Subjective Reasoning – Games with Unawareness

Yossi Feinberg[*][†]
Stanford University

Abstract

The subjective framework for reasoning is extended to incorporate the representation of unawareness in games. Both unawareness of actions and decision makers are modeled as well as reasoning about others' unawareness. It is shown that a small grain of uncertainty about unawareness with rational decision makers can lead to cooperation in the finitely repeated prisoner's dilemma. **JEL Classification: C72,D81,D82.**

1 Introduction

Extending games with the introduction of irrational types has led to the development of the notions of reputation in a wide variety of applications. Since Kreps, Milgrom, Roberts and Wilson (1982) showed that a grain of irrationality induces cooperation in the finitely repeated prisoner's dilemma, henceforth FRPD, and Milgrom and Roberts (1982) and Kreps and Wilson (1982) studied reputation in games such as an entry deterrence game, it has become popular to study the effects of a grain of irrationality in a variety of dynamic games. However, it seems that the irrational behavior exogenously added to the game is almost never arbitrary. Rather, it is a very specific kind of irrational behavior, a behavior that serves the interest of the *rational* player, a behavior that is worthwhile mimicking. Even in the celebrated studies mentioned above we find that an arbitrary irrational behavior need not generate the desired impact on the game's outcome. The question arises why should some specific irrational behavior emerge as a candidate for a grain of irrationality?

In this paper we suggest an alternative to the grain of irrationality approach. Namely, we offer a grain of unawareness as a possible modification to the game. We consider the

[*] *Stanford Graduate School of Business, Stanford CA 94305-5015.* e-mail: yossi@gsb.stanford.edu http://www.stanford.edu/~yossi/Main.htm

[†] I wish to thank Kyna Fong, Joe Halpern, David Kreps, Salvatore Modica, John Roberts and Tomasz Sadzik for helpful comments and suggestions. I gratefully acknowledge the support of the Center for Electronic Business and Commerce.

possibility that a decision maker is not fully aware of the game being played. This modification adds possible restricted views of the game but assumes the decision makers are rational subject to their unawareness of the full scope of the game. With this modification we are able to show that a grain of unawareness can generate cooperation in the FRPD.

Our FRPD game with a grain of unawareness begins with Nature choosing whether Alice will be aware or unaware that she is repeatedly playing the PD with Bob. More precisely, Alice may be unaware that the game allows her, or Bob, to defect. Bob, on the other hand is fully aware of the game; he is also aware that Alice can be potentially unaware. However, Bob is initially uncertain whether Alice is aware or not. If Alice is aware of the game she is confident that Bob is uncertain of her awareness, Bob is confident of that and so on. Naturally, if Alice is unaware that defection is possible she is also unaware that Bob is uncertain about her unawareness and unaware that she could have possibly been aware. In fact, she is unaware of Nature's move that determined her state of awareness. We also assume that the unaware Alice will become aware that defection is possible if and only if she observes Bob defecting and this is known to Bob. This corresponds to assuming that Alice and Bob commonly know that the game is symmetric and repeated. Hence, if the unaware Alice observes a defection then she becomes aware of the full scope of the game, and at that point Alice and Bob have common confidence that they are playing the FRPD.

From this verbal description of the FRPD with unawareness one anticipates that cooperation should emerge. Indeed, Bob has an incentive to cooperate since if Alice is unaware then defection would lead to the standard FRPD and defection throughout by both players. In addition, the aware Alice has an incentive to mimic the unaware Alice and cooperate since otherwise she is revealed to be aware and we are back in the standard FRPD leading to defection throughout. It seems that all the ingredients leading to cooperation in the FRPD with a grain of irrationality are present in our case. What we are missing is the formal framework for defining dynamic games with unawareness and solutions for these games. The purpose of this study is to fill this gap.

We translate the informal description of unawareness and reasoning about unawareness into a formal language. We then epistemically define dynamic games with unawareness. Throughout the paper the construction is applied to the specific example of the FRPD with unawareness described above. We show that the informal example translates into a rigorous unawareness construction and an epistemic form game with unawareness. The definition of the epistemic form of a game with unawareness has the desired property that from each subjective state of awareness the decision maker's perception is that a game with unawareness is played. Furthermore, everyone is aware that others view the situation as a game with unawareness and so on. We show that this language allows for reasoning about the unawareness of decision makers. Using the epistemic characterization in Feinberg (2004b) we extend sequential equilibria to the FRPD with a grain of unawareness. Our epistemic foundation for games with unawareness allows for a formal proof that the FRPD with unawareness leads to cooperation much like a grain of irrationality does.

In order to develop the concept of games with unawareness we first need to formalize the notion of unawareness itself. The importance of unawareness stems from the casual observation that an economic decision maker seldom has the privilege of comprehending the full scope of a decision problem at hand. It is often the case that some events or other decision makers influence the outcomes although the decision maker is unaware of these events and actors. It is not that she considers these influences unlikely, or improbable, it is that she does not consider them at all – from her subjective viewpoint they simply do not exist. Hence, these objects never enter her reasoning while making a decision, in other words, they are not part of her language and she is incapable of using them while reasoning. But we wish to capture more than that; we want to allow that another decision maker may want to reason about her unawareness. Hence, we want his reasoning about her to incorporate the fact that he may find her to be unable to reason about some events. Furthermore, it might be the case that he is unaware of an event that she *is* aware of, hence her reasoning about this event is beyond the scope of his reasoning.

We model unawareness as the inability to reason about fundamental events or the existence of other decision makers, in the sense that unawareness is interpreted as non-existence in the language. This is captured by formally excluding events or decision makers that a decision maker is unaware of from the language used by this decision maker. We use the restricted language to also capture higher orders of unawareness and reasoning of agents about the unawareness of others. This is achieved by restricting the expression of the agent's reasoning about others' in a way that reflects the agent's awareness of others' awareness – their subjectively restricted language. This representation of reasoning about unawareness is based on an extension of the subjective reasoning language presented in Feinberg (2004a).

A number of studies considered the modeling of unawareness in economic theory. Dekel, Lipman and Rustichini (1998) show the difficulties in modeling the unawareness of a decision maker with a partitional model. Modica and Rustichini (1994) point out the difficulties in modeling partial awareness without relaxing the inference rules. They are able to solve this difficulty for a single decision maker by providing a framework for describing unawareness and a determination theorem for the system which captures a decision maker's state of awareness. They capture unawareness via knowledge by defining unawareness of an event φ as $\neg k\varphi \wedge \neg k \neg k\varphi$, not knowing the event and not knowing that you don't know the event. Modica and Rustichini (1999) show how to relax the negative introspection axiom $\neg k\varphi \implies k\neg k\varphi$

to allow for unawareness and be able to represent knowledge with unawareness. The model of Modica and Rustichini is applied to general equilibrium settings by Modica, Rustichini and Tallon (1998) and Kawamura (2004) where individuals may be unaware of all possible future contingencies.

In the field of artificial intelligence the issue of modeling unawareness has been more widely studied. Fagin and Halpern (1987) provide a number of general formulations for representing unawareness and its interaction with knowledge and belief. Their systems not only deal with unawareness, but also expand on the approach proposed by Levesque (1984) towards the logical omniscience problem. The logical omniscience problem – the inability of the reasoning agent to know all the logical implications – is closely related to the awareness of an agent. Fagin and Halpern (1987) have also allowed for multi-agent reasoning with unawareness as well as dynamics for unawareness which considers agents that can learn and forget [1].

More recently, Heifetz ,Meier and Schipper (2004) have provided a semantic (state space) construction for interactive unawareness[2]. Their framework allows agents to reason about the awareness of others. By using a lattice structure for state spaces that represents the relative depth of perception of the agents, they can capture identities that are more aware than other identities as having a finer – more expressive – state space. They are able to show that this construction retains a host of desired properties of unawareness when unawareness is defined as not knowing and not knowing that you don't know. Using this model they are able to show that mere mutual unawareness can lead to speculative trade. We also note the work of Ewerhart (2001) where interactive unawareness is modeled in relation to the agreeing to disagree theorem and the set theoretic construction of interactive unawareness by Li (2004). The latter formulation begins with a possible worlds correspondence representing awareness as a primitive and then defines knowledge restricted to the states the agent is aware of.

We carry a variety of principles from the existing literature into our framework. The construction of unawareness from a restricted set of atomic statements appeared in Modica and Rustichini (1999), the syntactic approach and the relativism of necessitation can be found in Fagin and Halpern (1987), both Heifetz ,Meier and Schipper (2004) and Li (2004) provide approaches to interactive awareness. Extending the subjective framework for reasoning in games with these principles allows us to define games with unawareness and reason about behavior in these games. Games with unawareness, and in particular behavior and reasoning in these games, are the main objectives of this work.

[1] See also Halpern (2001) for relating Modica and Rustichini (1999) with Fagin and Halpern (1987). Also see Fagin et al. (1995) for more discussion on unawareness and logical omniscience.

[2] Halpern and Rêgo (2005) axiomatize this framework in a three valued logic framework.

An identity (a decision maker at a decision point in a dynamic game) is unaware of an event if that event is not part of the language that describes her reasoning. Hence, we associate with an identity her own subjective syntax that describes which atomic statements – the basic building blocks of the language at hand – are available to her[3]. Limiting the syntax of the reasoning identity allows us to capture unawareness of the *existence* of other decision makers by removing any reference to their confidence or belief operators from the language available to the reasoning identity. Higher order unawareness is captured by considering the set of atomic statements and identities that each identity finds that other identities are aware of. Some consistency requirements are postulated, for example, "if she is aware that he is aware of an atom then she must also be aware of that atom."

By extending the subjective framework developed in Feinberg (2004a) to reasoning about unawareness we retain the ability to epistemically represent games and reasoning in games within the same language. In a game with unawareness each identity perceives the game as consisting of the decision points and actions she is aware of. Unawareness of the full scope of the game is represented in the subjective framework with an epistemic description of the game as viewed from the limited language of a given identity. Furthermore, we can represent the awareness of identities of how other identities perceive the game, as well as higher order of awareness, by considering the reasoning of one identity about how others view the game. Our ability to account for unawareness of the existence of other identities plays a crucial role in the definition of games with unawareness. The epistemic form of a game with unawareness is the collection of statements that describe the view of each decision maker in the game, where this view might be only a partial view of the full game. It also describes how each identity views how others view the game and so on. We show that the description of a given state of awareness in a game, or higher order of awareness, is itself part of a game with unawareness: the game with unawareness as seen from this state of awareness. Conceptually speaking, we can say that the players understand that they are playing a game with unawareness, they understand that others understand *that*, and so on.

Once a game with unawareness is defined, we turn to the definition of a solution in such a game. Here we use an epistemic characterization to extend a solution to games with unawareness. In particular, we use the epistemic characterization of sequential equilibria in Feinberg (2004b) to define a solution for the game with unawareness. We restrict the beliefs and reasoning about rationality that characterize the solution in standard dynamic games to the language that reflects the unawareness at hand. The mapped conditions constrain the

[3] Our approach to unawareness via a restricted language for reasoning should not be confused with the work of Crawford and Haller (1990) where players use different languages in the sense that they cannot agree on the meaning of the language in order to coordinate their actions.

behavior in the game with unawareness, hence they characterize behavior which is defined as the extended solution. Conceptually, building on the subjective framework allows us to capture reasoning in games with unawareness as follows:

$$\begin{tikzcd}
Epistemic\ Game\ Form \arrow[r, "Unawareness\ Construction"] \arrow[d, "Epistemic\ Characterization"'] & Game\ with\ Unawareness \arrow[d, "Extended\ Characterization"] \\
Solution \arrow[r, "Unawareness\ Construction"] & Extended\ Solution
\end{tikzcd}$$

Armed with a solution concept for a game with unawareness we show that cooperation emerges in the FRPD with a grain of unawareness.

In Section 2 we define the framework for describing unawareness and reasoning about unawareness. In Section 3 we define games with unawareness and study their properties. Section 4 contains the extension of a solution to games with unawareness and the result stating that a grain of unawareness leads to cooperation in the FRPD. A discussion of unawareness as bounded rationality and concluding remarks appear in Section 5.

2 The Syntax for Unawareness

The syntactic formulation of a language for unawareness is based on the subjective reasoning framework as developed in Feinberg (2004a). Consider a given finite set $\alpha = \{a, b, c, ...\}$ the elements of which are called *atomic formulas* or *atomic statements*. We will call α the *global alphabet*. A finite set I is called the *global identity set*. The global identity set I contains all possible manifestations of each player, i.e., an identity for each decision point (vertex) in the game tree which includes an identity for any possible state of awareness for a player. The pair $\{I, \alpha\}$ is called the *global context*.

Each identity possess its own subjective view of the world, but departing from Feinberg (2004a), we now associate with each identity its subjective context. This context defines the set of atomic statements and identities that a given identity is aware of. Denote by $I_i \subset I$ the set of identities that $i \in I$ is aware of and by $\alpha_i \subset \alpha$ the set of atomic statements that i is aware of. We will assume that each identity is aware of its own existence $i \in I_i$ and of at least one atomic statement so $\alpha_i \neq \emptyset$. Since we are considering reasoning about unawareness we must allow an identity to reason about the awareness of another identity. For every pair of identities $i, j \in I$ we denote by $I_{i,j}, \alpha_{i,j}$ the set of identities that i is aware that j is aware of and the set of atomic statements that i is aware that j is aware of, respectively. If $j \notin I_i$ we let $I_{i,j} = \emptyset = \alpha_{i,j}$. We generalize this construction to any finite sequence of identities in $\Theta = \bigcup_{n=0}^{\infty} (I)^n$ where $(I)^0 = \{\emptyset\}$ and $(I)^n$ is the set of all n-tuples of identities. For every $\theta \in \Theta$, $\theta = (i_1, ..., i_n)$, we define $I_\theta \subset I$ and $\alpha_\theta \subset \alpha$ to be the set of identities and atomic statements that i_1 is aware that i_2 is aware that ... that i_n is aware of, respectively. We let $I_{\{\emptyset\}} = I$ and $\alpha_{\{\emptyset\}} = \alpha$. We assume that $I_\theta \neq \emptyset$ if and only if $\alpha_\theta \neq \emptyset$ and we denote $\bar{\Theta} = \{\theta \in \Theta | I_\theta \neq \emptyset \text{ and } \alpha_\theta \neq \emptyset\}$. We refer to $\theta \in \bar{\Theta}$ as an instance of higher order awareness, as a state of awareness or as iterated awareness. We postulate the following consistency conditions for higher order of unawareness:

1. For every $\theta = (i_1, ..., i_n) \in \Theta$, for every $\bar{\theta} = (i_{k_1}, i_{k_2}, ..., i_{k_m})$ with $1 \leq k_1 < k_2 < ... < k_m \leq n$ we have that $I_\theta \subset I_{\bar{\theta}}$ and $\alpha_\theta \subset \alpha_{\bar{\theta}}$.

