Awareness as an Equilibrium Notion: Normal-Form Games.

Jernej Čopič and Andrea Galeotti *

Abstract

We study normal-form games where parts of the games may not be common knowledge. Agents may be aware only of some facts describing the game. An awareness architecture is given by agents' awareness, and an infinite regress of conjectures about other agents and their conjectures. The problem is specified by the true underlying normal-form game, and by the set of possible awareness architectures. Awareness equilibrium is given by a feasible awareness architecture for each agent, strategies that are played and these strategies have to be consistent with the awareness architectures and agents' rationality. We first study games with complete information, where each player may be aware of a subset of the set of possible actions. We then study games with incomplete information, where each player may be aware of a subset of the set of types and probability over types. Our results illustrate how a departure from the assumption of common knowledge alters equilibium predictions.

*jernej@hss.caltech.edu; agaleo@essex.ac.uk. A conversation with Leeat Yariv helped us to get started on this project. We have also benefitted from discussions with Ken Binmore, Colin Camerer, Matt Jackson, Andrea Mattozzi, Debrah Meloso, Bill Zame, and the participants at the BWED conference 2005, especially Beth Allen, Andres Hervas and Tural Huseynov, and the Caltech SISL 2005 workshop.

1 Introduction

We define and analyze equilibrium situations where parts of the game may not be common knowledge. The general question we address is how departing from common knowledge alters the equilibrium predictions. We model the lack of common knowledge by using the notion of awareness. Each player may be aware of some, but perhaps not all, aspects of the game. A player makes conjectures about what others are aware of and what others will do, makes conjectures about others' conjectures, and so on *ad infinitum*.[1] This constitutes an awareness architecture. Heuristically, the awareness architecture of a player summarizes the theory that a player formulates about the world. Based on this theory, the player defines his plan of actions.

If the modeler has no information about agents' awareness, then he must consider all possible awareness architectures for the players. In other situations, the modeler may know something about players' awareness, e.g., that a player is aware of a certain fact. This is represented by the set of possible awareness architectures, which might be restricted.

Awareness architectures, and the play of the game, constitute an equilibrium situation if the following *three principles* are satisfied. The first principle is that conjectures must be consistent with awareness, that is, players may not make conjectures about facts that they are not aware of. The second principle is that conjectures must be consistent with the outcome that obtains. When such outcome is probabilistic, players' conjectures must be correct regarding the actual probabilities, otherwise they would realize that something was wrong with their inferences about the world. The third principle requires that the play of the game and conjectures must be consistent with optimizing behavior. In equilibrium, agents' awareness architectures and the play of the game must be self confirming. Thus, in equilibrium, the theory that each player formulates about the world must not be contradicted.

[1] Feinberg [2004] defines such construction as an awareness construction.

Although our approach to awareness is quite different, it is related to the epistemic models of unawareness, in particular those concerned with multi-person settings. After the seminal contribution of Dekkel et al. [1998], showing that standard state-space representation precludes non-trivial forms of unawareness, this literature has focused on providing general state-space models which are able to overcome this negative result. Recent papers on this are Heifetz et al [2005], Li [2006], Modica and Rustichini [1999]. These models propose state-space representations of unawareness; in specific situations, agents' unawareness structure is a primitive that models agents' states of minds. Feinberg [2004,2005] provides an approach that is more similar to ours - in particular, he proposes modelling agents' awareness as an awareness architecture. In his approach, agents' awareness might change after having observed an outcome, so that Feinberg's approach can be thought of as a dynamic one.

There are two key differences between these approaches and ours. First, in our approach, the agents' awareness is part of equilibrium and the equilibrium has to satisfy the three principles above. Our second innovation is that through the specification of the set of possible awareness architectures, we model the modeler's knowledge about the agents' awareness and conjectures.

