertake the study if and only if $(p_2 + p_L)(B - p_D L) > C$. We assume this condition is satisfied, so that the behavioral rule r^* - "undertake the study, approve the project if study undertaken and 'low-damage hazard' signal received , otherwise reject the project", maximizes

$V(r)$, among all possible behavioral rules and so constitutes an *optimal strategy or plan-of-action* for the fully aware DM.

Next consider the position of a naive DM whose model of the game is Γ' as shown in figure 2, and who does not take any account of the possibility that her model may be incomplete and so employs the Bayesian heuristic $\mathcal{H}_0$ discussed in section 2.

Conditional on not undertaking the study, that is, at the information set, $\{\langle a_L, a_{NS}\rangle, \langle a_2, a_{NS}\rangle\}$ the expected payoff from approval is $B - p_D L$, while the payoff from rejection is zero, so the proposal will be approved if and only if $B - p_D L > 0$. Conditional on undertaking the study, that is, reaching the information set, $\{\langle a_L, a_S\rangle, \langle a_2, a_S\rangle\}$ the expected return from approval is $B - p_D L - C$, while the return from rejection is $-C$, so once again the proposal will be approved if and only if $B - p_D L > 0$. In either case, the expected payoff is lower by C if 'study' is chosen at stage 1 (that is, the information value of the study is zero), so strategies involving 'study' are not optimal in the game Γ' and so 'study' will not be chosen by a DM following the heuristic $\mathcal{H}_0$.

Now consider a boundedly rational, but sophisticated DM. Such a DM will be conscious of her own limited awareness and of the possibility that there may exist hazards she has not considered that would cause her, *ex post*, to regret her decisions. The problem we face is how to represent this consciousness.

As outlined in section 2 above, Grant and Quiggin (2012) propose an approach in which the deductive analysis of decision theory is constrained by heuristics derived from inductive reasoning. In particular, given past experience of decisions that have turned out badly because of unconsidered possibilities, the DM may regard the proposition 'there may exist hazards I have not considered' as being *supported by historical induction.* In the simplified presentation here, we do not describe explicitly the history on which this inductive inference is based. Informally, we assume that the decision being modelled is one of a class for which the DM has some past experience, including experience of unfavorable surprises, as defined below.

We represent the DM's reasoning about propositions of which she is unaware in figure 3. The part of the game of which the DM is aware is presented in black and is the same as in figure 2. The upper part of the diagram, in red, represents the DM's consciousness that there may be hazards of which she is not aware. It consists of incompletely specified histories involving actions by nature that do not belong to the set $\{a_1, a_L\}$ and the incompletely specified payoffs associated with those histories. For histories where the DM chooses a_R 'reject' at her second or third information sets, the payoff is independent of actions by nature and is therefore specified in the diagram as $-C$ (if a_S 'study' was chosen at her first information set), or 0 (if a_{NS} 'no study' was chosen at her first information set). The payoffs from histories where the decisionmaker chooses a_A following an action choice by nature that does not belong to the set $\{a_1, a_L\}$ are unspecified, but the DM understands that the damage associated with such histories may exceed L.

<INSERT FIGURE 3 AROUND HERE>

From the perspective available to someone (say, a modeller, or the DM following a study) who has access to the full game Γ, the most important history omitted from Γ' is $\langle a_H, a_{NS}, a_A \rangle$. We will refer to a history of this kind, involving an unconsidered action by nature as a 'surprise', and say that any heuristic governing the DM 's choice that leads her to choose a_{NS} first and then always to choose a_A is '*subject to surprises*'. Hence for this example, the behavioral rule r_0 associated with the heuristic $\mathcal{H}_0$ subjects the DM to the possibility of being surprised.

Surprises may be favorable or unfavorable. In problems where the precautionary principle is considered, inductive reasoning will normally justify the proposition that the choices resulting from the application of a particular heuristic are subject to unfavorable surprises. To make this notion more precise, we assume that the restricted game Γ' includes a 'status quo' or 'secure' (behavioral) rule which is not subject to surprises. In the example, the natural choice is the behavioral rule 'no study, always reject' in which a_{NS} is chosen at her first information set and a_R is always chosen at all the other information sets controlled by the DM. We refer to the expected payoff from the 'secure' rule as the 'security level' for the restricted game, and adopt the convention (applicable to the example) that the security level is zero.[4]

We say the strategy β is *'subject to unfavorable surprises'* if the proposition 'there exist unconsidered moves by nature, against which the payoff from the behavioral rule r associated with the strategy β is less than zero' is *supported by historical induction.*

We consider the following *Strong Form of the Precautionary Principle:*

> '*In games with awareness for which there exists a secure rule, exclude any strategy that is subject to unfavorable surprises.*'

The following result is immediate from the definition of the ecological rationality of a heuristic and the expected payoff of the secure rule in the full game being normalized to zero, so we state it without proof.

Proposition 1 *Fix a game with awareness $\mathcal{G}$. Suppose the DM is initially only aware of the game Γ' which is a proper restriction of the full game Γ but there exists a secure rule $\bar{r}$, with $V(\bar{r}) = 0$. If the existence of unfavorable surprises is supported by historical induction, then the Strong Version of the Precautionary Principle is ecologically rational if and only if $V(r_0) < 0$, where r_0 is the behavioral rule associated with the Bayesian heuristic $\mathcal{H}_0$.*

[4] In this example, the payoff from the 'secure' rule is risk-free. In general, however, secure rules may be subject to risk arising from moves by nature that are explicitly represented in the restricted game Γ', and to which probabilities have been assigned.

The existence of a secure rule is essential to a coherent presentation of the Precautionary Principle. In the absence of such a rule it might well happen, as critics of the Precautionary Principle have suggested, that no decision is consistent with the Precautionary Principle. Recall that the Bayesian heuristic $\mathcal{H}_0$ results in the individual choosing the behavioral rule r_0 'no study, always approve'. This is subject to unfavorable surprises whenever the proposition 'there exist unconsidered acts of nature, with conditional expected cost greater (in absolute value) than the project value B' is supported by historical induction.