   This condition states that whenever Alice is aware that Bob is aware of an atomic statement (resp. identity) then Alice must also be aware of that statement (resp. identity). This means that if in the language for reasoning we have statements that describe Alice's reasoning about Bob's reasoning about an event then the statement about Bob's reasoning about the event is also a part of the language. It also assumes that if Alice is aware that Bob is aware of a statement then Bob – that subjective Bob that Alice is aware of – is indeed aware of that statement. There is also higher order awareness of these two properties. This does not imply that Bob is indeed aware of the statements. It is possible that Bob is unaware of the statement and is actually a different identity. However, since Alice is reasoning about an identity of Bob that *is* aware of the statement, the language must include a hypothetical identity for Bob that is aware of the statement. Note, that in this case Alice will be unaware of the existence of the other hypothetical identity of Bob that is unaware of the statement and the identity that is unaware of the statement will also be unaware of the identity that is aware of the statement. Hence, the subjective framework allows Alice to be aware that Bob is aware of a statement when Bob is actually not aware of that statement by using hypothetical identities.

2. For $\theta = (i_1, ..., i_k, i_{k+1}, ..., i_n)$ such that $i_k = i_{k+1}$ for some k we have $I_\theta = I_{\bar{\theta}}$ and $\alpha_\theta = \alpha_{\bar{\theta}}$ where $\bar{\theta} = (i_1, ..., i_{k-1}, i_{k+1}, ..., i_n)$.

   Here we require that Alice is aware of everything that she is aware that she is aware of. Note, that the first condition implies that what she is aware that she is aware of, she must be aware of. In addition, there is higher order awareness of that.

3. For all $\theta = (i_1, ..., i_n) \in \bar{\Theta}$ we have $i_n \in I_\theta$.

   The third condition states that Alice is aware of her own existence and that if there is higher order awareness of Alice being aware of something, i.e. of Alice's reasoning, then there is the same high order awareness of Alice's awareness of herself and in particular awareness of Alice.

Definition 1 *A collection $\mathcal{U} = \{I_\theta, \alpha_\theta\}_{\theta\in\Theta}$ which satisfies the consistency conditions $1. - 3.$ above is called an* unawareness construction *or an* awareness construction[4].

An awareness construction $\mathcal{U}$ defines a specific high order awareness relationship among a given set of identities. We now turn to define the syntax that corresponds to a given awareness construction $\mathcal{U}$. The language associated with $\mathcal{U}$ will be denoted $\mathcal{L}^{\mathcal{U}}$ and will be composed of a subset of the global language that includes all the statements generated by the global context.

As in Feinberg (2004a) we consider the operators $\neg, \wedge, C_i, P_i^r, u_i^x$ for every identity $i \in I$, rational $r \in [0, 1]$ and real $x \in [-M, M]$ for some positive bound M. All finite applications of these operators constitute the global language $\mathcal{L}^G$, i.e. all atomic formulas are in $\mathcal{L}^G$ and $\neg f$, $f \wedge g$, $C_i f$, $P_i^r f$, $u_i^x f$ are in $\mathcal{L}^G$ whenever $f, g \in \mathcal{L}^G$. The interpretations of these statements are "not f", "f and g", "i is confident of f", "i assigns to f a probability of at least r" and "i assigns a utility of at least x to f" respectively. In general, for every context $\{I, \alpha\}$ we denote the corresponding language by $\mathcal{L}^{\{I,\alpha\}}$.

In order to decide which statements in $\mathcal{L}^G$ are allowed in $\mathcal{L}^{\mathcal{U}}$ – the restricted language determined by the unawareness construction – we define a mapping ρ which associates with every $f \in \mathcal{L}^G$ a subset of $\Theta \times \alpha$. The collection $\rho(f) \subset \Theta \times \alpha$ describes the awareness implicit in f. The mapping ρ is defined inductively as

$$\begin{aligned} \rho(a) &= (\emptyset, a) \text{ for atomic statements } a \in \alpha \qquad (1) \\ \rho(\neg f) &= \rho(f) \\ \rho(f \wedge g) &= \rho(f) \cup \rho(g) \\ \rho(C_i f) &= \rho(P_i^r f) = \rho(u_i^x f) = \{(i\hat{}\theta, a) | (\theta, a) \in \rho(f)\} \end{aligned}$$

where for $\theta = (i_1, ..., i_n)$ and $\bar{\theta} = (j_1, ..., j_m)$ we define the concatenation $\theta\hat{}\bar{\theta} = (i_1, ..., i_n, j_1, ..., j_m)$.

The function ρ associates with each statement f the awareness assumed when the statement f is considered. For example, if $f = \neg(a \wedge C_i C_j b) \wedge C_k(a \wedge P_i^{0.5} d)$ where a, b, d are

[4] We freely interchange the terms "awareness construction" and "unawareness construction".

atomic statements, then $\rho(f) = \{(\emptyset, a), ((i, j), b), (k, a), ((k, i), d)\}$, i.e. f assumes that i is aware that j is aware of b, that k is aware of a and that k is aware that i is aware of d.

With the mapping ρ we can identify whether a statement respects an awareness construction. We define the language $\mathcal{L}^{\mathcal{U}}$:

$$\begin{aligned} \forall f &\in \mathcal{L}^{G} \text{ we let } f \in \mathcal{L}^{\mathcal{U}} \text{ if and only if} \\ \text{for all } (\theta, a) &\in \rho(f) \text{ with } \theta = (i_1, ..., i_n),\ n \geq 1 \text{ we have } a \in \alpha_\theta,\ i_n \in I_{(i_1,\ldots,i_{n-1})}. \end{aligned} \tag{2}$$

In other words, we say that f is in $\mathcal{L}^{\mathcal{U}}$ whenever the awareness implicitly assumed in f respects the awareness restrictions posed by $\mathcal{U}$.

Proposition 2 *For every awareness construction $\mathcal{U}$ we have::*

$f \in \mathcal{L}^{\mathcal{U}}$ if and only if $\neg f \in \mathcal{L}^{\mathcal{U}}$

$f, g \in \mathcal{L}^{\mathcal{U}}$ if and only if $f \wedge g \in \mathcal{L}^{\mathcal{U}}$.

If one of the operators $C_i f$, $P_i^\alpha f$ or $u_i^r f$ is in $\mathcal{L}^{\mathcal{U}}$ then so are the other two.

$C_i f, C_i g \in \mathcal{L}^{\mathcal{U}}$ if and only if $C_i(f \wedge g) \in \mathcal{L}^{\mathcal{U}}$.

$C_i f \in \mathcal{L}^{\mathcal{U}}$ if and only if $C_i \neg f$.

Proof. Follows directly from the definition of $\mathcal{L}^{\mathcal{U}}$. ■

Finally, the axiom scheme associated with the restricted language $\mathcal{L}^{\mathcal{U}}$ is based on the framework for subjective reasoning in Feinberg (2004a).

For all $f, g \in \mathcal{L}^{\mathcal{U}}$ we have the propositional calculus axioms:

PC1 $(f \vee f) \Longrightarrow f$

PC2 $f \Longrightarrow (f \vee g)$

PC3 $(f \vee g) \Longrightarrow (g \vee f)$

PC4 $(f \Longrightarrow g) \Longrightarrow ((f \vee h) \Longrightarrow (g \vee h))$

As usual $f \vee g$ is defined as $\neg(\neg f \wedge \neg g)$ and $f \Longrightarrow g$ stands for $\neg f \vee g$.

We postulate the following derivation rules for our axiomatic system: (all axioms are theorems)

MP Modus Ponens: If f and $f \Longrightarrow g$ are theorems then so is g

$\bar{\mathbf{N}}$ Subjective Necessity: If $f \in \mathcal{L}^{\mathcal{U}}$ is a theorem and $C_i f \in \mathcal{L}^{\mathcal{U}}$ then $C_i f$ is a theorem[5]

Whenever $C_i f, C_i g \in \mathcal{L}^{\mathcal{U}}$ we add the following axioms:

K $C_i(f \Longrightarrow g) \Longrightarrow (C_i f \Longrightarrow C_i g)$

D $C_i f \Longrightarrow \neg C_i \neg f$

4 $C_i f \Longrightarrow C_i C_i f$

U $C_i(C_i f \Longrightarrow f)$

and similarly for the axioms for the belief and utility operators we add all the axioms that relate to statements in $\mathcal{L}^{\mathcal{U}}$ following Heifetz and Mongin (2001) and Feinberg (2004a). Alternatively, we could have considered *all* the theorems in $\mathcal{L}^G$ (as in Feinberg 2004a) denoted $T \subset \mathcal{L}^G$ and define the theorems in $\mathcal{L}^{\mathcal{U}}$ simply as $T^{\mathcal{U}} = T \cap \mathcal{L}^{\mathcal{U}}$. We state the axioms above only to emphasize that they are applied in the same manner as in the global case when all players are fully aware of all statements. The main difference is the subjective necessity derivation rule. In particular, if the left hand side of an implication in any of the axioms K, D and 4 is in $\mathcal{L}^{\mathcal{U}}$ then so is the right hand side of the implication.

The motivation for the conditions defining an unawareness construction can be seen from the properties they imply for the language $\mathcal{L}^{\mathcal{U}}$. For example, consider the first condition. This condition requires that if a sequence of identities has an iterated awareness of an atom (or identity), then so does a partial sequence. This yields that if there is a statement in the language then any partial statement is also in the language. Where a partial statement is a statement used in construction of the original statement or awareness of a partial statement. For example if $C_i(a \wedge C_j(b \wedge C_k(a \vee \neg a)))$ is in the language so is $C_i(C_k(a \vee \neg a))$ since $C_k(a \vee \neg a)$ is used to construct the statement $a \wedge C_j(b \wedge C_k(a \vee \neg a))$ which is part of the language available to i.

We will sometimes refer to $\mathcal{L}^{\mathcal{U}}$ as the language with the axiom scheme generated by the awareness construction $\mathcal{U}$ as above and refer to $\mathcal{L}^{\{I,\alpha\}}$ as the language with context $\{I, \alpha\}$ with an axiom scheme as in Feinberg (2004a) which is unconstrained by an awareness construction.

3 Games with Unawareness

We now turn to games with unawareness. We begin with the FRPD with a grain of unawareness described in the introduction and provide an unawareness construction that captures

[5] This derivation rule is a relativization with respect to the unawareness construction (cf. Fagin and Halpern, 1987).

the verbal description of this game. The general definition of games with unawareness and their epistemic form follows. It is shown that for every game with unawareness for every state of awareness the definition yields yet another game with unawareness. This section concludes with the definition of the epistemic form of games with unawareness which allows us to analyze reasoning in such games using the same language that describes these game.

3.1 The FRPD with a Grain of Unawareness

Consider the following description of the FRPD with a grain of unawareness. Nature moves first choosing whether Alice is aware that the game being played is a repeated prisoner's dilemma (PD) or not. The probability that Alice is aware is $1-\varepsilon$ and with probability ε Alice is not aware that the players could do anything but cooperate. Unawareness of defection leads her to believe that the payoff is equal to the cooperative payoff. If an unaware Alice denoted AU_i observes an action other than the cooperative one she becomes fully aware that the game is the repeated PD and the identity that follows such a revelation is of the form A_j which denotes Alice's aware identities. In fact, if she observes defection Alice becomes aware of the whole situation including the fact that she was previously unaware. Obviously, the unaware Alice is not aware that she could possibly become aware in the future. We assume that Bob is fully aware throughout the game. Bob is initially uninformed (denoted $BU_{1,j}$) whether Alice is aware or not, but he knows that if she is unaware then by defecting he will *make* her aware.