This is conceptually the core of our approach and is consistent with other notions of equilibrium. Awareness equilibrium is close to the equilibrium models where off the equilibrium deviations are only conjectured but never actually observed (see Rubinstein and Wolinsky [1994], Fudenberg and Levine [1993], and Battigalli and Guatoli [1998]). It is worth noting that in normal-form games, these types of equilibria are equivalent to Nash equilibrium, which is not the case here. On the one hand, awareness equilibrium weakens the equilibrium notion further by not requiring that the model itself be common knowledge. On the other hand, by specifying the set of possible awareness architectures, the equilibria may have to satisfy additional restrictions. As all equilibrium models, awareness equilibrium can be thought of as characterizing steady states of dynamic processes - in this case, processes where agents adjust their actions and awareness architectures.

The main results of our paper can be summarized as follows. In Section 2 we study normal-form games with complete information. Agents may have limited awareness of strategies available to them and to others. All other aspects of the game are common knowledge. The outcome is a profile of players' mixed strategies. We provide conditions on the set of possible awareness architectures under which awareness equilibrium exists. We study a natural class of possible awareness architectures, satisfying existence conditions, where agents may have a cognitive bound on the number of strategies that they are aware of. This cognitive bound is a parameter of the model. We show that, in general, the cognitive bound of the players has to be sufficiently low, in order to obtain awareness equilibrium outcomes which are different from Nash equilibrium outcomes. When the cognitive bound of players increases the set of awareness equilibria converges to the set of Nash equilibria.

In Section 3 we define awareness equilibrium in the context of normal-form games with incomplete information. We depart from the standard Harsanyi's model by allowing agents not being fully aware of the set of possible types and the true distribution over types. In these environments, an awareness architecture consists of conjectures about the type space and the distribution over types. An awareness architecture is thus related to the Mertens and Zamir [1985] construction of infinite hierarchies of beliefs, except that an awareness architecture also concerns conjectures about the type space itself.[2] An outcome here is a randomization over actions, for each player, which results from players' type-contingent strategies and the distribution of players' types. Players can't verify the type-contingent strategy of the others, but just the joint distribution over own payoffs and opponents' actions. While players may be unaware of some types, and have erroneous probabilistic conjectures, in an awareness equilibrium, players' conjectures about the joint distribution over own

[2] Mertens and Zamir [1985] consistency axioms H1 and H2 are satisfied for awareness architectures where the type space is common knowledge.

payoffs and opponents' actions must be consistent with the truth.

We describe the relation between type-awareness equilibria and Bayesian equilibria. We first show that in private-values environments and strict-common values environments all awareness equilibria are *strongly outcome equivalent* to Bayesian equilibria of the true underlying game. That is, the joint distribution of actions and payoffs generated by the play of an awareness equilibrium equals to the one generated by a Bayesian equilibrium in the Bayesian game where the type space and the true prior are common knowledge. Thus, strong outcome equivalence describes a situation where an outside observer knows the payoff structure, the equilibrium payoffs that were attained by players and the true distribution over the states of the world and he can justify the observed behavior as a Bayesian equilibrium.

Strict common values and private values provide a large class of economically interesting games where the departure from common knowledge over the prior and type space does not alter the equilibrium predictions. But, this is not true in general. Our second result provides simple conditions on the payoff structure of the game under which strong outcome equivalence doesn't obtain. We illustrate this result in Example 3. In that example there exists an awareness equilibrium outcome which is not sustainable in any Bayesian equilibrium under the true prior, but is sustainable in a Bayesian equilibrium under some prior. Therefore, an outside observer who does not know the true prior could still justify observed behavior as a Bayesian equilibrium behavior selecting the appropriate prior.[3]

One might be tempted to believe that this holds generally. Example 4 shows that there exists games with awareness equilibria which are not supportable as Bayesian equilibria under any prior. Given players' rationality, in such a case, an outside observer verifies the hypothesis of unawareness.

[3]Example 3 is taken from Jackson and Kalai [1996]. Jackson and Kalai [2006] study learning of agents in recurring games who are uncertain over prior. Both the setup and the results there are very different from ours.