From the viewpoint of a fully aware external modeller, the proposition is true whenever $B - p_D H < 0$. Since the decisionmaker is not aware of the nature of the surprise, she cannot apply this test. Rather, she reasons inductively to the conclusion: the Bayesian heuristic $\mathcal{H}_0$ leads to the adoption of the behavioral rule r_0 that is subject to unfavorable surprises.

Perhaps less obviously, any strategy involving a choice of 'study' at the first information set is also subject to unfavorable surprises. At best, this leads to the behaviorial rule 'study, approve if no surprises, disapprove otherwise' in which case, if the study yields the possibility of 'high-damage hazard' leading to a decision not to proceed, the outcome will be $-C < 0$. Hence, the proposed interpretation of the Strong Form of the Precautionary Principle leads to the adoption of the secure rule, 'no study, always reject' with payoff zero. The application of the strong form of the precautionary principle to our example is illustrated in figure 4 with the choices corresponding to the decision maker adopting the secure rule appearing in blue.

<INSERT FIGURE 4 AROUND HERE>

The Strong Form of the Precautionary Principle ensures that the DM avoids unfavorable surprises whenever there exists a secure rule. However, as the example shows, in many circumstances, this form of the Principle may be too strong.

Consider the case of the decision not to undertake a study that would increase awareness. This case suggests that the Strong Form of the Precautionary Principle leads to excessive conservatism. Moreover, given that acceptance of the Precautionary Principle is motivated by the intuition that we should act cautiously in making decisions where the consequences are not fully understood, it appears somewhat perverse to rule out actions such as 'study' in the example. It is therefore, desirable to consider modifications of the Strong Form of the Precautionary Principle that yield more satisfactory implications.

In a typical problem of the kind considered in this example, we would normally expect that the cost of a study should be less than the expected benefit from the project, that is, $C < B - p_D L - C$. If the odds of unfavorable surprises are less than even then it seems consistent with the intuition behind the Precautionary Principle to propose a heuristic that leads to the behavioral rule, 'undertake the study, and choose whatever action appears optimal in light of the results'.

However, the reasoning leading to a notion of inductive justification does not provide a basis for well-defined subjective probabilities, comparable to those of a Bayesian decision-maker updating prior beliefs in the light of observed evidence. Hence, a decision maker cannot compute a net expected positive payoff. Rather, as discussed in Grant and Quiggin (2012), willingness to undertake a relatively low-cost study must be treated as a heuristic.

This modification of the precautionary principle applied to our example is illustrated in figure 5 with the choices corresponding to the DM following this strategy appearing in blue. At her first information set the DM chooses a_S (that is, she undertakes the study) with the intention of choosing a_A (that is, to approve the project) at her next information set in the game she perceives to be playing. In the event of low-damage hazard, she does indeed choose a_A and reaps a(n expected) payoff $B - p_D L - C > 0$ which is better than what she could have guaranteed herself with the secure rule. On the other-hand, if the study makes her aware of a high-damage hazard that she previously was not aware of, she has an opt-out (that is, fall-back) option to choose a_R (that is, she does not approve the project) thereby incurring a payoff of $-C < 0$. This is a worse outcome than she could have guaranteed herself by employing the secure rule but at least is bounded and *known* beforehand.

<INSERT FIGURE 5 AROUND HERE>

More generally, suppose the DM who has access to the game Γ' has available the choice of a strategy β' in Γ' which yields an expected payoff in Γ' of $V_{\Gamma'} > 0$ (that is, an expected payoff in the game Γ' that is greater than the payoff of the strategy corresponding to following the secure rule $\bar{r}$). Let r' denote the behavioral rule associated with this strategy. Furthermore, suppose she has access to a 'fallback option', that results in the behavioral rule r'' that is a modification of the behavioral rule r'. Suppose, in the event there are no surprises, following the rule r'' entails adopting the strategy β'' in Γ' that is costly relative to β' (the strategy adopted if following the behavioral rule r'). In the event there is a surprise, however, the decision maker can switch to a fall-back option β''' in Γ' thereby implementing a behavioral rule r'' that admits a *quantifiable upper bound on the expected cost that is less than or equal to* the expected benefit of r' relative to the secure rule $\bar{r}$ in the no-surprise event. More formally, suppose the rule r'' (the modification of the rule r') yields in the event there are no surprises an expected payoff (in the game Γ') equal to $V_{\Gamma'} - c^n > 0$, and yields in the event there are unfavorable surprises, an expected cost relative to the secure rule that is bounded above by c^s. We then designate the behavioral rule r'' to be a *fallback option* for r' if $V_{\Gamma'} - c^n > c^s$.

Relating this back to our example, where for the restricted game Γ', β' is the strategy 'no study, always approve', β'' is the strategy 'study, always approve', β''' is the strategy 'study, always disapprove', and for the full awareness game Γ, r' is the behavioral rule 'no study, always approve' and r'' is the behavioral rule 'study, approve if study and low-hazard signal, otherwise do not approve'. Furthermore, $V_{\Gamma'} = B - p_D L$, $c^n = c^s = C$. Application of the fallback option is illustrated in figure 5 with the resulting choices undertaken by the DM appearing in blue.

Hence we express the following *Modified Form of the Precautionary Principle:*

> *'In games with awareness for which there exists a secure rule, exclude any strategy that is subject to unfavorable surprises unless the strategy can be modified to incorporate a fallback option.'*

The next result formalizes our original intuition above that the Modified Form of the Precautionary Principal would be recommended in preference to the Strong Form by an outside observer if an unfavorable surprise was less likely than not. More formally:

Proposition 2 *Fix a game with awareness $\mathcal{G}$. Suppose the hypotheses of Proposition 1 apply. Further suppose, there exists a fall-back option for a behavioral rule that outperforms the secure rule from the perspective of the restricted game Γ'. Then the Modified Form of the Precautionary Principal is ecologically rational, relative to the Strong Form, relative to the Strong Form, if the occurrence of an unfavorable surprise is less likely than not.*