In Figure 1 we see the game as it is perceived from the point of view of all aware identities, i.e., all identities other than identities AU_i. These include all the identities – decision points – of Bob and all the identities of Alice, either those that are aware of the game after Nature's move, or those that become aware once they observe defection by Bob. In Figure 1 Bob's decision points $BU_{i,j}$ are in the information sets where Bob may be uncertain whether Alice is aware or not. The identities $B_{i,j}$ correspond to information sets where Alice is known to be aware and the game being played is the repeated PD. Note, that Bob may know that Alice is aware yet still be uncertain whether she was initially aware. For example, if Bob defects in the first round and Alice cooperated in the first round then Bob knows that Alice will now be aware of the FRPD, but he is uncertain whether she initially cooperated because of unawareness or not. Alice's identities who are aware of the game are denoted A_i. Note that some of these identities appear when it is certain that the game is the PD and some (such as A_3) make their choices when Bob maybe uncertain as to Alice's awareness. The shaded part of the game displays that the continuation after the information set $BU_{2,i}$ is the same as depicted for the information set $BU_{1,i}$. The circled PD denotes that the game

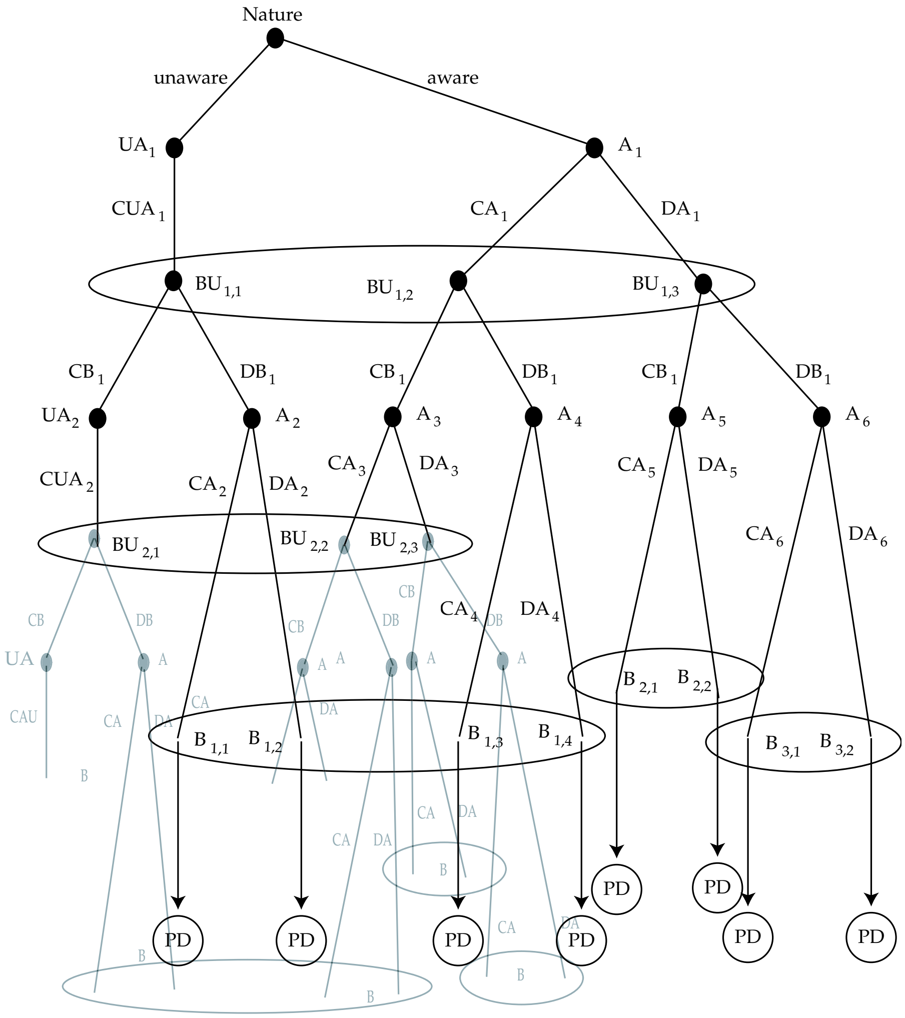

FIGURE 1: The game as viewed from the subjective view of the fully aware identities.

continues according to the FRPD from that point onwards although some information sets remain when Bob cannot determine whether Alice was initially aware or not. We omit the payoffs in order not to overcomplicate the graph.

It is important to note that the extensive form depicted in Figure 1 does not fully capture the game with unawareness since it does not provide information about unawareness at various decision points and reasoning about unawareness. In our example, the game as viewed by the unaware identities of Alice is depicted in Figure 2. Furthermore, all the *aware* identities – Bob's and Alice's – realize that the unaware Alice views the game as in Figure 2. Moreover, there is common confidence (as defined in Feinberg 2004a) among all aware identities that this is how the unaware identities view the game. The unaware identities, on the other hand, are confident that there is common confidence among them and the other identities in the game in Figure 2 that this is the the game at hand. Common confidence among the aware identities of this fact is the formal expression of their agreement of how the unaware identities view the game. At the end of Subsection 3.3 we show that this collection of statements about common confidence as to how the various identities view the game characterizes the epistemic form of this game with unawareness.

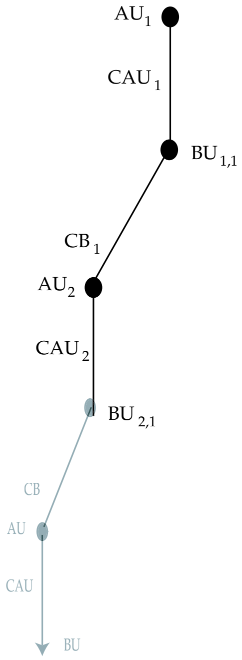

FIGURE 2: The game as seen by the unaware Alice.

This example captures the main properties required for representing a game with unawareness; there is an underlying game being played, some identities may be unaware of the full extent of the game while other identities may be uncertain about unawareness but may be aware of others' unawareness. The important feature of this example is that all identities view other identities and their actions as constituting a game, furthermore, they view others' view as constituting a game, and so on. We now translate the underlying game and an unawareness construction associated with it into an epistemic description of the game with unawareness. We do this for our example and show that the description of reasoning about unawareness in the game as suggested above is satisfied by the epistemic form of the game with unawareness. A general definition will follow.

Formally, we define the global context for the FRPD with unawareness game as the set of identities

$$I = \{A_1, A_2, ..., BU_{1,1}, BU_{1,2}, BU_{1,3}, BU_{2,1}, ..., B_{1,1}, B_{1,2}, ..., UA_1, UA_2, ...\}$$

and the set of names and actions

$$\begin{aligned}\alpha &= \{Alice, Bob, aware, unaware, CUA_1, CUA_2, ..., CA_1, DA_1, CA_2, DA_2, ... \\ &\quad , CBU_1, DBU_1, ..., CB_1, DB_1, ...\}.\end{aligned}$$

Note that this context differs from the context of the FRPD by adding Nature's move choosing *aware* or *unaware*, and adding the identities corresponding to an unaware Alice and a Bob that is uncertain about Alice's awareness.

The unawareness construction $\mathcal{U}$ is formally defined according to the description above as follows. Let $I_{(UA_1)} = \{UA_1, UA_2, ..., BU_{1,1}, BU_{2,1}, ...\}$, note that this set includes only Bob's identities $BU_{i,1}$ that follow an unaware identity of Alice in the game. We define $\mathcal{U} = \{I_\theta, \alpha_\theta\}_{\theta\in\Theta}$ by setting for each $\theta = (i_1, i_2, ..., i_n)$ the following:

$$I_\theta = \left\{ \begin{array}{ll} I & i_k \neq UA_j \text{ for all } k, j \\ I_{(UA_1)} & \begin{array}{l} \exists k, k \leq n \text{ with } i_k = UA_j \text{ and for such a } k \\ \text{that is minimal we have } i_{k+1}, .., i_n \in I_{(UA_1)} \end{array} \\ \emptyset & \text{otherwise} \end{array} \right\} \tag{3}$$

and

$$\alpha_\theta = \left\{ \begin{array}{ll} \alpha & i_k \neq UA_j \text{ for all } k, j \\ \{Alice, Bob, CUA_1, ..., CBU_1, ...\} & \begin{array}{l} \exists k, k \leq n \text{ with } i_k = UA_j \text{ and for such a } k \\ \text{that is minimal we have } i_{k+1}, .., i_n \in I_{(UA_1)} \end{array} \\ \emptyset & \text{otherwise.} \end{array} \right\} \tag{4}$$

The first part of the definition in (3) states that all identities other than Alice's unaware identities are aware that all identities other than Alice's unaware identities are aware that ... are aware of *all* identities. The second part states that Alice's unaware identities are only aware of the identities in $I_{(UA_1)}$ (the identities that appear in Figure 2), that Alice's unaware identities are aware that all identities in $I_{(UA_1)}$ are aware of the identities in $I_{(UA_1)}$, and so on. Moreover, all other identities are aware of each of these constructions of awareness for Alice's unaware identities. The third line of the definition states that these exhaust the awareness of all identities, i.e., any other order of awareness of identities is empty and will not be allowed in the language restricted to this unawareness construction.

The definition of awareness of atomic statements in (4) follows the same description. Identities are only aware of the actions of other identities if they are aware of those identities and these actions are present in the game as far as they are aware. It should be noted that

Alice's unaware identities are only aware of a single action for each of the identities $BU_{i,1}$ even though these identities have two actions in the original game. This captures the notion that these unaware identities of Alice are unaware of Bob's ability to defect[6].

The unawareness construction $\mathcal{U}$ defined in (3) and (4) generates a language $\mathcal{L}^{\mathcal{U}}$ in which the unaware identities of Alice UA_i reason only about themselves and about the identities $BU_{i,1}$ of Bob. Hence, which game is being played depends on the awareness of each identity. We define games with unawareness and their epistemic form in the next sections. The epistemic form is defined in the language $\mathcal{L}^{\mathcal{U}}$. For the FRPD with unawareness, we want the epistemic form to capture common confidence among all fully aware identities that the game follows Figure 1 and that the unaware identities are confident it follows Figure 2, and that the unaware identities are confident that there is common confidence (among the identities within their scope of awareness) that the game follows Figure 2.

3.2 Defining a Game with Unawareness

The primitives that are used for the definition of a game with unawareness are a finite extensive form[7] game $\Gamma = (N, H, P, \{\mu_h\}_{h \notin P^{-1}(N) \cup Z}, \{G_i\}_{i \in N}, \{\tilde{u}_i\}_{i \in N})$ and an unawareness construction $\mathcal{U}$. Recall that the extensive form game Γ is defined as a finite set of players N, a set of histories H which is a finite set of finite sequences closed under elimination of the last member of a sequence including the empty sequence, P is a function from a subset of $H \setminus Z$ to N with Z being all maximal sequences in H. For every non-terminal history $h \in H \setminus Z$ we define $A(h) = \{a | (h, a) \in H\}$ – the set of continuations of the history h. μ_h is a probability distribution over $A(h)$ for every history $h \in H \setminus Z$ that is not mapped under P – for every nature move. G_i is a partition of $P^{-1}(i)$ that satisfies $A(h) = A(h')$ whenever $h, h' \in H \setminus Z$ are in the same partition member and $\tilde{u}_i$ are utility functions over Z. We denote by $G(h) \subset H \setminus Z$ the set of histories that are in the same information set as h.

It is important to note that the game and the unawareness construction are not independent objects. For example, the global game in Figure 1 is not the repeated PD, it is a game in which Nature moves first to determine if Alice is aware or not. Hence, the game is already set to capture uncertainty about unawareness. On the other hand, there are further restrictions on the unawareness construction that we need to impose since we wish every identity to view the situation as a game with unawareness. This implies that we cannot arbitrarily allow the identities to be aware of any combination of atoms and identities, we must

[6] The game in Figure 1 does not fall into the class of games characterized in Feinberg (2004a) because it has decision points with a single possible action. We incorporate those decision points by assuming that the identities aware of only one action simply are confident that this action is taken.

[7] We note that the notation Γ is also used for the epistemic form of the extensive form game.

maintain the relationships between actions, names and identities that constitute a game and agree with the game Γ. For example, we do not wish to have an identity aware of actions $A(h)$ but not aware of the identity h to whom these actions belong. Such constraints are captured in the following definition.

Let Γ be an extensive form game $\Gamma = (N, H, P, \{\mu_h\}_{h \notin P^{-1}(N) \cup Z}, \{G_i\}_{i \in N}, \{\tilde{u}_i\}_{i \in N})$ with perfect agent recall and $\mathcal{U}$ denote the unawareness construction $\mathcal{U} = \{I_\theta, \alpha_\theta\}_{\theta \in \Theta}$.

Definition 3 *An extensive form game Γ with perfect agent recall and an unawareness construction $\mathcal{U}$ are called* a game Γ with unawareness $\mathcal{U}$ *if the following conditions hold*

1. *The context of the epistemic form of Γ coincides with the global context of $\mathcal{U}$.*

   Formally,
$$\{I, \alpha\} = \{\{P^{-1}(N)\}, N \cup \bigcup_{h \in H \setminus Z} A(h)\}. \tag{5}$$

2. *Conditions a. $-$ e. below hold for all levels of high order awareness:*

   (a) *Every identity is aware of her own action set:*
$$A(h) \subset \alpha_h \textit{ for all } h \in P^{-1}(N) \tag{6}$$

   (b) *Every identity is aware of the names of all the identities she is aware of:*
$$P(h') \in \alpha_h \textit{ whenever } h' \in I_h \tag{7}$$

   (c) *Every identity is aware of at least one action for each identity she is aware of:*
$$\alpha_h \cap A(h') \neq \emptyset \textit{ for all } h' \in I_h \tag{8}$$

   (d) *Every identity is aware of atomic statements that are either names or actions of identities she is aware of:*
$$\alpha_h \subset P(I_h) \cup \bigcup_{h' \in I_h} A(h') \cup \bigcup_{h'' \in (H \setminus Z) \setminus P^{-1}(N)} A(h'') \tag{9}$$

   (e) *The sets of identities and actions that an agent is aware of form a tree. This tree adds no terminal nodes to Γ.*

Formally, for all $h \in H$ such that there exists $h' \in H$ where $h' = (h, a)$ and $a \in A(h)$. We define

$$H_h = \left\{ (a_1, .., a_n) \;\middle|\; \begin{array}{c} a_i \in \alpha_h \ \forall i \\ (\bar{b}_1, a_1, \bar{b}_2, a_2, ...,, \bar{b}_n, a_n) \in H \\ \textit{where } \bar{b}_i \textit{ are possibly empty vectors of actions not in } \alpha_h \end{array} \right. \tag{10}$$

Recall that H is a collection of finite sequences closed under elimination of the last element of a sequence and note that $\emptyset \in H_h$. We require that:

i. H_h is a tree, i.e., $(a_1, .., a_n) \in H_h$ implies $(a_1, .., a_{n-1}) \in H_h$.

ii. No two distinct paths are identified in H_h. For every $(a_1, .., a_n) \in H_h$ there is a unique sequence of vectors $\bar{b}_1, \bar{b}_2, ..., \bar{b}_n$ such that $\bar{b}_i$ contains no member of α_h and $(\bar{b}_1, a_1, \bar{b}_2, a_2, .., \bar{b}_n, a_n) \in H$.

iii. Every branching of the tree occurs either at a decision point or at a nature move. If $(a_1, .., a_{n-1,}a_n) \in H_h$ and $(a_1, .., a_{n-1,}a'_n) \in H_h$ then $(\bar{b}_1, a_1, \bar{b}_2, a_2, .., \bar{b}_n, a_n) \in H$ and $(\bar{b}_1, a_1, \bar{b}_2, a_2, .., \bar{b}_n, a'_n) \in H$, i.e. the same unique sequence of vectors $\bar{b}_1, \bar{b}_2, ..., \bar{b}_n$ such that $\bar{b}_i$ contains no member of α_h leads to the decision point where a_n and a'_n are possible actions. Note that $\bar{b}_1, \bar{b}_2, ..., \bar{b}_{n-1}$ were the same because the sequence of vectors is unique for $(a_1, ..., a_{n-1})$ and since H_h is a tree.

iv. The set of identities coincides with the non-terminal histories in H_h that are not nature moves. For all $(a_1, .., a_n) \in H_h$ the unique sequence $(\bar{b}_1, a_1, \bar{b}_2, a_2, .., \bar{b}_n, a_n) \in H$ satisfies $(\bar{b}_1, a_1, \bar{b}_2, a_2, .., \bar{b}_n) \in I_h$ whenever $(\bar{b}_1, a_1, \bar{b}_2, a_2, .., \bar{b}_n) \in P^{-1}(N)$. Note that the opposite direction follows from (5) *and* (9) *and the definition of H_h.*

v. H_h introduces no new terminal nodes, i.e., if $(a_1, .., a_n) \in H_h$ is a terminal sequence then the unique sequence $(\bar{b}_1, a_1, \bar{b}_2, a_2, .., \bar{b}_n, a_n) \in Z$ is terminal in H.