2 Complete information and action-awareness equilibrium

In this section we study normal-form games of complete information where action sets might not be common knowledge. The set of agents is common knowledge. Each agent may be aware only of some actions and the corresponding outcomes. The mapping from outcomes into payoffs is common knowledge, so that if an agent is aware of a profile of actions, he is aware of the corresponding payoffs to all agents. Awareness equilibrium builds on Nash Equilibrium in the sense that agents play and conjecture best responses (given their awareness). We present this section in terms of 2-player finite games to make the section easier to read.

Let $\mathbf{N} = \{1, 2\}$ be the set of agents, let $\mathbf{A} = \mathbf{A}_1 \times \mathbf{A}_2$ be the set of action profiles, where $\mathbf{A}_n$ is finite for each $n \in \{1, 2\}$, $\mathbf{A}_1 = \{\underline{1}, ..., \underline{K}\}$ and $\mathbf{A}_2 = \{\bar{1}, ..., \bar{K}\}$, $a = (\underline{a}, \bar{a})$ is a typical element of $\mathbf{A}$, and $\sigma = (\underline{\sigma}, \bar{\sigma})$ is a mixed strategy profile. The set of pure-strategy outcomes corresponds to $\mathbf{A}$, and denote by $\Delta(\mathbf{A})$ the set of mixed-strategy outcomes, i.e. corresponding to lotteries over pure strategies. Payoffs over pure-strategy outcomes are represented by a mapping $u : \mathbf{A} \to I\!R^2$, $u(a) = (u_1(a), u_2(a)), \forall a \in \mathbf{A}$. Agents have Von Neumann-Morgenstern expected utilities, and payoffs associated to a mixed strategy profile σ are $U(\sigma) = (U_1(\sigma), U_2(\sigma)) = E_\sigma[u(a)]$.

Players' awareness restricts the set of actions of each player, and players make conjectures about others' awareness and others' conjectures and so on, which we call an awareness architecture. Denote by $A^{(n)} = A_1^{(n)} \times A_2^{(n)} \subset \mathcal{A}$ the action-awareness of player n, so that $A_1^{(n)}$ are actions of player 1 that player n is aware of, $n \in \{1, 2\}$. A player is aware of $u(a) = (u_1(a), u_2(a))$ if and only if he is aware of a. A first-order conjecture of agent n about m's awareness $A^{(m)}$ is denoted by $A^{(n,m)} = A_1^{(n,m)} \times A_2^{(n,m)}$, and so on. Define the awareness architecture of agent n

by $c_n = (A^{(n)}, A^{(n,m)}, A^{(n,m,n)}, ...)$, $n, m \in \{1,2\}$. The set of all possible awareness architectures for player n is $C_n \subset \{0,1\}^A \times \{0,1\}^A \times ...$ and the space of possible awareness architectures is $\mathbf{C} = C_1 \times C_2$. A normal form game with action awareness is $\mathcal{U} = (\mathbf{N}, \mathbf{A}, \mathbf{u}, \mathbf{C})$.

For a finite sequence k of length x let (k,n) be a sequence of length $x+1$, such that $(k,n)_i = k_i$ for each $i \leq x$ and $(k,n)_{i+1} = n$.

Definition 1. The *Action-awareness Equilibrium* (AAE) is an outcome $\sigma \in \Delta(\mathbf{A})$ and awareness architectures $(c_1, c_2) \in C_1 \times C_2$ such that

AAE1 $A_l^{(k,m)} \subset A_l^k$, and if $k = (n, ..., m)$, then $A^{(k,m)} = A^k$, $\forall l \in \mathbf{N}$, for all $k \in \mathbf{N}^x, \forall x < \infty$, $n, m \in \mathbf{N}$.

AAE2 $a \in A^k, \forall k \in \mathbf{N}^x, \forall x < \infty, \forall a \in \mathbf{supp}(\sigma)$.

AAE3 $\underline{\sigma} = \arg\max_{\underline{\sigma}' \in \Delta(A_1^k)} U_1(\underline{\sigma}', \bar{\sigma})$, $\bar{\sigma} = \arg\max_{\bar{\sigma}' \in \Delta(A_2^k)} U_2(\underline{\sigma}, \bar{\sigma}'), \forall k \in \mathbf{N}^x, \forall x < \infty$.