Proof. Let r' denote the behavioral rule associated with the choice of the strategy β' in the restricted game Γ' that outperforms the secure rule in Γ'. Let $V_{\Gamma'}$ denote the expected payoff from β' in Γ'. Since the expected payoff to the secure rule is zero in both Γ and Γ', we have $V_{\Gamma'} > 0$. Let r'' denote the modification of r' to incorporate the fall-back option which corresponds to the strategy β'' in the restricted game Γ' as well as the 'fall-back' option β'''. Since r'' is a modification of r' that incorporates the fall-back option, there exists c^n and c^s such that the expected payoff from β'' in Γ' is given by $V_{\Gamma'} - c^n > 0$ and if there is an unfavorable surprise, any 'loss' from the 'fall-back' option β''' is bounded from above by $c^s < V_{\Gamma'} - c^n$. Since the hypotheses of Proposition 1 apply it follows that $0 > V(r_0)$. So to establish the ecological rationality of the Modified Form of the Precautionary Principal it is enough to show that $V(r'') > 0$. Let π denote the probability an unfavorable surprise obtains in the full game Γ when following r''. Then by construction $V(r'') \geq (1-\pi)(V_{\Gamma'} - c^n) + \pi(-c^s)$. So for $\pi < 0.5$ it follows that $V(r'') > (0.5)(V_{\Gamma'} - c^n - c^s) > 0$, as required. ■

Informally, in a game with awareness where the existence of unfavorable surprises and a fall-back option are justified inductively, we may say that the Modified Form of the Precautionary Principal is preferable to the Strong Form if a surprise is unlikely, a condition that is, in some sense, inherent in the idea of 'surprise'. The Strong Form of the Precautionary Principal is preferable to the Modified Form only in the case where the model available to the decision-maker is so limited in its representation of the problem that she is more likely than not to be surprised.

5 The Precautionary Principle for interactive decisions

Our discussion so far has treated the Precautionary Principle in terms of its relationship to models of individual decisionmaking. Moreover, much of the criticism of the Precautionary Principle has implicitly treated it as being a guide for a unitary decisionmaker, seeking to determinine an optimal public policy. Sunstein and other critics argue that such decisions should be made on the basis of an objective process such as benefit–cost analysis, with respect to which the Precautionary Principle may be treated as an alternative or variant.

However, supporters of the Precautionary Principle typically advance it in the context of contested public decisions, involving interactions between parties with different beliefs and interests. A common process of this kind is one in which a project is proposed by one actor (the proponent) and another actor (say, a regulator) must decide whether to approve or reject it, possibly taking into account the views of other parties, such as environmental organizations. In this context, the adoption of (the Strong Form of) the precautionary principle by the regulator would entail rejection of all projects where, given the regulator's level of awareness, the strategy of pursuing the project was subject to unfavorable surprises.[5]

In some decision processes, regulators may simply approve or reject proposed projects based on their own assessment. It is not uncommon, however, to require the project proponent to submit information, such as the results of environmental impact studies, with the aim of showing that the possibility of adverse surprises has been considered, or that the possible loss from such surprises is small enough to be consistent with the requirement for no strongly adverse surprises.

In this context, the precautionary principle may be seen as placing the burden of proof on project proponents. If the proponent can produce sufficient information, and therefore a sufficient increase in awareness, to satisfy the regulator that the project is not subject to unfavorable surprises, then the project can proceed. Otherwise it is rejected. This is a reversal of the position, criticised by advocates of the precautionary principle, in which there is a presumption that projects should proceed unless it can be shown that they are environmentally damaging.

5.1 The burden of proof model

We consider modifications of the model of figure 1, in which there are now two players, the project proponent (player 1) and the regulator (player 2). The game $\hat{\Gamma}$ illustrated in figure 6, represents the external (and fully aware) perspective of the strategic interaction between the proponent and the regulator. It is similar to the game of figure 1, except that at the first information set, it is the proponent who decides whether or not to undertake the study. As before, the regulator becomes

[5] We will not give a formal description of multi-player games of awareness, since this requires a discussion of reasoning about other players' awareness that is not relevant to the concerns of this paper. A detailed formal development of multi-player games of awareness is given in Grant and Quiggin (2012).

aware of the full game, and sees the realization of the signal regarding nature's action (either 'high-damage hazard' $\{a_H, a_1\}$ or 'low-damage hazard' $\{a_L, a_2\}$) only if the study is undertaken. In this case, the regulator will choose to approve or to reject the project according to whether the signal realization is 'high-damage hazard' or 'low-damage hazard'. On the other hand, if the study is not undertaken and the regulator remains unaware, she may adopt either the naive plan, approving the project in all cases, or the precautionary principle, rejecting the project whenever she believes it is subject to unfavorable surprises. We assume that that the proponent receives the benefit B whenever this is realised, while the loss from any damage (either H or L) is externalized to society at large. The cost of the study, C, if undertaken, is borne by the proponent. The regulator, as before, wishes to maximize social welfare, taking account of benefits and costs both to the proponent and to society at large.

<INSERT FIGURE 6 AROUND HERE>

The game $\hat{\Gamma}'$ in figure 7 represents the boundedly aware perspective of the regulator (and possibly also the proponent) who is unaware of the possibility of high damage.

<INSERT FIGURE 7 AROUND HERE>

We first examine the case in which initially both parties are unaware that there may exist unconsidered possibilities. For games with awareness, where unconsidered possibilities are not taken into account as is illustrated in the game $\hat{\Gamma}'$ in figure 7, it is straightforward to see that the strategy pair ('no study', 'always approve') is an equilibrium.

The equilibrium outcome is unchanged if the proponent, but not the regulator, initially believes that there may exist unconsidered possibilities. This corresponds to the situation in which the perspective of the proponent corresponds to the game in figure 8 that augments $\hat{\Gamma}'$ with a virtual node to account for the proponent's consciousness of the possible existence of hazards of which he is not aware. The reasoning is a little more involved, however, since in order to predict an outcome for this game with awareness, it is necessary to consider the proponent's prediction of the regulator's decision. This prediction will in turn depend on the regulator's level of awareness, as perceived by the proponent. For the proponent, the study incurs the cost C but it can only create possibilities for rejection of the project. Hence, the Precautionary Principle (in either its Strong or Modified form) reinforces the proponent's choice of 'no study', given the prediction by the proponent that, with the regulator's awareness remaining limited if 'no study' is chosen, the regulator will choose 'always approve'. However, if after the choice of 'study', the regulator's awareness is expanded to include now the possibility of high damage then, the proponent should predict the regulator will follow the behaviorial rule "approve if no study or if study and signal realization is 'low-damage hazard', disapprove if study and signal realization is 'high-damage hazard'."