The formal expression of having these conditions hold for any high order level of unawareness is provided in Appendix A.

The first condition states that the iterated awareness of all identities in the game about the actions and identities in the game is detailed in the unawareness construction and no other iterated awareness is.

The first four parts of condition 2 require that identities are aware of their actions, their names and that other identities have at least some action. This mostly simplifies notation since if an identity is unaware of one of her own actions then no one else can be aware that she is aware of such an action rendering such an action irrelevant. If an identity is unaware of her own name we will not be able to associate an epistemic form with the game as she perceives it and if an identity is aware of another identity but is unaware that the other identity participates in the game then the latter identity is not part of the context for the epistemic form of the game.

Part e of condition 2 contains two requirements. The first assures that each identity perceives the identities she is aware of and their action as constituting a game. The second requirement guarantees that the extensive form game Γ contains all the information needed for describing the game restricted to the identities state of awareness. Otherwise, if new terminal nodes are added then payoff at these nodes needs to be defined ex-ante since it might depend on actions of identities she is unaware of. We note that for games where it is natural to consider such unawareness the payoffs can be naturally defined, for example if a player is unaware of the full length of a repeated game she can still be assigned payoffs according to the number of stages she is aware of.

As an example for a game with unawareness we turn once again to the FRPD game with unawareness. Let Γ be the extensive form game depicted in Figure 1 and $\mathcal{U}$ be the unawareness construction defined in (3) and (4). Consider the unaware identity UA_1. From the unawareness construction $\mathcal{U}$ we have that she is only aware of the identities $I_{(UA_1)}$ and atomic statements $\alpha_{(UA_1)} = \{Alice, Bob, CUA_1, CUA_2, ..., CBU_1, CBU_2, ...\}$. It is easy to check that all the parts of condition 2 above are satisfied for $h = (UA_1)$ – here the history that corresponds to UA_1 is $(unaware)$. Furthermore, the tree $H_{(UA_1)}$ defined in (10) corresponds to the game tree depicted in Figure 2. Similarly, for every $\theta \in \bar{\Theta}$ the conditions specified in 2 hold for higher order iterations and the game tree from each awareness state corresponds either to the game in Figure 1 or the game in Figure 2.

We now derive how an awareness construction is viewed by each identity from her state of awareness and how each identity views each other identity's view of the unawareness construction as well as higher orders of awareness. We justify our definition of a game with unawareness by showing that every identity views the game with unawareness as a game with unawareness with respect to her state of unawareness. Similarly, higher order iterations yield a game with unawareness as well.

For a given unawareness construction $\mathcal{U}$ we define for each $\theta \in \bar{\Theta}$ the awareness construction $\mathcal{U}^\theta$ as follows:

$$\text{set } \Theta_\theta = \bigcup_{n=0}^{\infty} (I_\theta)^n \text{ and define } \mathcal{U}^\theta = \{(I^\theta_{\bar{\theta}}, \alpha^\theta_{\bar{\theta}})\}_{\bar{\theta}\in\Theta_\theta} \text{ where } I^\theta_{\bar{\theta}} = I_{\theta\hat{}\bar{\theta}} \text{ and } \alpha^\theta_{\bar{\theta}} = \alpha_{\theta\hat{}\bar{\theta}}. \tag{11}$$

For $\theta = (i_1, i_2, ..., i_n)$ the awareness construction $\mathcal{U}^{\theta}$ describes the awareness construction that i_1 finds that i_2 finds that ... that i_n considers from his state of awareness. Note that $\mathcal{U}^{(\emptyset)} = \mathcal{U}$. The proof of the following Lemma is left to the reader.

Lemma 4 *If $\mathcal{U}$ is an awareness construction then so is $\mathcal{U}^{\theta}$ for all $\theta \in \bar{\Theta}$.*

Let $\Gamma = (N, H, P, \{\mu_h\}_{h \notin P^{-1}(N) \cup Z}, \{G_i\}_{i \in N}, \{\tilde{u}_i\}_{i \in N})$ be an extensive form game. Throughout we will only consider extensive form games with agent perfect recall. Let $\mathcal{U}$ be an unawareness construction such that Γ is a game with unawareness $\mathcal{U}$. Consider $\theta \in \bar{\Theta}$. We define the game Γ^{θ}, $\Gamma^{\theta} = (P(I_{\theta}), H_{\theta}, P^{\theta}, \{\mu_h^{\theta}\}_{h \notin I_{\theta} \cup Z}, \{G_i^{\theta}\}_{i \in P(I_{\theta})}, \{\tilde{u}_i^{\theta}\}_{i \in P(I_{\theta})})$, as the game viewed in the state θ, where H_{θ} is defined in (36) (see Appendix A), and $P^{\theta}, G_i^{\theta}, \tilde{u}_i^{\theta}$ are the restrictions of $P, G_i, \tilde{u}_i$ to H_{θ}. Here μ_h^{θ} is the conditional distribution μ_h over the the actions in $A(h)$ of which there is awareness[8], i.e. $\mu_h^{\theta}(\cdot) = \mu_h(\cdot | A(h) \cap \alpha_{\theta})$. Let $A^{\theta}(h) = A(h) \cap \alpha_{\theta}$ denote the actions of h in Γ as viewed by θ. Also note that $\tilde{u}_i^{\theta}$ is well defined on the terminal nodes of H_{θ} since the conditions for higher order awareness in the definition of a game with unawareness require that H_{θ} introduces no new terminal nodes. This amounts to the natural restriction of the game Γ to the state of awareness θ.

The following proposition states that the definition of a game with unawareness induces a game with unawareness for any high order state of awareness. Hence, the description of a game with unawareness has every identity viewing the situation as a game with unawareness, every identity perceives that other identities perceive the game as a game with unawareness and so on.

Proposition 5 *If Γ is a game with unawareness $\mathcal{U}$ then for each $\theta \in \bar{\Theta}$ we have that Γ^{θ} is a game with unawareness $\mathcal{U}^{\theta}$.*

Proof. See Appendix B. ■

3.3 The Epistemic Form of a Game with Unawareness

The epistemic form of a game allows us to express the game in the same language used for reasoning about the game and its solutions. We ask that the definition of an epistemic form of a game with unawareness satisfy a number of immediate requirements. First and foremost, the epistemic form should be expressed in the language which reflects the unawareness construction, i.e., it should be stated as a collection of statements in $\mathcal{L}^{\mathcal{U}}$. We also require that the epistemic form of a game Γ with unawareness $\mathcal{U}$ coincide with the epistemic form of the *extensive* form game Γ when the unawareness construction $\mathcal{U}$ states that all identities

[8] We assume that for every Nature move each action is chosen with positive probability in Γ.

are aware of the same game, as well as iterated awareness of the same game, i.e. coincide in the case where there is full iterative awareness of the game. Finally, we require the epistemic form to satisfy consistency with respect to every state of awareness: We would like the epistemic form of the game Γ^θ with construction $\mathcal{U}^\theta$, when $\theta \in \bar{\Theta}$, to include the statements in the epistemic form of Γ with the unawareness construction $\mathcal{U}$ which are expressible in the language $\mathcal{L}^{\mathcal{U}^\theta}$. In other words, any statement in the description of the game Γ with unawareness $\mathcal{U}$ which can be expressed in the restricted language $\mathcal{L}^{\mathcal{U}^\theta}$ should appear in the description of the game Γ^θ with unawareness $\mathcal{U}^\theta$.

The epistemic form for a game with unawareness we purpose closely mimics the epistemic form of a dynamic game presented in Feinberg (2004a). The epistemic form of an extensive form game was defined as the epistemic and logical closure of the epistemic description of the building blocks of the game – actions, identities, Nature moves, utilities, information sets and dynamic structure. Similarly, the epistemic form of the game with unawareness is defined as the epistemic and logical closure of the same building blocks in the game with unawareness, where the closure is constructed in the restricted language $\mathcal{L}^{\mathcal{U}}$. We epistemically describe actions, identities, Nature moves, utilities, information sets and the dynamic structure as it is perceived at every state of awareness and the closure of all these descriptions constitutes the epistemic form of a game with unawareness.

Consider an extensive form game $\Gamma = (N, H, P, \{\mu_h\}_{h \notin P^{-1}(N) \cup Z}, \{G_i\}_{i \in N}, \{\tilde{u}_i\}_{i \in N})$ with perfect agent recall and an unawareness construction $\mathcal{U} = \{I_\theta, \alpha_\theta\}_{\theta \in \Theta}$. The epistemic form of Γ (see Feinberg 2004a) is the epistemic and logical closure of the following sets of statements:

$$C_h(i \wedge \neg j) \qquad \forall h \in P^{-1}(N), i = P(h), j \in N \setminus \{i\} \qquad \text{(naming)} \tag{12}$$

$$C_h a \iff a \qquad \forall h \in P^{-1}(N), \forall a \in A(h) \qquad \text{(h's actions)} \tag{13}$$

$$a \Longrightarrow \neg a' \qquad \forall h \in H \setminus Z, \forall a, a' \in A(h) \qquad \text{(actions are precise)} \tag{14}$$

$$\bigvee_{a \in A(h)} a \qquad \forall h \in H \setminus Z \qquad \text{(action sets are proper)} \tag{15}$$

$$C_{\bar{h}}(C_h f \iff C_{h'} f) \qquad \forall \bar{h}, h \in P^{-1}(N), \forall h' \in G(h) \qquad \text{(information sets)} \tag{16}$$

$$C_h(\bigvee_{h' \in G(h)} s(h')) \qquad \forall h \in P^{-1}(N) \qquad \text{(dynamic knowledge structure)} \tag{17}$$

$$U_h^r(s(h')) \qquad \forall h \in P^{-1}(N), \tilde{u}_{P(h)}(h') = r, \forall h' \in Z \qquad \text{(utilities)} \tag{18}$$

$$
\begin{aligned}
&P_h^{\alpha}(\pi|_a) \iff P_h^{\beta}(\pi|_b) \qquad \text{whenever } \mu_{\bar{h}}(a)\beta = \mu_{\bar{h}}(b)\alpha, \\
&\forall \pi, \forall h \in \mathcal{I}, \text{ every Nature move } \bar{h} \text{ with acts } a, b \in A(\bar{h}) \\
&\text{such that } \pi|_a \Longrightarrow \bigvee_{h' \in G(h)} s(h') \text{ and } \pi|_b \Longrightarrow \bigvee_{h' \in G(h)} s(h'),
\end{aligned}
\tag{19}
$$

where π denotes a pure strategy profile $\pi = \bigwedge_{h \in H \setminus Z} a_h$ such that $a_h = a_{h'}$ for $h' \in G(h)$, and $\pi|_a$ denotes the conjunction generated by replacing the action in π for the corresponding members of $G(h)$ with a.

For every $\theta = (h_1, ..., h_n)$ such that $\theta \neq \emptyset$ and $I_\theta \neq \emptyset$ we denote by γ^θ the collection of statements below in (20) − (27):

$$C_{h_1}...C_{h_n}(i \wedge \neg j) \qquad i = P(h_n), j \in P(I_\theta) \setminus \{i\} \qquad \text{(naming)} \tag{20}$$

$$C_{h_1}...C_{h_n}(C_h a \iff a) \qquad \forall h \in I_\theta, \forall a \in A^\theta(h) \qquad \text{(}h\text{'s actions)} \tag{21}$$

$$C_{h_1}...C_{h_n}(a \Longrightarrow \neg a') \qquad \forall h \in H_\theta \setminus Z, \forall a, a' \in A^\theta(h) \qquad \text{(actions are precise)} \tag{22}$$

$$C_{h_1}...C_{h_n}\left(\bigvee_{a \in A^\theta(h)} a\right) \qquad \forall h \in H_\theta \setminus Z \qquad \text{(action sets are proper)} \tag{23}$$

$$C_{h_1}...C_{h_n}(C_h f \iff C_{h'} f) \qquad \forall h \in I_\theta, \forall h' \in G^\theta(h), \forall f \in \mathcal{L}^{\mathcal{U}^{\theta\hat{}h}} \qquad \text{(information sets)} \tag{24}$$

note that (24) implies that $\mathcal{L}^{\mathcal{U}^{\theta\hat{}h}} = \mathcal{L}^{\mathcal{U}^{\theta\hat{}h'}}$ for two identities for which there is awareness in θ which are in the same information set. Hence, from every state of awareness θ, h and h' are aware of the same identities, actions and names, hence are aware of the same game. In particular, the two identities h and h' are aware of each other. This interpretation of an information set states that, not only are the two identities seen to have the same information, but they are seen to have the same state of awareness including the awareness of being in an information set.[9]

$$C_{h_1}...C_{h_n}(\bigvee_{h' \in G^\theta(h_n)} s(h')) \qquad \text{(dynamic knowledge structure).} \tag{25}$$

Note that $s(h')$ for $h' \in G^\theta(h_n)$ is the conjunction of actions in h' as a history in H_θ.

$$C_{h_1}...C_{h_{n-1}}\left(U_{h_n}^r(s(h'))\right) \qquad \text{for } \tilde{u}_{P^\theta(h_n)}^\theta(h') = r, \text{ for every terminal } h' \text{ in } H_\theta \qquad \text{(utilities)} \tag{26}$$

[9] Note that this definition allows for the existence of identities that are unaware of an information set.

$$
\begin{gathered}
C_{h_1}...C_{h_{n-1}}\left(P^{\alpha}_{h_n}(\pi|_a) \iff P^{\beta}_{h_n}(\pi|_b)\right) \qquad \text{whenever } \mu_{\bar{h}}(a)\beta = \mu_{\bar{h}}(b)\alpha, \\
\forall\pi,\ \text{every Nature move } \bar{h} \in H_\theta \text{ with acts } a, b \in A^{\theta}(\bar{h}) \\
\text{such that } \pi|_a \Longrightarrow \bigvee_{h' \in G^{\theta}(h)} s(h') \text{ and } \pi|_b \Longrightarrow \bigvee_{h' \in G^{\theta}(h)} s(h'),
\end{gathered} \tag{27}
$$

where π is as in (19) with H_θ and $G^\theta(h)$ replacing H and $G(h)$ respectively.