An AAE is a situation where the agents' perception of the world is internally consistent (AAE1), consistent with the outcome (AAE2), and consistent with the aspects that are common knowledge (AAE3). The requirement AAE1 is that agents cannot reason about actions that they are not aware of. For example, if player 1 is not aware of action $\underline{a}$, then he cannot conjecture that player 2 is aware of that action. This is very different from knowledge, where a player may not know that an action is available, but is allowed to make conjectures about this action. AAE1 also requires that if 1 is aware of some actions, he cannot conjecture otherwise about himself. AAE1 has been proposed by Feinberg [2005] (Weak Awareness Axiom), who also describes agents' subjective reasoning as an awareness architecture.

AAE2 requires that in equilibrium the players are aware of the action profile that is realized, and correctly conjecture that others are aware of that action profile, and so on at all orders of conjectures. AAE3 requires that the action profile that obtains is consistent with agents' optimization, at every order of conjectures. Equivalently, agents must conjecture that at every order of awareness, each player is playing a best reply to the actions of the other players.

Note that if $\mathbf{N}, \mathbf{A}, \mathbf{u}$ were all common knowledge then this would be a standard game. Then, AAE would coincide with the standard definition of Nash equilibrium. Here, our only departure is to relax the assumption of common knowledge of the action space.

While $\mathcal{U}$ is a different and more complex object than the standard game $\Gamma = (\mathbf{N}, \mathbf{A}, \mathbf{u})$, there is a relationship between Nash equilibria of Γ and AAE of $\mathcal{U}$. Nash equilibria of Γ are AAE of $\mathcal{U}$ that are not sensitive to the details of the specification of $\mathbf{C}$. Observe that in general, by virtue of AAE1, for every $\mathbf{C}$, we can restrict attention to $\mathbf{C}_E \subset \mathbf{C}$, such that AAE1 holds for every element of $\mathbf{C}_E$. We from now on restrict attention to architecture spaces $\mathbf{C}$ such that $\mathbf{C}_E \neq \emptyset$.[4]

Proposition 1. Given $\Gamma = (\mathbf{N}, \mathbf{A}, \mathbf{u})$, the profile σ is a Nash equilibrium of Γ if and only if it is supportable in AAE for every $\mathcal{U} = (\mathbf{N}, \mathbf{A}, \mathbf{u}, \mathbf{C})$, such that there exist $(c_1, c_2) \in \mathbf{C}_E$ with $\mathbf{supp}(\sigma) \subset \cap_{x<\infty} \cap_{k \in N^x} A^k$.

Proof. We provide the proof for pure strategies, the proof for mixed strategies is analogous. Let $(\underline{a}, \bar{b})$ be a Nash equilibrium of Γ and suppose $\exists (c_1, c_2) \in \mathbf{C}_E$ such that $(\underline{a}, \bar{b}) \in \cap_{x<\infty} \cap_{k \in N^x} A^k$. Since $(\underline{a}, \bar{b})$ is a Nash equilibrium there are no profitable deviations to either of the two players even if their action sets are restricted, so that AAE3 is satisfied. AAE1 and AAE2 are satisfied by assumption. For the converse, $(\underline{a}, \bar{b})$ is supportable on the architecture space $\mathbf{C}$, where $A^{(1)} = A^{(2)} = A$, in which case players must be playing a Nash equilibrium by AAE3. $\square$

The above proposition states that a Nash equilibrium profile of Γ is the only

[4] It is very easy to provide examples of $\mathbf{C}$ such that $\mathbf{C}_E = \emptyset$. For instance, that is true if $A^{(n)} \cap A^{(n,m)} = \emptyset$.

one for which players can make *any* conjectures that are internally consistent (in the sense of AAE1), and consistent with the given profile (in the sense of AAE2), and such conjectures along with the action profile constitute an AAE. That is, for a strategy profile that is not a Nash equilibrium in Γ we can find an architecture space $\mathbf{C}$ such that even if this strategy profile satisfies AAE2 (is feasible) it does not constitute a part of an AAE.