<INSERT FIGURE 8 AROUND HERE>

In the case where the regulator, but not the proponent believes that there may exist unconsidered possibilities, adoption of either form of the Precautionary Principle will lead the regulator to reject the project in the absence of a study. Once again, to predict an outcome for the game with awareness, it is necessary to consider the proponent's prediction of the regulator's decision. There are two possibilities here. On the one hand, a naive proponent will anticipate approval, and will therefore fail to undertake the study, suffering an unpleasant surprise when the proposal is rejected. Alternatively, even though the proponent himself does not believe an unfavorable surprise is possible, he may be aware of the regulator's likely behavior, and therefore anticipate that the decision 'no study' will lead to rejection of the project. By contrast, he believes that the decision 'study' will reveal the absence of unfavorable surprises, and therefore lead to approval by the regulator. Hence, provided $B > \max(p_D L, C)$ the study will be conducted.[6]

Finally, and most significantly, consider the case where both the regulator and the proponent regard the proposition that the project may be subject to unfavorable surprises is supported by historical induction. If the proponent chooses 'no study', the regulator, applying the Precautionary Principle, will choose 'reject'. We assume, as in the previous case, that the proponent understands this, and is therefore faced with a choice between undertaking the study, which may or may not result in acceptance, or not undertaking the study, leading to the regulator choosing 'reject'.

As in the one-player game with awareness, analyzed in section 4 above, where the regulator decides on whether or not to undertake the study, adoption by the Proponent of the Strong Form of the Precautionary Principle leads him to choose the action 'no study' and the regulator adopting the behavioral rule "approve if study and signal is 'low-damage hazard', otherwise reject" leading to the play of the game in which the proponent chooses 'no study' and the regulator chooses 'reject'. On the other hand, given the belief of the proponent that the regulator is following the rule "accept if study and signal is 'low-damage hazard', otherwise reject," and if $B - C > C$ then application of the Modified Form of the Precautionary Principle *by the proponent* leads him to choose 'study'. Furthermore, from proposition 2 it follows if it is less likely than not that the study will show there is a risk of high-damage (that is, $p_H + p_1 < p_2 + p_L$) then following this course of action is ecologically rational for the proponent.

Compared to the case of a unitary decision, the proponent has a greater incentive to undertake the study, since he captures the benefits if the project is successful, but the losses in the event of a moderate hazard being realised are externalized.

---

[6] $B > C$ means it is profitable for the proponent to incur the cost of the study should this induce the regulator to accept the project and $B > p_D L$ ensures that the regulator will approve the project after a study has been undertaken and the realization of the signal is 'low-damge hazard'.

6 Concluding comments

Informally stated, the precautionary principle has strong intuitive appeal, particularly in the context of environmental regulation. In dealing with complex, fragile and poorly understood natural systems, it seems to make sense 'to err on the side of caution.' However, this way of putting things points up the difficulties in formalizing the precautionary principle. 'To err' means to commit an error, and it is obviously difficult to include a prescription for error in a formal theory of decision under uncertainty. Yet decisions are prone to errors arising from an incomplete understanding of the problem at hand, and of the likelihood that some contingencies will not be taken into account. It seems desirable to take account of this reality in formulating a procedure for making decisions.

In this paper, we have addressed the question in relation to the standard Bayesian model of decision theory, developed in the framework of an extensive-form game. We have argued that the precautionary principle is best understood, not as a modification of Bayesian decision theory, but rather as a heuristic constraint on the application of that theory; that is, as a response to the recognition that the outcomes of decisions may be affected by unforeseen contingencies. Heuristic constraints such as the precautionary principle must be satisfied before it is appropriate to apply the tools of Bayesian decision theory.

The precautionary principle is most commonly applied in relation to interactive decisions, involving judgements as to whether or not to proceed with projects or innovations that may pose unforeseen risks. In this context, the precautionary principle may be regarded as a procedural rule that places the burden of proof on to advocates of decisions involving poorly-understood risks.

References

Grant, Simon, and John Quiggin (2012), "Inductive Reasoning About Unawareness," Working Paper WP R12_2, Risk and Sustainable Management Group, University of Queensland.

Gollier, C., Jullien, B. and Treich, N. (2000), "Scientific progress and irreversibility: An economic interpretation of the 'precautionary principle'," *Journal of Public Economics*, 75, 229–53.

Goldstein, D. and Gigerenzer, G. (2002), "Models of Ecological Rationality: The Recognition Heuristic," *Psychological Review*, 109(1), 75-90.

Halpern, J. and Rêgo, L. (2006), "Extensive games with possibly unaware players," Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems, Hakodate, Japan, May 8-12.

Halpern, J. and Rêgo, L. (2008), "Interactive unawareness revisited," *Games and Economic Behavior*, 62(1), 232–262.

Halpern, J. and Rêgo, L. (2009), "Reasoning about knowledge of unawareness," *Games and Economic Behavior*, 67(2), 503–525.

Halpern, J. and Rêgo, L. (2011), "Reasoning about knowledge of unawareness revisited," Proceedings of the Tenth International Conference on Principles of Knowledge Representation and Reasoning.

Heifetz, A., Meier, M. and Schipper, B. (2006), "Interactive unawareness," *Journal of Economic Theory*, 130(1), 78–94.

Heifetz, A., Meier, M. and Schipper, B. (2009), "Dynamic unawareness and rationalizable behavior," Working Paper, UC Davis.

Rumsfeld, D. (2002), Statement to Press, Washington, February.

Osborne, M.J. and Rubinstein, A. (1994), *A Course in Game Theory*, MIT Press, Cambridge, Mass.

Sunstein, C.R. (2003), "Beyond the Precautionary Principle," *University of Pennsylvania Law Review* 151(January): 1003-58.

Walker, O. (2012), "Unawareness with 'Possible' Possible Worlds," mimeo London School of Economics.