For $\theta = \emptyset$ we define γ^θ to be the collection of statements in $(21), (22)$ and (23).

Definition 6 *The epistemic form of a game Γ with unawareness $\mathcal{U}$ is the epistemic and logical closure in $\mathcal{L}^{\mathcal{U}}$ of the union of sets of statements γ^θ:*

$$\bigcup_{\theta \in \bar{\Theta}} \gamma^\theta. \tag{28}$$

The epistemic form is denoted by $\Gamma^{\mathcal{U}}$.

We first note the following:

Lemma 7 *An epistemic form of a game with unawareness is well defined, in the sense that all the statements in* (28) *are in $\mathcal{L}^{\mathcal{U}}$.*

Proof. See Appendix B. ■

This establishes our first requirement from the epistemic form. The second requirement is also satisfied due to the following proposition:

Proposition 8 *Consider an extensive form game Γ with unawareness construction $\mathcal{U}$ where all identities are fully aware of the game, as are all higher order awareness, i.e. $\{I_\theta, \alpha_\theta\} = \{I, \alpha\}$ for all $\theta \in \Theta$. The epistemic form of the game Γ with unawareness $\mathcal{U}$ coincides with the epistemic form of Γ.*

Proof. For a game Γ with unawareness $\mathcal{U}$ where there is common awareness of all identities and atomic statements we have that $\Gamma^\theta = \Gamma$ for all θ. Note that $I_\theta \neq \emptyset$ for all θ in this case. In particular, the conditions $(20) - (27)$ are applied for the same set of histories, actions, nature moves, terminal histories etc. Combining with $\gamma^\emptyset$, the collection (28) which resides in the global language $\mathcal{L}^{\{I,\alpha\}}$ is identical to the collection of statement corresponding to common confidence of the statements in $(12) - (19)$. Since $\mathcal{L}^{\{I,\alpha\}} = \mathcal{L}^{\mathcal{U}}$ in this case, we have that the epistemic form of Γ generated by the epistemic and logical closure of $(12) - (19)$ coincides with the epistemic and logical closure of (28) for Γ with unawareness $\mathcal{U}$. ■

We turn to showing that our definition of the epistemic form of a game with unawareness satisfies the central requirement: the epistemic form describes a game with unawareness from every identity's view point, as well as higher order states of awareness. In other words, any statement in the epistemic description of the game that is allowed in the restricted language for θ, is also part of the description of the game with unawareness as viewed from the state of awareness θ.

Proposition 9 *Let Γ be an extensive form game with unawareness $\mathcal{U}$ and let $\theta \in \bar{\Theta}$, then*

$$\Gamma^{\mathcal{U}} \cap \mathcal{L}^{\mathcal{U}^{\theta}} \subset \left(\Gamma^{\theta}\right)^{\mathcal{U}^{\theta}}. \tag{29}$$

Proof. See Appendix B. ■

We have established that the epistemic form of a game with unawareness incorporates the description of the game with unawareness as viewed by each and every identity given her awareness, as well as capturing how each identity views others' view of the game with unawareness.

From the description of the FRPD game Γ with the unawareness construction $\mathcal{U}$ defined in $(3)-(4)$, the game Γ^{θ} is either the game depicted in Figure 1 or the game depicted in Figure 2 whenever $I_{\theta} \neq \emptyset$. We now relate the epistemic form of the FRPD with unawareness to these games. This establishes the informal claims made in Subsection 3.1 in which we stated how the players view the game as one of the two forms in Figures 1 and 2, and that higher order states of awareness also view the game in one of these two forms.

Proposition 10 *Let $\Gamma^{\mathcal{U}}$ denote the epistemic form of the game Γ depicted in Figure 1 with unawareness $\mathcal{U}$ defined in $(3)-(4)$. $\Gamma^{\mathcal{U}}$ is logically equivalent in $\mathcal{L}^{\mathcal{U}}$ to the union of the following collection of statements which belong to $\mathcal{L}^{\mathcal{U}}$:*

(a) The identities $\{UA_i\}_i$ view the game as the (epistemic form of the) game $\Gamma^{(UA_1)}$ depicted in Figure 2.

(b) The identities $\{UA_i\}_i$ believe that there is common confidence among the identities in $I_{(UA_1)}$ of the (epistemic form of the) game $\Gamma^{(UA_1)}$.

(c) All other identities $j \in I \setminus \{UA_i\}_i$ view the game as $\Gamma^{\mathcal{U}}$.

(d) There is common confidence among the identities in $I \setminus \{UA_i\}_i$ in the statements in $(a)-(c)$ above.

Proof. See Appendix B. ■

4 A Grain of Unawareness Leading to Cooperation

In order to derive a conclusion on the players behavior in games with unawareness we need to define a solution concept for these games. We do so using an epistemic characterization. We consider the epistemic characterization of a solution to extensive form games and and define the solution to games with unawareness using the same epistemic characterization in the language restricted by the unawareness construction. The behavior (or beliefs) corresponding to the epistemic characterization in the constrained language define the solution for games with unawareness.

Using the epistemic characterization we provide an extension of sequential equilibria to games with unawareness and show that the epistemic conditions on reasoning that lead to a sequential equilibrium imply that the *aware* players in the FRPD with unawareness follow the exact same behavior as the *rational* players in the FRPD with irrationality. Hence, a grain of unawareness leads to cooperation in the FRPD.

Recall that in the subjective framework (as in other epistemic frameworks) an epistemic characterization of a solution is a collection of sets of statements. Such a collection Υ characterizes a solution for a class of games. In other words, for each game Γ in the relevant class of games with a collection of sets of statements Υ, one considers whether each set of statements $\upsilon \in \Upsilon$ is consistent with the epistemic game form Γ. When υ and Γ are logically consistent then the conjectures (beliefs) consistent with *both* are included in the solution. A characterization usually contains a collection of statements regrading the players' rationality, beliefs about others' rationality and behavior and so on. In the case of sequential equilibria the characterization in Corollary 3 in Feinberg (2004b) states that an assessment (μ, σ) is an equilibrium with sequential rationality and convex structural consistency for the game Γ if and only if the following epistemic statements are logically consistent:

a) The statements that correspond to the epistemic form of Γ.
b) The statements that describe for every h that h's conjectures about others strongly follow (μ, σ).
c) Common confidence of all the beliefs stated in b).
d) Common confidence among σ-possible identities that *all* identities are rational.

Also recall that Theorem 2 in Feinberg (2004b) adds the consistency of the following:
e) Common confidence (among all identities) that *all* identities find future identities to be conditionally rational – $CFCR$.

This epistemic characterization of sequential equilibria is extended to games with unawareness by considering each of these epistemic conditions within the language for unawareness. By mapping the epistemic characterization to conditions that respect the unawareness construction we derive a solution for the game with unawareness.

Definition 11 *An assessment* (μ, σ) *is called* an extended sequential equilibrium *for a game with unawareness* $\Gamma^{\mathcal{U}}$ *if the the following epistemic conditions are consistent in* $\mathcal{L}^{\mathcal{U}}$*:*

a) The statements that correspond to the epistemic form of $\Gamma^{\mathcal{U}}$*.*
b) The statements that describe for every h *that* h*'s conjectures about others strongly follow* (μ, σ) *when* (μ, σ) *is conditioned on the identities and actions that* h *is aware of.*
c) For all $\theta = (h_1, ..., h_n)$ *with* $h \in I_\theta$ *(in particular* $I_\theta \neq \emptyset$*) we have*

$$C_{h_1} C_{h_2} ... C_{h_n} \left(\begin{array}{r} h\text{'s conjectures follow } (\mu, \sigma) \\ \text{conditioned on } (I_{\theta\hat{}h}, \alpha_{\theta\hat{}h}) \end{array} \right). \tag{30}$$

d) For all $\theta = (h_1, ..., h_n)$ *with* $h \in I_\theta$ *such that* $h_1, ..., h_n, h$ *are all* σ*-possible, we have*

$$C_{h_1} C_{h_2} ... C_{h_n} (h \text{ is rational}). \tag{31}$$

The extended solution constrains the epistemic characterization to the unawareness construction. It requires that an identity only reason and conjecture about identities and actions she is aware of. Hence, conjectures are conditioned on identities and actions within the scope of awareness as is confidence in rationality. For example, the statement "h is rational" in (31) stands for the statements that describe h not choosing a dominated action in the action set $A^\theta(h)$ which is restricted to the state of awareness θ and with respect to his conjecture as perceived in (30).

We note that since the epistemic characterization of a solution for extensive form games need not be unique, the extension to games with unawareness need not be unique. For our purpose we chose the minimal epistemic conditions that characterize sequential equilibria as provided in Feinberg (2004b). If additional epistemic conditions are added, such as adding the collection of statements *e*) from Theorem 2 in Feinberg (2004b), our result will still hold since additional restrictions do not enlarge the set of conjectures the solution defines.

Our main result in this section states that a grain of unawareness can generate cooperation in the FRPD much like a grain of irrationality does. We note however, that although behavior is identical, namely, the aware Alice is mimicking the unaware Alice, just like the rational type mimics the irrational type, Bob cooperates for different reasons in our framework. In the case of a grain of irrationality, it is the irrational strategy of Tit-for-Tat or grim trigger (defect forever once the opponent does) which forces Bob to consider cooperation. Only such *special* irrational types would lead to the cooperative outcome. In contrast, with a grain of unawareness Bob's "punishment" for defection comes from revealing this action to the unaware Alice and not from the ad-hoc irrational reaction to defection. If Bob defects then he reveals the game as the FRPD and defection occurs throughout. In this sense the threat comes from the nature of the game once Alice becomes fully aware of it and not from an ad-hoc irrational behavior.

Theorem 12 *The behavior of the aware identities in all the extended sequential equilibria of the FRPD with unawareness is identical to the behavior of the rational types in the FRPD with a grain of irrationality as studied by Kreps, Milgrom, Roberts and Wilson (1982). In particular, the aware types will cooperate except for a finite number of periods which does not depend on the overall length of the game.*

Proof. See Appendix B. ■

5 Final Remarks

We have provided a framework for reasoning about unawareness, defined games with unawareness and provided a method to extend a solution to such games. This allowed us to show that a grain of unawareness can lead to the same impact on behavior as a grain of irrationality did for the FRPD.

The question arises in what way does unawareness differ from irrationality?

Conceptually, unawareness is perceived as a form of bounded rationality. Bounded rationality refers to a variety of relaxations on the assumptions of a rational economic agent. Among these reductions we find costly decision making, the inability to deduce all logical implications (know all theorems), mistakes that occur with small probability (trembling hand), and unawareness of the full scope of the decision situation at hand. Hence, unawareness allows for optimal behavior but in a restricted environment and without realizing the possibility that the full scope of the situation is unobserved.

Operationally, adding unawareness to a game only allows us to restrict the game as viewed by the reasoning identities. While one can generate any pure strategy in an extensive form game by limiting a player's awareness to the unique action determined at each stage by the pure strategy, some of these restrictions can be ruled out as nonsensical. For example, to implement the Tit-for-Tat strategy in the FRPD with unawareness, we would need Alice to be aware of defection only if Bob defected in the previous round and to be only aware of cooperation if he didn't. In such a case Alice must be very forgetful and Bob can strategically determine which action is available to Alice. If, as we assumed here, Alice recognizes the game is symmetric and repeated, and she always expects to have an identical set of actions in each round for both Bob and her, then she can only be unaware of cooperation or defection as long as the other action is not revealed. Hence, cooperation emerges not as a result of specifically tailored irrational behavior, but as one of two possible simple cases of first order unawareness when the game is known to be symmetric and repeated.

A natural extension to our result is the study of the impact of higher order of unawareness. As was shown in Milgrom and Roberts (1982, Appendix B) (see also Kreps, Milgrom, Roberts and Wilson,1982 for the FRPD), a grain of uncertainty about the uncertainty ... about the uncertainty of rationality suffices for the generating the reputational effect. It is not clear whether higher order unawareness possess this property.

It is quite difficult to identify which irrational behavior leads to cooperation in the FRPD, or in many other applications of a grain of irrationality and reputation. Although Aumann and Sorin (1989) have been able to show that full support of the irrational types viewed as automata leads to cooperation in mutual interest games, such general results are rare. In contrast, a repeated game where players remember previous and observed actions, and recognize that the game is repeated, hence the same action sets are available at every period, allows for a much smaller set of variations of a grain of unawareness. Hence, a grain of unawareness might be more amicable to a characterization of its induced behavior.

Finally, we recall that unawareness differed from irrationality in the manner in which cooperation emerged. In our example unawareness turns to awareness if and only if an action is observed. Hence it is Bob's defection that causes the unaware Alice to become aware. Once Alice becomes aware, the game turns into a standard FRPD from that point onwards. In fact, it is this rational defection by Alice when she is aware that causes Bob to cooperate in the first place. The threat that leads to his cooperation is not some arbitrary irrational behavior tailored for that purpose, it is simply the threat of revealing the full details of the actual game to a possible unaware Alice, a revelation that leads to full awareness and common confidence that there is awareness. In particular, Bob strategic impact on Alice's state of awareness drives our result.