We now turn to the question of existence of AAE. Existence of AAE may depend on the specification of $\mathbf{C}$. There are two very different situations at the opposite extremes of the possible specifications of $\mathbf{C}$. The first situation is one where $C_n = \{0,1\}^A \times \{0,1\}^A \times ....$ This corresponds to the case where an omniscient outside observer sees the game, but has no indication on the agents' awareness of the game. In this case, existence is not an issue, since for instance the outcomes associated with Nash equilibria of Γ will be supported in AAE of $\mathcal{U}$. However, in this case, an outside observer could justify every realized pure-strategy outcome in AAE - simply take $A^k = \{\underline{a}, \bar{b}\}, \forall k$. At the other extreme is the situation where $C_n = \{0,1\}^{\{\underline{a},\bar{b}\}} \times \{0,1\}^{\{\underline{a},\bar{b}\}} \times ..., n \in \mathbf{N}$, then the outcome corresponding to $\{\underline{a}, \bar{b}\}$ is the unique outcome supportable in AAE. This is a very restrictive case where agents' awareness is trivial. The interesting cases are somewhere in between, where some restriction on $\mathbf{C}$ is exogenously specified. For example, an experimenter *tells* each player something about A, in which case $C_n \subset \{0,1\}^A \times \{0,1\}^A \times ...$, where $A^{(n)}$ has to equal to what player n was told. The following example illustrates that in such a situation, an AAE may fail to exist.

Example **1.** Let Γ be described by the following normal-form representation.

$_1\backslash^2$	$\bar{1}$	$\bar{2}$
$\underline{1}$	$6,4$	$8,7$
$\underline{2}$	$5,9$	$\mathbf{10},\mathbf{10}$

Observe that Γ has a unique pure-strategy Nash equilibrium, $(\underline{2}, \bar{2})$. If $\mathbf{C}$ is such

that $A^{(1)} = \{\underline{1}, \underline{2}, \bar{1}\}$ and $A^{(2)} = \{\underline{1}, \bar{1}, \bar{2}\}$, then no AAE exists. The reason is that regardless of $A^{(2,1)}$, player 2 would always play $\bar{2}$, which would violate AAE2.
In contrast, if $\mathbf{C}$ is such that $A^{(1)} = \{\underline{1}, \bar{1}, \bar{2}\}$ and $A^{(2)} = \{\underline{1}, \underline{2}, \bar{1}\}$, then $\{\underline{1}, \bar{1}\}$ can be supported in an awareness equilibrium. An awareness architecture that supports it is $A^{(12)} = A^{(21)} = A^k = \{\underline{1}, \bar{1}\}, \forall k$, s.t. $k \in N^x, x \geq 3, k = (n, m, ...), n \neq m$. Note that not all awareness architectures will be equilibrium architectures, for instance if $A^{(12)} = A^{(1)}$, no such AAE will exist.
Finally, we remark that if $\mathbf{C}$ is such that $A^{(1)} = A^{(2)} = A$ then the only AAE outcome is the Nash equilibrium outcome of Γ.

It is natural to ask what sets of possible awareness architectures will give existence of AAE.

PROPOSITION **2.** Given $\mathcal{U} = (\mathbf{N}, \mathbf{A}, \mathbf{u}, \mathbf{C})$, an AAE exists, if and only if there exists $(c_1, c_2) \in \mathbf{C}_E$, and $\exists \sigma$, $\mathbf{supp}(\sigma) \subset \cap_{x=1,2,...} \cap_{k \in \mathbf{N}^x} A^{(k)}$, with $\underline{\sigma} = \arg\max_{\underline{\sigma}' \in \Delta(A^k)} U_1(\underline{\sigma}', \bar{\sigma})$, for $k \in \{(1), (21)\}$ and $\bar{\sigma} = \arg\max_{\bar{\sigma}' \in \Delta(A^k)} U_2(\underline{\sigma}, \bar{\sigma}')$, for $k \in \{(2), (12)\}$.