Wingspread Conference (1998), "Wingspread Statement on the Precautionary Principle," Press release, Racine, Washington, February.

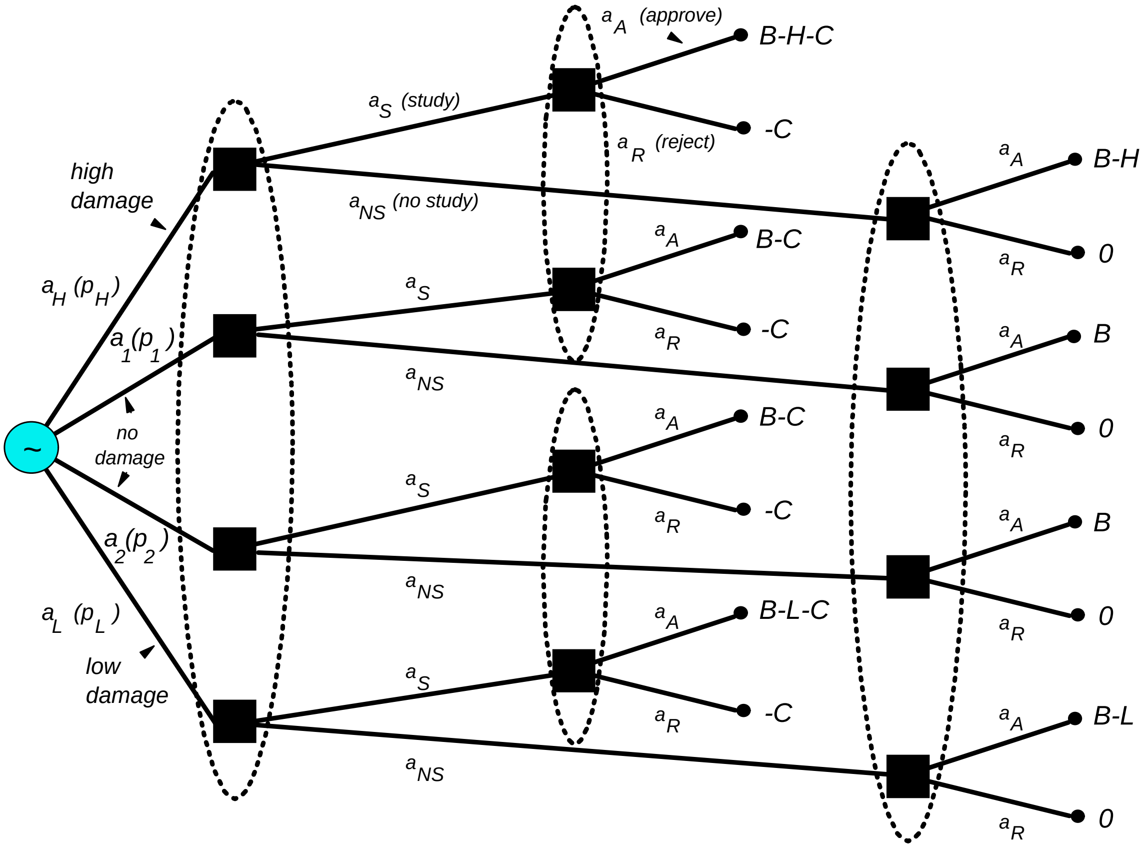

Figure 1: The game Γ represents the external (fully aware) perspective.

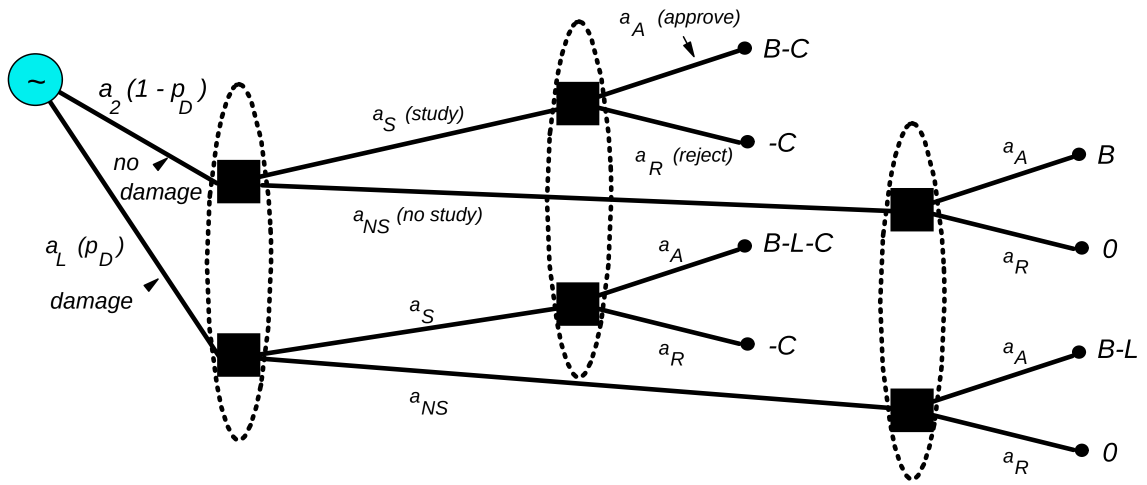

Figure 2: The game Γ′ represents the (boundedly aware) perspective of a decision who is unaware of the possibility of High damage.