References

Aumann, R. J., and Brandenburger, A., 1995. Epistemic conditions for Nash equilibrium. Econometrica 63(5), 1161–1180.

Aumann, R.J., and Sorin, S., 1989. Cooperation and bounded recall, Games and Economic Behavior, 1, 5-39.

Crawford, V. P., and Haller, H., 1990. Learning How to Cooperate: Optimal Play in Repeated Coordination Games. Econometrica 58(3), 571-595.

Dekel, E., Lipman, B. L., and Rustichini, A., 1998. Standard State-Space Models Preclude Unawareness. Econometrica 66(1), 159-174.

Ewerhart, C., 2001. Heterogeneous Awareness and the Possibility of Agreement, Discussion paper 01-30, Sonderforschungsbereich 504, Universitt Mannheim.

Fagin, R., and Halpern, J. Y., 1987. Belief, awareness, and limited reasoning. Artificial Intelligence 34(1), 39–76.

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

Feinberg, Y., 2004a. Subjective reasoning - dynamic games. Games and Economic Behavior, forthcoming.

Feinberg, Y., 2004b. Subjective reasoning - solutions. Games and Economic Behavior, forthcoming.

Fudenberg, D., and Maskin, E., 1986. The Folk Theorem for Repeated Games with Discounting and Incomplete Information. Econometrica, 54, 533-54.

Fudenberg, D., and Tirole, J., 1991. Game Theory. MIT Press, Cambridge, MA.

Halpern, J. Y., 2001. Alternative Semantics for Unawareness. Games and Economic Behavior 37(2), 321-339

Halpern, J. Y., and Rêgo, L. C., 2005. Interactive Unawareness Revisited. Mimeo.

Heifetz, A., Meier, M., and Schipper, B. C., 2004. Interactive Unawareness. Journal of Economic Theory, forthcoming.

Kawamura, E., 2004. Competitive Equilibrium with Unawareness in Economies with Production. Journal of Economic Theory, forthcoming.

Kreps, D. M., Milgrom, P., Roberts, J., and Wilson, R. (1982). Rational cooperation in the finitely repeated prisoners' dilemma. J. Econom. Theory 27(2), 245–252.

Kreps, D. M., and Wilson, R. (1982). Reputation and imperfect information. J. Econom. Theory 27(2), 253-279.

Levesque, H. J. (1984). A logic of implicit and explicit belief. In Proceedings of the Fourth National Conference on Artificial Intelligence (AAAI-84), pages 198-202, Austin, TX.

Li, J., 2004. Unawareness. unpublished manuscript.

Milgrom, P., and Roberts, J. 1982. Predation, reputation and entry deterrence, Journal of Economic Theory 27, 280-312.

Modica, S., and Rustichini, A., 1994. Awareness and partitional information structures. Theory and Decision 37(1), 107–124.

Modica, S., and Rustichini, A., 1999. Unawareness and partitional information structures. Games and Economic Behavior. 27(2), 265–298.

Modica, S., Rustichini, A., and Tallon, J., 1998. Unawareness and Bankruptcy: A General Equilibrium Model. Economic Theory 12, 259–292.

Appendix A

The definition of the conditions for high order awareness for the definition of a game with unawareness is given below:

For every $\theta = (h_1, ..., h_n)$ such that $I_\theta \neq \emptyset$ we require the following.

A generalization of (6):

$$A(h_n) \cap \alpha_{(h_1,...,h_{n-1})} \subset \alpha_\theta. \tag{32}$$

A generalization of (7):

$$P(h) \in \alpha_\theta \ \forall h \in I_\theta. \tag{33}$$

The association of at least one action for each identity – condition (8):

$$\alpha_\theta \cap A(h) \neq \emptyset \ \forall h \in I_\theta, \tag{34}$$

note that (34) implies that the set $A(h_n) \cap \alpha_{(h_1,...,h_{n-1})}$ which appears in (32) is not empty. We generalize (9) with:

$$\alpha_\theta \subset P(I_\theta) \cup \bigcup_{h \in I_\theta} A(h) \cup \bigcup_{h' \in (H \backslash Z) \backslash P^{-1}(N)} A(h'). \tag{35}$$

We recursively define

$$H_\theta = \left\{ (a_1, .., a_m) \;\middle|\; \begin{array}{c} a_i \in \alpha_\theta \ \forall i \\ (\bar{b}_1, a_1, \bar{b}_2, a_2, ..., , \bar{b}_m, a_m) \in H_{(h_1,...,h_{n-1})} \\ \text{where } \bar{b}_i \text{ are possibly empty vectors of actions not in } \alpha_\theta, \end{array} \right. \tag{36}$$

and require that

i. H_θ is a tree:

$$(a_1, .., a_n) \in H_\theta \text{ implies } (a_1, .., a_{n-1}) \in H_\theta. \tag{37}$$

ii. No two distinct paths are identified in H_θ:

$$\begin{gathered} \forall (a_1, .., a_n) \in H_\theta \text{ there is a unique sequence of vectors } \bar{b}_1, \bar{b}_2, ..., \bar{b}_n \\ \text{such that } \bar{b}_i \text{ contains no member of } \alpha_\theta \text{ and } (\bar{b}_1, a_1, \bar{b}_2, a_2, .., \bar{b}_n, a_n) \in H_{(h_1,...,h_{n-1})}. \end{gathered} \tag{38}$$

iii. Every branching of the tree occurs at a decision point or nature move:

$$\begin{aligned} \text{If } (a_1, .., a_{n-1,} a_n) &\in H_\theta \text{ and } (a_1, .., a_{n-1}, a_n') \in H_\theta \text{ then} \\ (\bar{b}_1, a_1, \bar{b}_2, a_2, .., \bar{b}_n, a_n) &\in H_{(h_1,...,h_{n-1})} \text{ and } (\bar{b}_1, a_1, \bar{b}_2, a_2, .., \bar{b}_n, a_n') \in H_{(h_1,...,h_{n-1})}. \end{aligned} \tag{39}$$

iv. The set of identities coincides with the non-terminal histories in H_θ that are not nature moves:

$$\begin{gathered}\forall(a_1,..,a_n) \in H_\theta \text{ the unique sequence } (\bar{b}_1, a_1, \bar{b}_2, a_2, .., \bar{b}_n, a_n) \in H_{(h_1,...,h_{n-1})} \\ \text{satisfies } (\bar{b}_1, a_1, \bar{b}_2, a_2, .., \bar{b}_n) \in I_\theta \text{ whenever } (\bar{b}_1, a_1, \bar{b}_2, a_2, .., \bar{b}_n) \in P^{-1}(N).\end{gathered} \tag{40}$$

v. H_θ introduces no new terminal nodes:

$$\begin{gathered}\text{If } (a_1,..,a_n) \in H_\theta \text{ is a terminal sequence then the unique} \\ \text{sequence } (\bar{b}_1, a_1, \bar{b}_2, a_2, .., \bar{b}_n, a_n) \in Z \text{ is terminal in } H_{(h_1,...,h_{n-1})}.\end{gathered} \tag{41}$$

Appendix B

Proof of Proposition 5. Let $\theta \in \bar{\Theta}$, i.e. $I_\theta \neq \emptyset$. Consider the unawareness construction $\mathcal{U}^\theta$ as defined in (11) and Γ^θ as defined in the text. We need to show that the conditions in the definition of a game with unawareness hold for $(\Gamma^\theta, \mathcal{U}^\theta)$. It suffices to show that condition 1 in the definition holds and that for every $\bar{\theta}$ such that $I^\theta_{\bar{\theta}} \neq \emptyset$ the conditions in Appendix A hold.

From the definition of H_θ in (36) we have that all the actions in the game Γ^θ belong to α_θ. Similarly, the identities are all in I_θ and hence by (33) applied to Γ the names of these identities are also in α_θ. We only have actions and names from Γ^θ in α_θ due to (35) applied to Γ. By the definition of Γ^θ the set of identities in Γ^θ is $I_\theta = \left(P^\theta\right)^{-1}(P(I_\theta))$ and we conclude that

$$\{I_\theta, \alpha_\theta\} = \{\{\left(P^\theta\right)^{-1}(P(I_\theta))\}, P(I_\theta) \cup \bigcup_{h \in H_\theta \setminus Z_\theta} A^\theta(h)\}. \tag{42}$$

Here we denote the terminal elements of H_θ by Z_θ. Note that these correspond to a subset of Z. We conclude that condition 1 holds. Also recall that $A^\theta(h) = A(h) \cap \alpha_\theta$.

We now show that for every $\bar{\theta} = (h_1, ..., h_n)$ such that $I^\theta_{\bar{\theta}} \neq \emptyset$ we have that $(32) - (35)$ and $(37) - (41)$ hold for Γ^θ and $\mathcal{U}^\theta$.

From the definition in (11) we have that $I^\theta_{\bar{\theta}} = I_{\theta\hat{}\bar{\theta}} \neq \emptyset$ hence since Γ is a game with unawareness $\mathcal{U}$ we have that for $\theta\hat{}\bar{\theta} = \theta\hat{}(h_1, ..., h_n)$ applying (32) for Γ yields

$$A(h_n) \cap \alpha_{\theta\hat{}(h_1,...,h_{n-1})} \subset \alpha_{\theta\hat{}\bar{\theta}} \tag{43}$$

and since $\alpha_{\theta\hat{}(h_1,...,h_{n-1})} = \alpha^{\theta}_{(h_1,...,h_{n-1})}$ and $\alpha_{\theta\hat{}\bar{\theta}} = \alpha^{\theta}_{\bar{\theta}} \subset \alpha_{\theta}$ we have

$$A^{\theta}(h_n) \cap \alpha^{\theta}_{(h_1,...,h_{n-1})} \subset \alpha^{\theta}_{\bar{\theta}} \tag{44}$$

which implies that (32) holds for the game Γ^{θ} with awareness construction $\mathcal{U}^{\theta}$ since the actions of h_n in Γ^{θ} are a subset of $A(h_n)$ as defined in Γ.

Consider $h \in I^{\theta}_{\bar{\theta}} = I_{\theta\hat{}\bar{\theta}}$ which implies $P(h) \in \alpha_{\theta\hat{}\bar{\theta}}$ according to (33) for Γ. Since $h \in I^{\theta}_{\bar{\theta}} \subset I_{\theta}$ we have $P^{\theta}(h)$ is well defined and $P^{\theta}(h) \in \alpha_{\theta\hat{}\bar{\theta}} = \alpha^{\theta}_{\bar{\theta}}$ so (33) holds for Γ^{θ} with $\mathcal{U}^{\theta}$.

Let $h \in I^{\theta}_{\bar{\theta}} = I_{\theta\hat{}\bar{\theta}}$, then $\alpha_{\theta\hat{}\bar{\theta}} \cap A(h) \neq \emptyset$. But since $\alpha_{\theta\hat{}\bar{\theta}} \subset \alpha_{\theta}$ we have $\alpha^{\theta}_{\bar{\theta}} \cap A^{\theta}(h) \neq \emptyset$ and (34) holds.

Consider the game $\Gamma^{\theta\hat{}\bar{\theta}}$. Applying (42) to this game we have that

$$\alpha_{\theta\hat{}\bar{\theta}} = P(I_{\theta\hat{}\bar{\theta}}) \cup \bigcup_{h \in H_{\theta\hat{}\bar{\theta}} \setminus Z} A^{\theta\hat{}\bar{\theta}}(h) = P(I_{\theta\hat{}\bar{\theta}}) \cup \bigcup_{h \in I_{\theta\hat{}\bar{\theta}}} A^{\theta\hat{}\bar{\theta}}(h) \cup \bigcup_{h' \in (H_{\theta\hat{}\bar{\theta}} \setminus Z) \setminus P^{-1}(N)} A^{\theta\hat{}\bar{\theta}}(h') \tag{45}$$

and since $I_{\theta\hat{}\bar{\theta}} \subset I_{\theta}$ we have that $P(I_{\theta\hat{}\bar{\theta}}) = P^{\theta}(I_{\theta\hat{}\bar{\theta}})$. We also have

$$(H_{\theta\hat{}\bar{\theta}} \setminus Z) \setminus P^{-1}(N) \subset (H_{\theta} \setminus Z) \setminus \left(P^{\theta}\right)^{-1} (P(I_{\theta})) \tag{46}$$

and $A^{\theta\hat{}\bar{\theta}}(h) \subset A^{\theta}(h)$. Hence we conclude

$$\alpha^{\theta}_{\bar{\theta}} \subset P^{\theta}(I^{\theta}_{\bar{\theta}}) \cup \bigcup_{h \in I^{\theta}_{\bar{\theta}}} A^{\theta}(h) \cup \bigcup_{h' \in (H_{\theta} \setminus Z) \setminus \left(P^{\theta}\right)^{-1} (P(I_{\theta}))} A^{\theta}(h') \tag{47}$$

and condition (35) holds.

For $\bar{\theta} = (h_1, ..., h_n)$ with $I^{\theta}_{\bar{\theta}} \neq \emptyset$, the definition of the induced game tree for the game Γ^{θ} with respect to $\bar{\theta}$ follows (36):

$$(H_{\theta})_{\bar{\theta}} = \left\{ (a_1, .., a_m) \;\middle|\; \begin{array}{c} a_i \in \alpha^{\theta}_{\bar{\theta}} \; \forall i \\ (\bar{b}_1, a_1, \bar{b}_2, a_2, ..., , \bar{b}_m, a_m) \in (H_{\theta})_{(h_1,...,h_{n-1})} \\ \text{where } \bar{b}_i \text{ are possibly empty vectors of actions not in } \alpha^{\theta}_{\bar{\theta}}. \end{array} \right. \tag{48}$$

Note that the vectors $\bar{b}_i$ are in H_{θ}. By induction we have that

$$(H_{\theta})_{\bar{\theta}} = H_{\theta\hat{}\bar{\theta}}. \tag{49}$$

We have that (37) for Γ^{θ} follows from (37) for Γ. Consider a sequence of vectors $\bar{b}_1, \bar{b}_2, ..., \bar{b}_n$ such that $\bar{b}_i$ contains no member of $\alpha_{\theta\hat{}\bar{\theta}}$ and contains only actions in H_{θ}. Let $\theta = (g_1, ..., g_m)$.