Proof. The only if part follows from the fact that if such (c_1, c_2) didn't exist, then for every outcome satisfying AAE2, there would be a player $n \in \mathbf{N}$, such that either n would deviate given $A^{(n)}$, or $m \neq n$ would deviate under n's conjecture $A^{(n,m)}$. In either of these cases, AAE3 is violated.
We now prove the if part. If n doesn't have a profitable deviation under $A^{(n)}$ and under $A^{(m,n)}$, then he does not have a profitable deviation under S, $\forall S \subset A^{(n)}$ nor under P, $\forall P \subset A^{(m,n)}$, so that the claim follows by AAE1. □

Proposition 2 shows that generically, a restriction on $\mathbf{C}$ for which AAE will exist, will be much stronger than just requiring that there exist an outcome in the intersection of all the players' conjectures. There must exist such outcome, which is also consistent with players' own optimization, and the first order conjecture that the other agent optimizes.

Proposition 2 also illustrates that admissible restrictions on **C** in general depend on the specification of Γ. As we noted earlier, one class of such restrictions on **C** is to impose the awareness of the agents.[5] In the next section we consider a model where **C** is restricted in a natural way, independently of Γ.

2.1 A model of cognitive bounds

In this section we study a model where the number of actions that a player can be aware of is a parameter, but not which actions these are. Such restriction on **C** has a natural interpretation that the agents have bounded cognitive abilities.[6] Moreover, the existence of AAE will not be in question, regardless of Γ.

We study how the number of outcomes sustained in a AAE changes with the cognitive bound. This illustrates the effect of unawareness of actions on the predictions of the standard model. It provides a very stylized metric of how far from the set of Nash equilibrium outcomes the outcomes sustainable in AAE might be.

We focus on pure strategies. Let ℓ be the number of actions of each agent that agent i is aware of, and assume that the number of actions that agents are aware of is common knowledge.

Definition **2.** Fix an $\ell \geq 1$. An ℓ-Action-awareness equilibrium, ℓ-AAE, is an AAE where $|A_l^{(n)}| = \ell$, and this is common knowledge.

By Proposition 1, a Nash equilibrium profile would be an ℓ-AAE whenever the corresponding Nash-equilibrium action profile a is in the awareness sets of both

[5] One could also consider the possibility that the experimenter also tells the players what he told the other player, and possibly lies about that, but this might not be enough to control the players' conjectures if they see a reason not to trust the experimenter. This consideration does not apply to their awareness, since if the experimenter tells them something he is sure that they are aware of it.

[6] For example, the subjects in an experiment might be confronted with a very large normal-form game, only a fraction of which fits on the computer screen.

players. This observation only holds for Nash equilibrium profiles of Γ. This suggests a refinement of ℓ-AAE, which we call ℓ^*-AAE. An ℓ^*-AAE is an ℓ-AAE which is an equilibrium even if some Nash-equilibrium profile is in the awareness sets of both players. We carefully define and study ℓ^*-AAE in Appendix A.

We now compare how the sets of ℓ-AAE change as we vary ℓ. Such comparison is useful for providing a measure of how strengthening the restriction on the architecture space strengthens the equilibrium notion. This comparative-statics approach also provides a method of estimating the cognitive bound ℓ of the agents. In the absence of other considerations, if a certain outcome is observed, then ℓ has to be low enough, in order to support that outcome as an ℓ-AAE. The following result simplifies our analysis.

LEMMA **1.** A profile of actions a is an ℓ-AAE if and only if it is an ℓ-AAE with $A^k = A^{k'}, \forall k \in N^x, \forall x < \infty$.