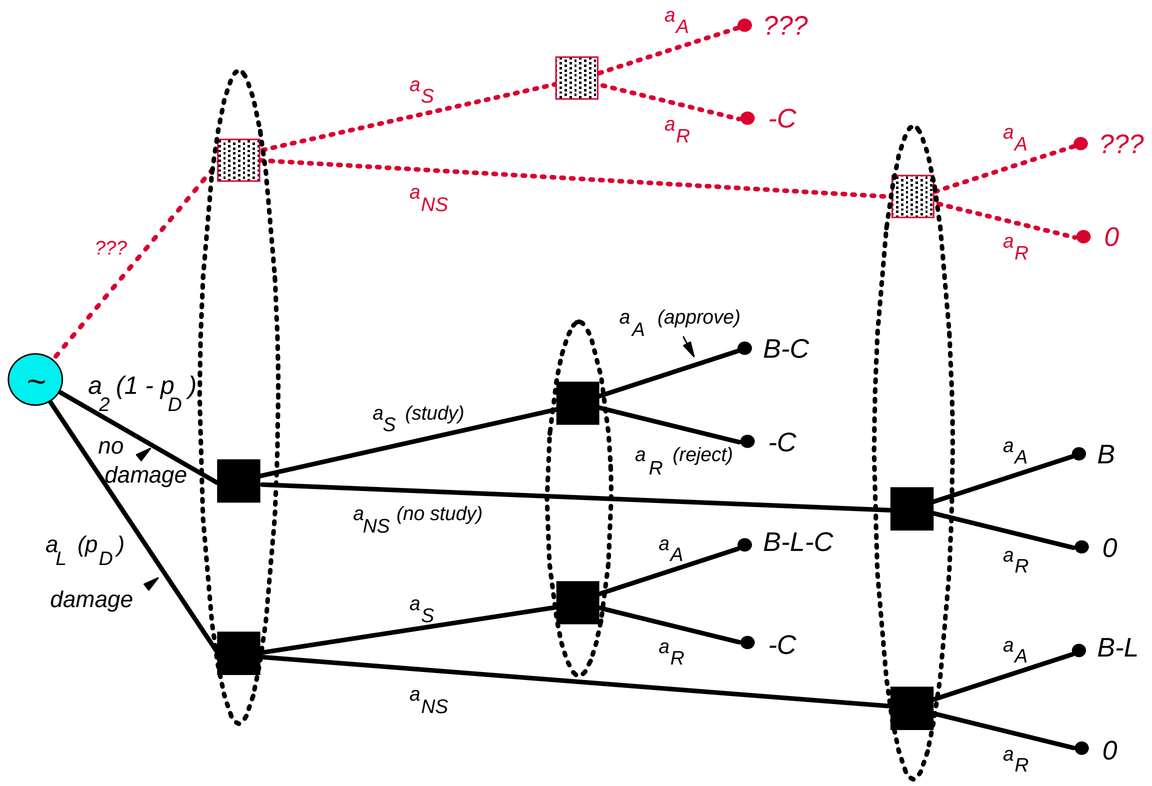

Figure 3: The game Γ′ from figure 2 has been augmented to account for the decisionmaker's implicit awareness of hazards of which she is not *explicitly* aware.

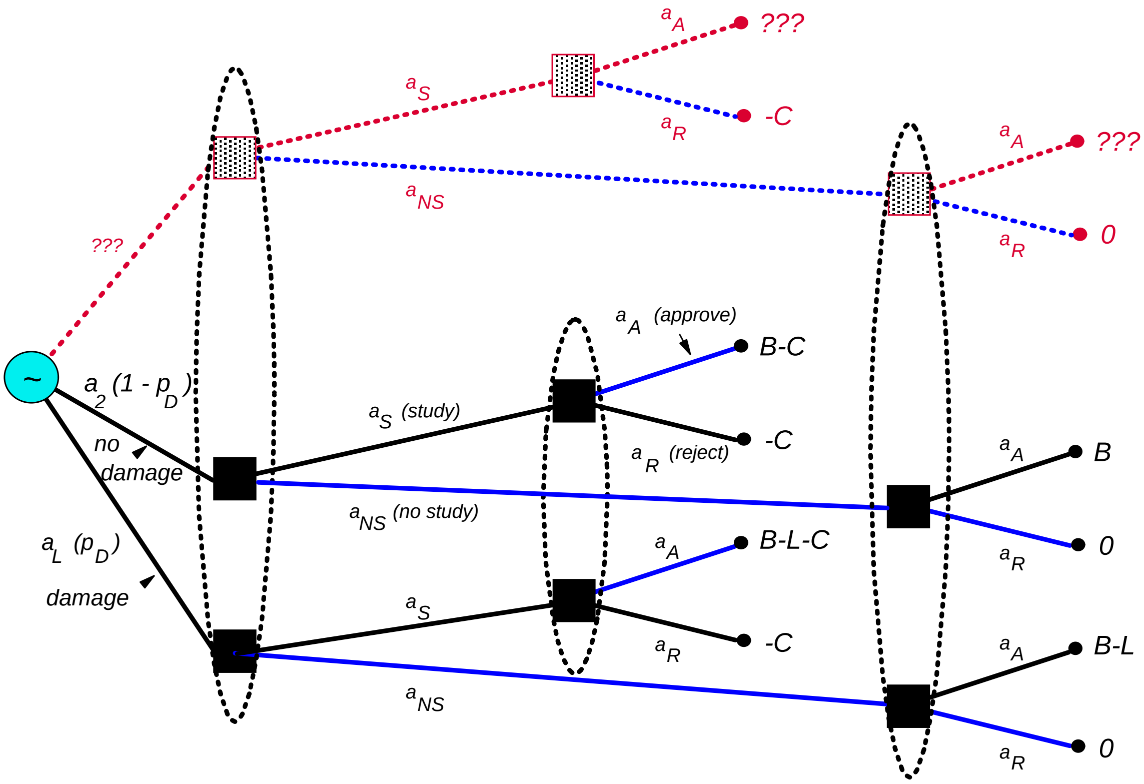

Figure 4: Precautionary principle (strong form).