By induction, there is a unique sequence of vectors $\bar{c}_j$ $(\bar{c}_1, \bar{b}_1, \bar{c}_2, a_1, \bar{c}_3, \bar{b}_2, \bar{c}_4, a_2, .., \bar{b}_n, \bar{c}_{2n}, a_n) \in H_{\theta\hat{}(h_1,...,h_{n-1})}$. From the uniqueness property (38) for the game Γ, we have that the uniqueness of the vectors $\bar{d}_1 = \bar{c}_1\hat{}\bar{b}_1\hat{}\bar{c}_2, ..., \bar{d}_n = \bar{c}_{2n-1}\hat{}\bar{b}_n\hat{}\bar{c}_{2n}$ implies that there is a unique sequence of vectors from $(H_\theta)_{(h_1,...,h_{n-1})}$ with no elements from $\alpha^{\theta}_{\bar{\theta}}$ for every member of $(H_\theta)_{\bar{\theta}}$ and (38) holds for Γ^θ. Condition (39) follows using the same argument and (49). This argument with $I^{\theta}_{\bar{\theta}} = I_{\theta\hat{}\bar{\theta}}$, (49) and $P^{-1}(P(I_\theta)) \subset P^{-1}(N)$ also yields (40) for Γ^θ. Finally, if $(a_1, .., a_n)$ is terminal in $(H_\theta)_{\bar{\theta}}$ then it is terminal in $H_{\theta\hat{}(h_1,...,h_{n-1})}$ from (41) for Γ and (49). If, by way of contradiction, the unique extension $(\bar{b}_1, a_1, \bar{b}_2, a_2, .., \bar{b}_n, a_n)$ of $(a_1, .., a_n)$ in $(H_\theta)_{(h_1,...,h_{n-1})}$ was not terminal in H_θ we could have added an action c from H_θ and find $(\bar{b}_1, a_1, \bar{b}_2, a_2, .., \bar{b}_n, a_n, c) \in H_\theta$. Extending $(\bar{b}_1, a_1, \bar{b}_2, a_2, .., \bar{b}_n, a_n, c)$ to H would now imply that there is an extension of $(a_1, .., a_n)$ from $H_{\theta\hat{}\bar{\theta}}$ to H which is not terminal in H. This contradicts an inductive application of (41) for Γ. We conclude that (41) must hold for Γ^θ and the proof is complete. ■

Proof of Lemma 7. Let $\theta = (h_1, .., h_n) \in \bar{\Theta}$ hence $h_n \in I_{(h_1,..,h_{n-1})}$. By (33) we have that $P(I_\theta) \subset \alpha_\theta$. From the definition of ρ in (1) we have $\rho(C_{h_1}...C_{h_n}(i \wedge \neg j)) = \{(\theta, i), (\theta, j)\}$ and from (2), $i, j \in \alpha_\theta$ and $h_n \in I_{(h_1,..,h_{n-1})}$ we deduce that the statements in (20) are all in $\mathcal{L}^{\mathcal{U}}$. For the statements in (21) note that the definition of $A^\theta(h)$ assures that when $a \in A^\theta(h)$ then $a \in \alpha_\theta$ and since $h \in I_\theta$ from (32) applied to $\theta\hat{}h$ we have that $a \in \alpha_{\theta\hat{}h}$. We conclude that both $C_{h_1}...C_{h_n}(a)$ and $C_{h_1}...C_{h_n}(C_h a)$ are in $\mathcal{L}^{\mathcal{U}}$ which assures the statements in (21) are included in $\mathcal{L}^{\mathcal{U}}$. In the exact same manner we also have that the sets of statements in (23), (22), (24) and (25) are included in $\mathcal{L}^{\mathcal{U}}$. For the statements in (26) and (27) we simply note that the definition of ρ for the operators U^r_h and P^α_h is constructed in the same manner as for the operator C_h. Since the actions in statements of the form $s(h'), \pi|_a$ are in the corresponding α_θ these sets of statements are also in $\mathcal{L}^{\mathcal{U}}$. Note that for $\theta = \emptyset$ a smaller collection of sets of statements is considered and the same argument assures these are in $\mathcal{L}^{\mathcal{U}}$. ■

Proof of Proposition 9. Consider a game Γ with unawareness $\mathcal{U}$ and $\theta \in \bar{\Theta}$. We fix θ and assume throughout that $\theta \neq \emptyset$ since otherwise the statement is trivial.

Let $\theta' \in \bar{\Theta}$. We show that if $f \in \gamma^{\theta'}$ as in (28) and $f \in \mathcal{L}^{\mathcal{U}^\theta}$ then f is implied by the epistemic form of Γ^θ with unawareness $\mathcal{U}^\theta$. This will prove that

$$\gamma^{\theta'} \cap \mathcal{L}^{\mathcal{U}^\theta} \subset \left(\Gamma^\theta\right)^{\mathcal{U}^\theta}. \tag{50}$$

From (50) the conclusion will follow since any member of the closure of (28) which belongs to $\mathcal{L}^{\mathcal{U}^\theta}$ can be deduced in $\mathcal{L}^{\mathcal{U}^\theta}$. This is a result of the nature of the axiom scheme as discussed at the end of Section 2, since if a derivation implies a statement in $\mathcal{L}^{\mathcal{U}^\theta}$ then by the axiom scheme and derivation rules the implication is also derived from a statement in $\mathcal{L}^{\mathcal{U}^\theta}$.

From the definition of $\gamma^{\theta'}$ the statement f appears in one of the collections of statements $(20)-(27)$ where $\theta' = (h'_1, ..., h'_m)$. Consider first a statement f from (20). Assume $f = C_{h'_1}...C_{h'_m}(i \wedge \neg j)$ with $i = P(h'_m), j \in P(I_{\theta'}) \setminus \{i\}$. We need to show that if $f \in \mathcal{L}^{\mathcal{U}^\theta}$ then $f \in \left(\Gamma^\theta\right)^{\mathcal{U}^\theta}$. Assume $f \in \mathcal{L}^{\mathcal{U}^\theta}$ and let $\left(\gamma^\theta\right)^{\theta'}$ denote the collection of statements $(20)-(27)$ as defined for θ' for the game Γ^θ with unawareness $\mathcal{U}^\theta$. We will show that $f \in \left(\gamma^\theta\right)^{\theta'}$ and more precisely that f is stated in (20) for $\left(\gamma^\theta\right)^{\theta'}$. Since $f \in \mathcal{L}^{\mathcal{U}^\theta}$ we have that $i, j \in \alpha^\theta_{\theta'}$ and since $h'_m \in I_\theta$ we have $P^\theta(h'_m) = i$. In particular $I^\theta_{\theta'} \neq \emptyset$. From (35) we have

$$\alpha^\theta_{\theta'} \subset P^\theta(I^\theta_{\theta'}) \cup \bigcup_{h \in I^\theta_{\theta'}} A^\theta(h) \cup \bigcup_{h' \in (H_\theta \setminus Z) \setminus \left(P^\theta\right)^{-1}(N)} A^\theta(h'), \tag{51}$$

and since $A^\theta(h) = A(h) \cap \alpha_\theta$ and j is a name we have that $j \notin A^\theta(h)$. We conclude that $j \in \alpha^\theta_{\theta'}$ implies that $j \in P^\theta(I^\theta_{\theta'})$. We have shown that $P^\theta(h'_m) = i$ and $j \in P^\theta(I^\theta_{\theta'}) \setminus \{i\}$ which together with $i, j \in \alpha^\theta_{\theta'}$ and $I^\theta_{\theta'} \neq \emptyset$ imply that f is one of the statements in (20) for $\left(\gamma^\theta\right)^{\theta'}$.

Next assume that f is of the form that appears in (21). Let $f = C_{h'_1}...C_{h'_m}(C_h a \iff a)$ with $h \in I_{\theta'}$ and $a \in A^{\theta'}(h)$. From our assumption that $f \in \mathcal{L}^{\mathcal{U}^\theta}$ we have that $a \in \alpha^\theta_{\theta'}, \alpha^\theta_{\theta'\hat{}h}$ and $h \in I^\theta_{\theta'}$ hence also $I^\theta_{\theta'} \neq \emptyset$. We now show that f is stated in (21) for $\left(\gamma^\theta\right)^{\theta'}$. Since $A^{\theta\hat{}\theta'}(h) = A(h) \cap \alpha_{\theta\hat{}\theta'}$ we have that $a \in A^{\theta\hat{}\theta'}(h)$. From the first condition of consistency of an unawareness construction we have $\alpha_{\theta\hat{}\theta'} \subset \alpha_\theta$ which implies

$$A^{\theta\hat{}\theta'}(h) = A(h) \cap \alpha_{\theta\hat{}\theta'} = A(h) \cap \alpha_\theta \cap \alpha_{\theta\hat{}\theta'} = A^\theta(h) \cap \alpha^\theta_{\theta'} = \left(A^\theta\right)^{\theta'}(h). \tag{52}$$

From $h \in I^\theta_{\theta'}$ and (52) we have that $f \in \left(\gamma^\theta\right)^{\theta'}$ as required.

For f as in (22) we have $a, a' \in \alpha^\theta_{\theta'}$ since $f \in \mathcal{L}^{\mathcal{U}^\theta}$. As in (51) above, there is an $h \in H_{\theta'}$ such that $a, a' \in A^\theta(h)$ since a, a' are not names and A^θ is a restriction of A. From (52) we have $a, a' \in \left(A^\theta\right)^{\theta'}(h)$. Since no new terminal nodes are added to $(H_\theta)_{\theta'}$, which coincides with $H_{\theta\hat{}\theta'}$ according to (49), there is an $h' \in (H_\theta)_{\theta'} \setminus Z$ with $a, a' \in \left(A^\theta\right)^{\theta'}(h')$ and the conditions for $f \in \left(\gamma^\theta\right)^{\theta'}$ are satisfied.

Let

$$f = C_{h'_1}...C_{h'_m}\left(\bigvee_{a \in A^{\theta'}(h)} a\right) \tag{53}$$

with $h \in H_{\theta'} \setminus Z$. From $f \in \mathcal{L}^{\mathcal{U}^\theta}$ we have that $A^{\theta'}(h) \subset \left(A^\theta\right)^{\theta'}(h)$ since $A^{\theta'}(h) \subset \alpha^\theta_{\theta'}$ and from (52). Here we abuse the notation h for the corresponding member of both $H_{\theta'}$ and $H_{\theta\hat{}\theta'}$ whose existence was established above. Since $A^{\theta'}(h) \supset \left(A^\theta\right)^{\theta'}(h)$ from the definition of the restriction of the action set, we conclude that $A^{\theta'}(h) = \left(A^{\theta}\right)^{\theta'}(h)$ which implies that $\bigvee\limits_{a\in\left(A^{\theta}\right)^{\theta'}(h)} a = \bigvee\limits_{a\in A^{\theta'}(h)} a$ and f appears in (23) for $\left(\gamma^{\theta}\right)^{\theta'}$.

Let $f = C_{h'_1}...C_{h'_m}(C_h g \iff C_{h'} g)$ for some $h \in I_{\theta'}$ and $h' \in G^{\theta'}(h)$ and $g \in \mathcal{L}^{\mathcal{U}^{\theta\hat{}h}}$. From $f \in \mathcal{L}^{\mathcal{U}^{\theta}}$ we deduce that $g \in \mathcal{L}^{\mathcal{U}^{\theta\hat{}\theta'\hat{}h}}, \mathcal{L}^{\mathcal{U}^{\theta\hat{}\theta'\hat{}h'}}$ and $h, h' \in I^{\theta}_{\theta'}$. Since $\left(G^{\theta}\right)^{\theta'}(h)$ is a restriction of $G(h)$ to $I^{\theta}_{\theta'}$ we have that $h' \in \left(G^{\theta}\right)^{\theta'}(h)$. We have that $\mathcal{U}^{\theta\hat{}\theta'\hat{}h} = \left(\mathcal{U}^{\theta}\right)^{\theta'\hat{}h}$ from the definition of the unawareness constructions and so

$$g \in \mathcal{L}^{\mathcal{U}^{\theta\hat{}\theta'\hat{}h}} = \mathcal{L}^{\left(\mathcal{U}^{\theta}\right)^{\theta'\hat{}h}} \tag{54}$$

and the conditions for f appearing in (24) for $\left(\gamma^{\theta}\right)^{\theta'}$ are satisfied.

For the cases where f is of the form (25), (26) or (27) the proof is similar to the the cases studies above. ■

We note that the other direction need not hold and we might have a strict inclusion:

$$\Gamma^{\mathcal{U}} \cap \mathcal{L}^{\mathcal{U}^{\theta}} \subsetneqq \left(\Gamma^{\theta}\right)^{\mathcal{U}^{\theta}}. \tag{55}$$

For example, at the state of awareness θ the game might describe the set of actions that h has as $A^{\theta}(h)$. From (23) we have that the statement $f = \bigvee\limits_{a\in A^{\theta}(h)} a$ holds in $\left(\Gamma^{\theta}\right)^{\mathcal{U}^{\theta}}$. However, in the game Γ the identity h might have more available actions in $A(h)$, hence the statement $f \in \mathcal{L}^{\mathcal{U}^{\theta}}$ will not be implied by $\Gamma^{\mathcal{U}}$.