Proof. The if part is trivial and we omit it. The only if part is as follows. Let a be an ℓ-AAE outcome under *some* awareness architecture, different from those specified in the claim. This implies that there are at least $\ell - 1$ deviations by each player to which a is a best reply. But then a can be supportable also with some awareness architecture where all agents are aware of the same actions and make the correct conjectures about others. □

We will further restrict our analysis to generic games. The reason is that it is possible to construct non-generic and non-trivial $(K \times K)$ games where there is a unique Nash equilibrium in the game $(\mathbf{N}, \mathbf{A}, \mathbf{u})$, but for every $\ell < K$ every outcome can be supported in an ℓ-AAE (and even in ℓ^*-AAE). We illustrate this with Example 5 in the Appendix A.

We say that a $K \times K$ game Γ is *generic* if it satisfies the following *no-indifference* condition,

$$u_1(\underline{p}, \bar{q}) \neq u_1(\underline{p}', \bar{q}), \forall p \neq p', \forall q,$$

and similarly for player 2.

THEOREM **1.** Let Γ be a generic $K \times K$ game. Denote by $e_\ell(\Gamma)$ the number of distinct ℓ-AAE outcomes of Γ, for each $\ell \in \{1, ..., K\}$. Then $K^2 - 2(\ell - 1)K \leq e_\ell(\Gamma) \leq K^2 - (\ell - 1)K$, $\forall \ell \in \{1, ..., K\}$.

Proof. See Appendix A. □

The bounds in Theorem 1 are tight. Theorem 1 shows that as ℓ increases, in a generic game the set of ℓ-AAE shrinks. When ℓ converges to K, the set of ℓ-AAE generically converges to the set of Nash equilibria of Γ. However, it is a simple corollary that when ℓ is substantially smaller than K, the set of ℓ-AAE is strictly larger than the set of Nash equilibria.

COROLLARY **1.** Let $K \geq 3$, and let Γ be a generic $K \times K$ game, then the set of ℓ-AAE outcomes of Γ is a strict superset of the set of pure-strategy Nash Equilibrium outcomes of Γ, $\forall \ell \leq \frac{K}{2}$.

Proof. A generic $K \times K$ game can have at most K pure-strategy Nash Equilibrium outcomes, and the claim follows. □

Theorem 1 and Corollary 1 tell us that in general, to obtain outcomes that are different from Nash equilibria, players have to be unaware of a large number of actions. Theorem 1 says nothing about the bounds on the number of ℓ-AAE relative to the number of pure-strategy Nash Equilibria. When pure strategy Nash Equilibria exist, the lower bound on the number of ℓ-AAE may be in some cases improved, since every Nash Equilibrium is also an ℓ-AAE, for all ℓ.[7] Nevertheless, Theorem 1 shows

[7]In contrast, Theorem 2 in Appendix A shows that the lower bound on the number of ℓ^*-AAE is linked to the number of Nash equilibria of Γ.

that it is impossible to provide general conditions on game forms which would assure that the set of ℓ-AAE equals the set of Nash Equilibria, without tying ℓ to K.

3 Incomplete information and type-awareness equilibrium

In this section we define awareness equilibrium in the context of games with incomplete information. Our starting point is Harsanyi's model, and we depart from it by considering the possibility of agents not being fully aware of the set of possible types and the true distribution over types. If types were common knowledge and players were allowed to be unaware of some actions, we could in the usual way think of this as a normal-form game between different types, and the results would be the same as in the previous section. What is truly different in a setting with uncertainty is that it allows for unawareness of types and the true distribution over these. Players here can't verify the type-contingent strategy of the others, but just the joint distribution over own payoffs and opponents' actions. While players may be unaware of some types, and have erroneous probabilistic conjectures, in equilibrium, players' conjectures about the joint distribution over own payoffs and opponents' actions must be consistent with the truth. The model where agents are not fully aware of the set of types can be interpreted as a specific case of unawareness of players.