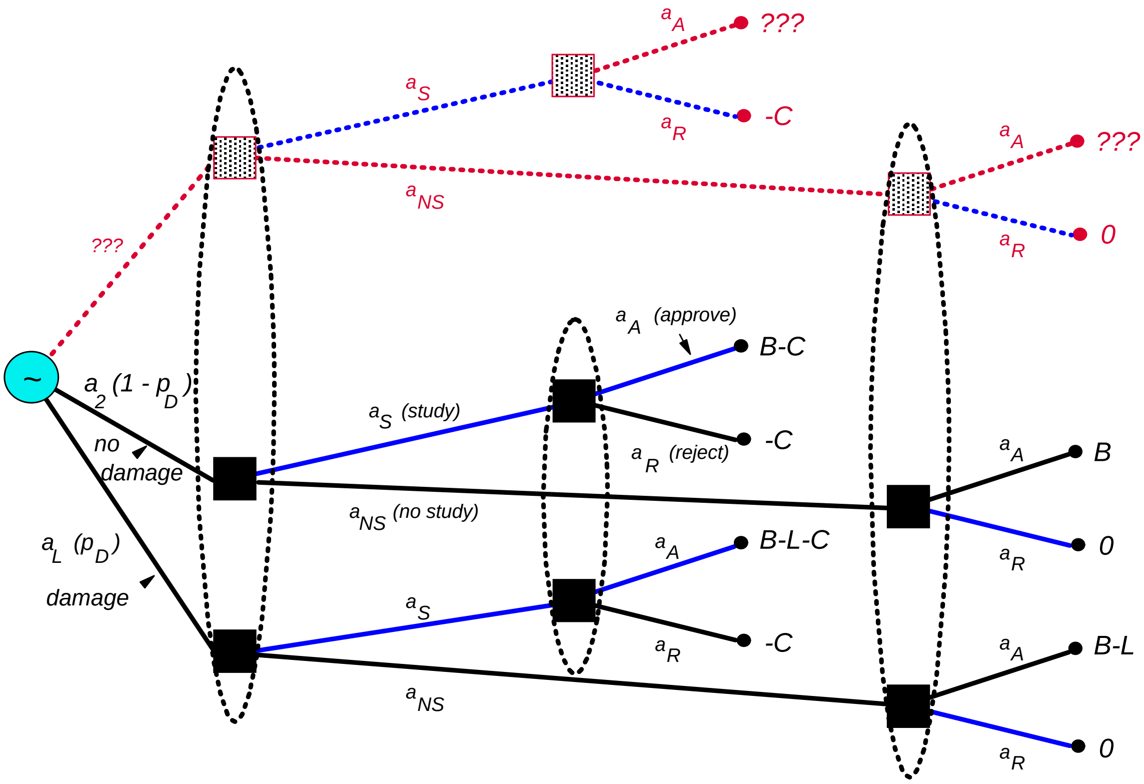

Figure 5: Precautionary principle (moderate form).

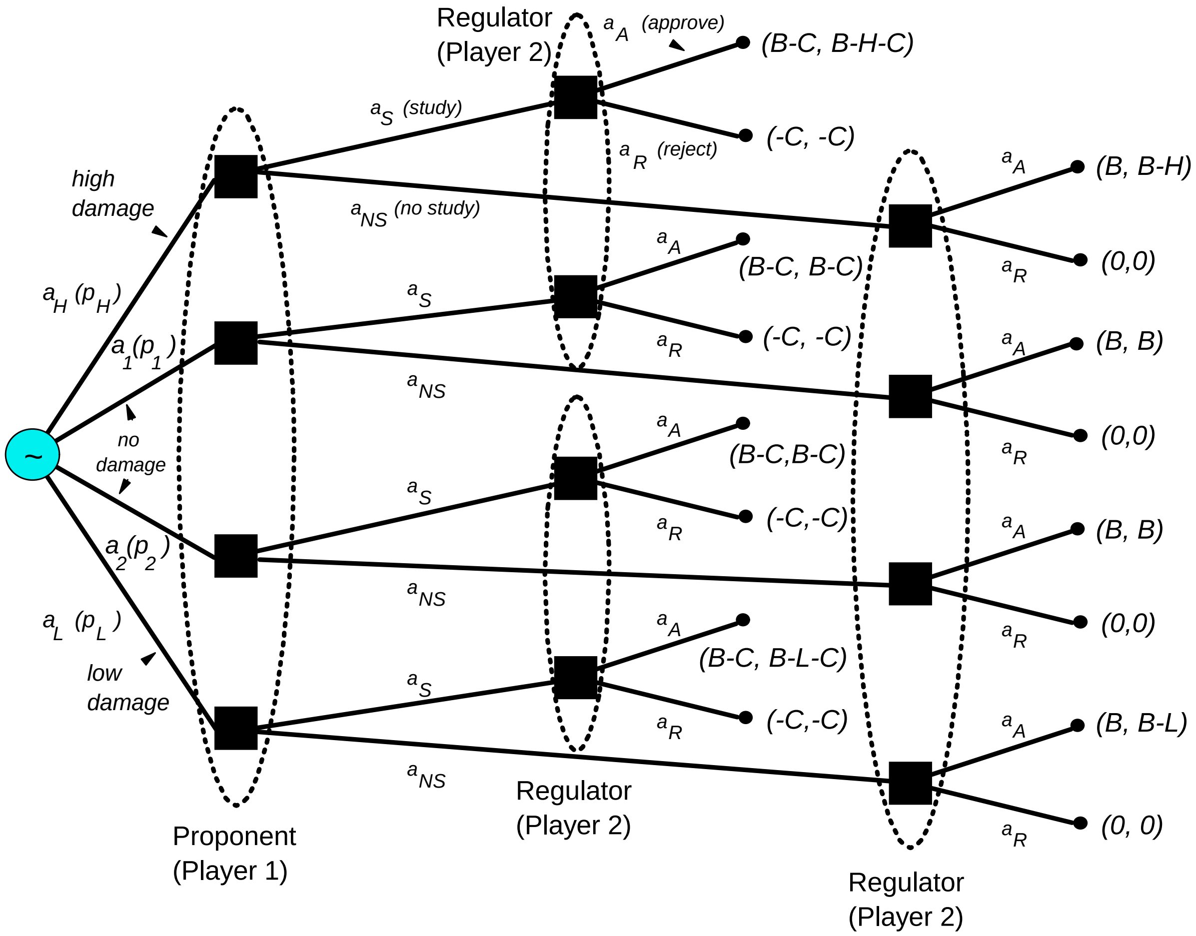

Figure 6: The game $\hat{\Gamma}$ represents the external (fully aware) perspective of the strategic interaction between a Proponent (Player 1) and a Regulator (Player 2).