Proof of Proposition 10. Consider the game with unawareness for Γ as depicted in Figure 1 and $\mathcal{U}$ as defined in $(3)-(4)$. Let $\theta = (i_1, ..., i_n) \in \bar{\Theta}$. We have either $I_{\theta} = I$ or $I_{\theta} = I_{(UA_1)}$. In the first case $i_k \notin \{UA_j\}_j$ for every $k = 1, .., n$. From the definition of a game with unawareness we have that γ^{θ} as defined in $(20)-(27)$ describes that i_1 is confident that i_2 is confident that ... that i_n is confident that the game is as described in Γ^{θ}. Since $I_{\theta} = I$ we have that $\Gamma^{\theta} = \Gamma$ and hence γ^{θ} is implied by the set of statements described in (c). For $I_{\theta} = I_{(UA_1)}$ we have that $\theta = (i_1, ..., i_{k-1}, i_k, ..., i_n)$ where for some $k \leq n$ we have $i_{k+1}, ..., i_n \in I_{(UA_1)}$, $i_k = UA_j$ and $i_1, ..., i_{k-1} \notin \{UA_j\}_j$ when $k > 1$. In particular, the statements in γ^{θ} state that i_1 is confident that i_2 is confident that ... that i_{k-1} is confident that i_k is confident of some mutual confidence of members in $I_{(UA_1)}$ that the game follows the description in Γ^{θ} which is as depicted in Figure 2 since $I_{\theta} = I_{(UA_1)}$. This collection of statements is implied by (b) when $k = 1$ and therefore by (d) when $k > 1$. For the converse direction the proof follows similarly by noting that for each of the statements of mutual confidence of the game being played as described in $(a)-(d)$, the corresponding statements can be found in γ^{θ} where θ is describes the identities ordered by the mutual confidence level considered. ■

Before we prove Theorem 12 we note that in Feinberg (2004b) we provided an epistemic characterization for equilibria with sequential rationality and convex structural consistency. However, for FRPD as well as the game in Figure 1 (viewed as a standard extensive form game) we find that this solution coincides with sequential equilibria since in both cases any assessment off the equilibrium path can be obtained as a limit of perturbations.

For the proof of the theorem we relate the sequential equilibria of the FRPD with a grain of irrationality to the sequential equilibria of the extensive form game depicted in Figure 1. Throughout the remainder of the paper we let Γ denote the game in Figure 1. Recall that Kreps, Milgrom, Roberts and Wilson (1982) considered the FRPD with a grain of irrationality with an irrational type that plays tit-for-tat however their result is known to hold when one considers an irrational type that plays the grim trigger strategy - cooperate if and only if there was never a defection. We assume that the irrational type *must* follow this strategy and not that she prefers this strategy, in the sense that this is the only course of action available to her in the corresponding extensive form game. This extensive form game has Nature move and select whether Alice is rational or not. In Figure 3 we depict the extensive form of the FRPD with a grain of irrationality with an irrational type that must play the grim trigger strategy. We denote this game by ϑ. We note that the game ϑ differs from Γ by the continuations after an irrational Alice denoted IA_i observes a defection in ϑ versus an identity that becomes aware because of a defection in Γ. Also some information sets that involve such identities are rearranged. For example, consider the difference between the identity IA_3 depicted in Figure 3 and the identity A_2 in Figure 1. In Γ a defection leads to a continuation following the FRPD while in ϑ the continuation leads to Alice defecting forever and Bob being possibly uncertain if Alice is rational or not. The results of Kreps, Milgrom, Roberts and Wilson (1982) extend to the game ϑ as depicted (it is actually a simpler case since the irrational type has no choice but to follow the grim trigger behavior. In fact, a grim trigger strategy was used by Fudenberg and Maskin (1986) to prove a folk theorem for two person games with a grain of irrationality (see also Theorem 9.2 in Fudenberg and Tirole ,1991). The similarity between the games ϑ and Γ will allow us to prove that the epistemic characterization of sequential rationality in $\Gamma^{\mathcal{U}}$ leads to cooperation. Although the extensive form games ϑ and Γ are similar we remind the reader that we are interested in analyzing behavior in the game with unawareness $\Gamma^{\mathcal{U}}$ and not the game with full awareness Γ.

The following Lemma characterizes behavior in a sequential equilibrium after a defection has occurred in the game ϑ.

Lemma 13 *Every sequential equilibrium of the game ϑ satisfies:*

Once a defection occurs either by Alice (either rational or not) or by Bob, either on the equilibrium path or off the equilibrium path, the sequential equilibrium dictates that both players will defect from that point onwards.

Proof. By the definition of the game, the irrational Alice defects once defection occurs. Assume that defection occurred and that a rational Alice cooperates after defection occurs. By the extensive form game ϑ depicted in Figure 3 we have that if Alice cooperates after defection then after that point Bob knows that Alice is rational. Hence, after Alice cooperates following some past defection the game continues as a subgame which is identical to a FRPD. Here we use the fact that the irrational Alice must defect after defection hence a cooperation by Alice after defection can only occur in a branch of the game tree that follows Nature choosing a rational Alice. Since after she cooperates following past defection we are at a subgame of the game ϑ which is identical to a FRPD, we conclude by subgame perfection that the unique sequential equilibrium behavior from that point onwards is defection for both Alice and Bob.

We found that after defection occurs if Alice cooperates she will get the stage payoff for cooperating and the continuation payoff from both players defecting from that point onwards in a sequential equilibrium. In particular, no matter what Bob does at the stage she cooperated (after defection occurred), she is strictly better off defecting. Hence, the rational Alice will also always defect in a sequential equilibrium once defection has occurred. Since once defection occurred Alice will be defecting forever whether she is rational or not we have that Bob's best response is to defect forever after defection occurred as claimed. ■

We now relate the behavior in sequential equilibria in the games ϑ and Γ treated as a standard extensive form game.

Lemma 14 *Consider the FRPD game with a grain of irrationality ϑ depicted in Figure* 3*, where there is an irrational type for Alice that is chosen with a small probability $\varepsilon > 0$ and this irrational type plays the grim trigger strategy (she cooperates as long as Bob does and defects forever once Bob defects). The sequential equilibria behavior strategies for decision points following nature choosing a rational Alice coincide with the sequential equilibria behavior for decision points following nature choosing an aware Alice for the extensive form game Γ depicted in Figure* 1*. Furthermore, in both games in every sequential equilibrium, once a defection occurs both players will defect forever.*

Proof. Note that in Figures 1 and 3 we have already denoted a one to one mapping of the identities (and their actions) following Nature choosing "*rational*" in ϑ onto the the identities following nature choosing "*aware*" in Γ. The correspondence is given by the naming

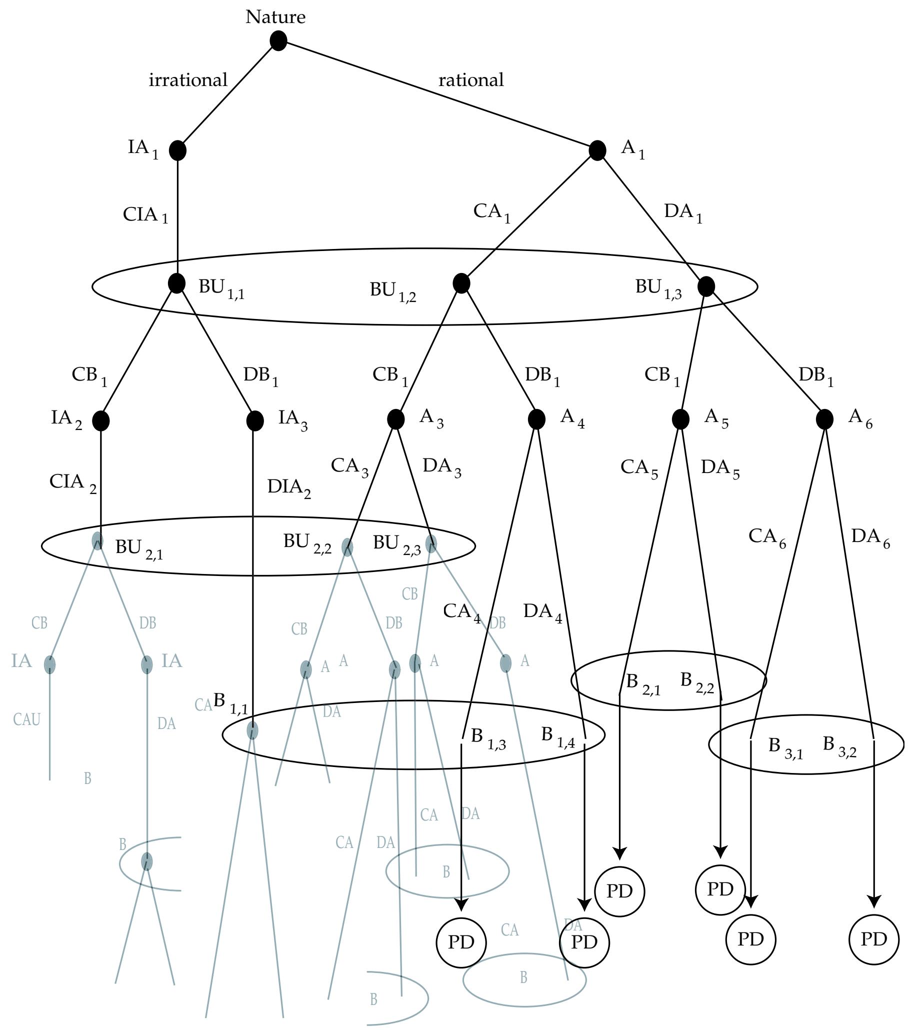

FIGURE 3: The FRPD with a grain of irrationality in the form of a grim trigger strategy.

of identities and indices. In all sequential equilibria of ϑ the irrational Alice cooperates as long as Bob does and defects forever if he doesn't. From Lemma 13 we also have that after Nature chooses "*rational*" once Alice or Bob defect then defection occurs forever in ϑ. Hence, even at information sets where Bob is uncertain whether Alice is irrational or not, when he observes defection he will choose to defect. Given a sequential equilibrium for ϑ consider the following strategy profile for Γ:

> In information sets where either all members follow nature choosing "*aware*" or where no defection has occurred so far, play the same strategy as in the given sequential equilibrium for ϑ. In all other information sets defect. The assessment associated with this strategy profile in Γ is identical to the assessment for the equilibrium for ϑ at information sets where no defection has occurred or where all members follow Nature choosing "*aware*". At other information sets choose

any arbitrary consistent assessment.

We need to show that the strategy profile and assessment defined for Γ constitute a sequential equilibrium. For information sets where no defection has previously occurred we have that the players' expected payoff is based on the assessment and the continuation payoff from cooperation versus defection. If the equilibrium dictates defection then the expected continuation payoff is identical to the expected continuation payoff for the corresponding information set in ϑ since the assessments coincide and defection leads in both cases to defection forever from the next stage. Now, consider an information set which has no continuation in which the equilibrium dictates positive probability for cooperation and such that no defection has previously occurred, by induction the expected payoff from defection and cooperation follows that of the game ϑ and hence the same distribution over the actions as in the equilibrium for ϑ is a sequentially rational behavior in Γ. By induction we have that the strategy mapped from the sequential equilibrium in ϑ is a sequential equilibrium in Γ.

The proof that a sequential equilibrium in Γ induces a sequential equilibrium in ϑ with the same behavior for the aware identities is exactly the same. The strategy and assessment are mapped for corresponding information sets where no defection has previously occurred and defection forever with an arbitrary consistent assessment is dictated for all other information sets. The mapping between the sequential equilibria in these two games induces a one to one mapping of the behavior of the aware identities onto the behavior of the rational identities in the respective games. Together with Lemma 13 which states that the behavior for the unaware and irrational types is uniquely determined in a sequential equilibrium, we have defined a one to one onto mapping between the sets of sequential equilibria of these games where the behavior of rational and aware identities coincide. ■

For the proof of Theorem 12 it now is sufficient to show that the epistemic conditions characterizing sequential equilibria in Γ as described in Feinberg (2004b) retain their implications on the aware identities when constrained to the game with unawareness $\Gamma^{\mathcal{U}}$.

Proof of Theorem 12. We will show that if (μ, σ) is an assessment such that the set of statements $a) - d)$ in Definition 11 of an extended sequential equilibrium are logically consistent in $\mathcal{L}^{\mathcal{U}}$ then the behavior of the aware identities according to σ coincides with their behavior in a sequential equilibrium of the game Γ. From Lemma 14 this will prove the theorem.

Assume by way of contradiction that for some aware identity (μ, σ) does not imply the behavior of a sequential equilibrium in Γ even though the statements $a) - d)$ are consistent. In particular, we have an identity $h \notin \{UA_j\}_j$ such that σ assigns positive probability to an action of h which is not a best response given μ and σ. Consider another aware identity $\bar{h}$ that does not follow h and is on a σ possible play path. If $h \neq A_1$, where A_1 is the aware identity of Alice immediately following Nature's move in Figure 1, we let $\bar{h} = A_1$. If $h = A_1$ we let $\bar{h}$ be the first identity of Alice which becomes aware following a defection of an uncertain identity $BU_{j,1}$ of Bob, where this defection occurs with positive probability according to σ. We need to show that such an identity exists, i.e. that with positive probability according to σ an uncertain Bob will defect prior to the last round. Suppose this was not the case, then Bob cooperates for sure as long as Alice does until the last round according to σ, but this implies that an aware Alice is strictly better off defecting in the one before last round if no previous defection occurred. From A_1's viewpoint the statements in *d*) imply that she finds this future identity of hers to be rational and since they are fully aware of the game from *b*) we find that A_1's conjecture about her identity in this prior to last round coincides with σ and dictates defection. Since, by our assumption, the prior to last information set of the uncertain Bob is reached with positive probability according to σ, we find that Bob's identities in this information set expect Alice's aware identity to defect in this round according to σ. We assumed these identities are cooperating in this round which implies that μ in this information set assigns a very high probability to Alice being unaware. Backtracking, we must conclude that μ (which agrees with σ since this information set is σ possible) was previously updated by a positive probability of defection by an aware Alice in a previous round. Backtracking until the round where such a defection occurs with positive probability according to σ we find an information set where an aware Alice defects with positive probability even though cooperation will lead Bob to cooperate until the one before last round by our assumption that only at that point Bob will defect if no defection occurred earlier. This defection by Alice is not a best response, but it is dictated by σ and is part of the conjecture of an identity of Bob which is aware of the game and hence a contradiction to *d*). We conclude that we can always find an aware σ possible identity $\bar{h}$ that does not follow an identity h, such that $\bar{h}$'s conjecture about h coincides with σ. Since in both cases $\bar{h}$ is σ possible we have from *d*) that $\bar{h}$ finds h to be rational and from h and $\bar{h}$'s full awareness, h's rationality implies that h plays a best response to (μ, σ) according to *b*) and *c*). Since $\bar{h}$ conjecture coincides with σ and $\bar{h}$ is confident that h's conjecture follows σ we find that σ cannot assign positive probability to an action that is not a best response to (μ, σ) and the proof is complete. ■