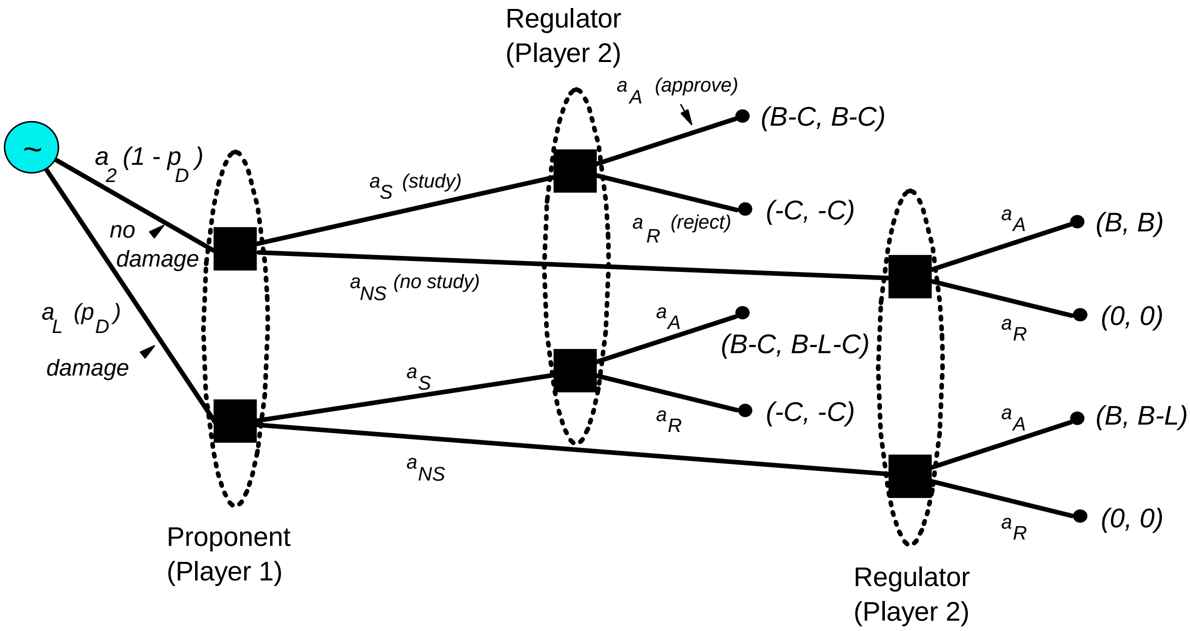

Figure 7: The game $\hat{\Gamma}'$ represents the boundedly aware perspective of a player (possibly the proponent or the regulator or both) who is unaware of the possibility of High damage.

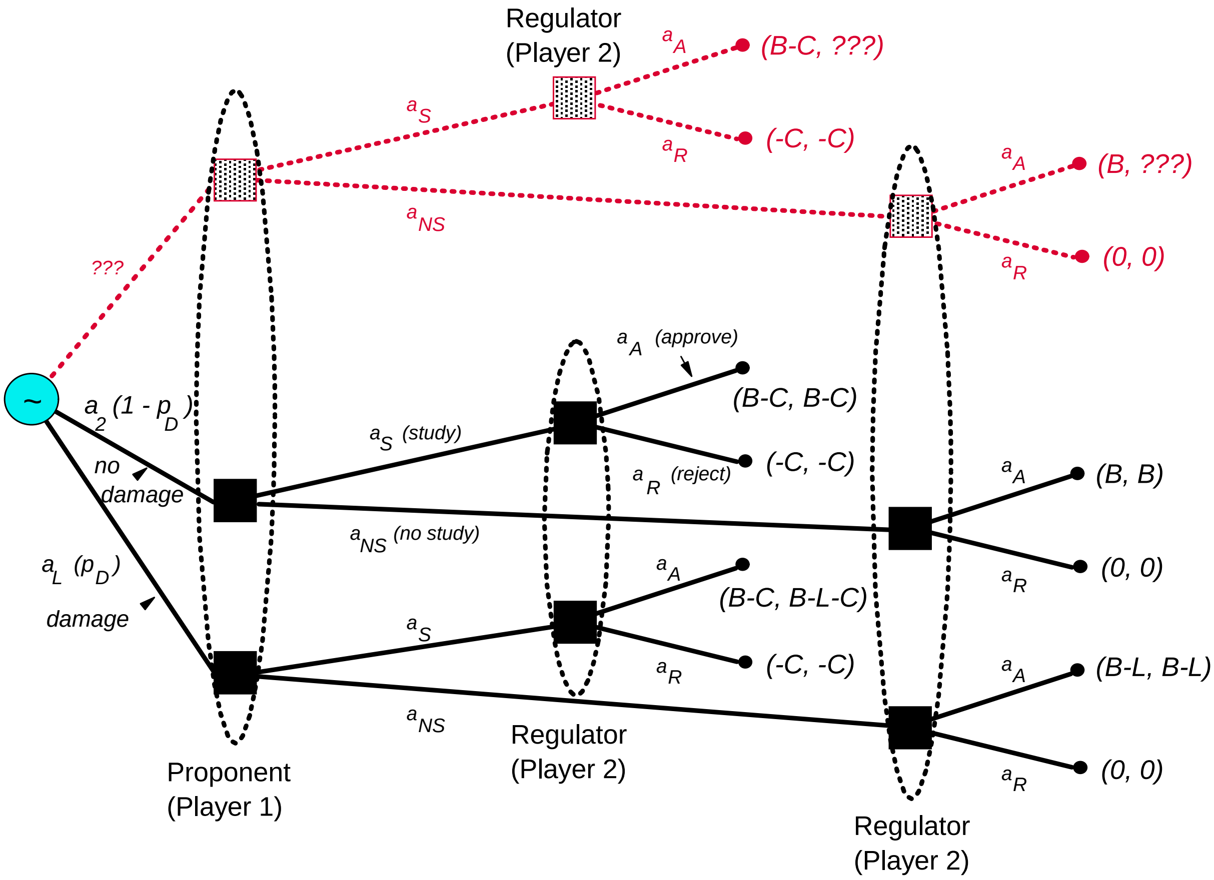

Figure 8: The game $\hat{\Gamma}'$ from figure 7 has been augmented to account for a player's implicit awareness of hazards of which she is not *explicitly* aware.