Let $\mathbf{N} = \{1, 2\}$ be the set of agents, let $\mathbf{A} = \mathbf{A}_1 \times \mathbf{A}_2$ be the set of action profiles, where $\mathbf{A}_n$ is finite for each $n \in \{1, 2\}$, $\mathbf{A}_1 = \{\underline{1}, ..., \underline{K}\}$ and $\mathbf{A}_2 = \{\bar{1}, ..., \bar{K}\}$, $a = (\underline{a}, \bar{a})$ is a typical element of $\mathbf{A}$. Let $\mathbf{T} = \mathbf{T}_1 \times \mathbf{T}_2$ be the set of players' types, where $\mathbf{T}_n$ is finite. We refer to t_n as an element of $\mathbf{T}_n$, and $t = (t_1, t_2)$ is an element of $\mathbf{T}$. There is a $P : \mathbf{T} \to [0, 1]$, which is the true joint probability distribution over players' types.

A strategy for player n is a mapping $s_n^{(n)} : \mathbf{T}_n \to \Delta(\mathbf{A}_n)$, where $\Delta(\mathbf{A}_n)$ is the set of probability distributions over $\mathbf{A}_n$. Player n conjectures a strategy of player m, which is a mapping $s_m^{(n)} : \mathbf{T}_m \to \Delta(\mathbf{A}_m)$. Let $s^{(n)} = (s_n^{(n)}, s_m^{(n)})$. At the next order, n conjectures that m conjectures a strategy of n, $s_n^{(n,m)} : \mathbf{T}_n \to \Delta(\mathbf{A}_n)$, and own strategy $s_m^{(n,m)} : \mathbf{T}_m \to \Delta(\mathbf{A}_m)$. Let $s^{(n,m)} = (s_n^{(n,m)}, s_m^{(n,m)})$, and so on at higher orders of conjectures. We define $s_n = (s^{(n)}, s^{(n,m)}, ...)$, as the *strategy architecture* of player n, and we call $s = (s_n, s_m)$ the strategy-architecture profile (from now on, a profile). The payoff relevant components of a profile s are $(s_n^{(n)}, s_m^{(m)})$.

The set of outcomes is $\mathbf{T} \times \Delta(\mathbf{A}_1) \times \Delta(\mathbf{A}_2)$, and the outcome corresponding to a profile s depends only on the payoff-relevant components $(s_n^{(n)}, s_m^{(m)})$. The payoffs corresponding to deterministic outcomes are given by $u : \mathbf{T} \times \mathbf{A} \to R^2$. Payoffs to a profile s are $U_i(s, t_i) = E_{P|t_i} E_{(s_n^{(n)}, s_m^{(m)})}[u(a, t_i, t_j)]$, where $i, j \in \mathbf{N}$, $j \neq i$.

Again, players make conjectures about others' awareness and others' conjectures and so on (awareness architecture). Denote by $T^{(n)} = T_1^{(n)} \times T_2^{(n)} \subset \mathbf{T}$ the type-awareness of player n, so that $T_1^{(n)}$ are types of player 1 that player n is aware of, $n \in \{1, 2\}$. A player is aware of $u(a, t) = (u_1(a, t), u_2(a, t))$ if and only if he is aware of t. Along with $T^{(n)}$, player n also conjectures a distribution $P^{(n)}$ over $T^{(n)}$. Zeroth order conjecture of agent n (or his awareness) is thus $(T^{(n)}, P^{(n)})$. As before, a first-order conjecture of agent n about m's awareness $(T^{(m)}, P^{(m)})$ is denoted by $(T^{(n,m)}, P^{(n,m)}) = (T_1^{(n,m)} \times T_2^{(n,m)}, P^{(n,m)})$, where $P^{(n,m)}$ is a joint distribution over $T^{(n,m)}$ and so on. Define the awareness architecture of agent n by $c_n = (T^{(n)}, P^{(n)}, T^{(n,m)}, P^{(n,m)}, T^{(n,m,n)}, P^{(n,m,n)}, ...)$, $n, m \in \mathbf{N}$. The set of all possible awareness architectures for player n is $C_n \subset \left(\{0,1\}^{\mathbf{T}} \times \{0,1\}^{\Delta(\mathbf{T})}\right) \times \left(\{0,1\}^{\mathbf{T}} \times \{0,1\}^{\Delta(\mathbf{T})}\right) \times ....$ It is more convenient to write $C_n = C_{n,T} \times C_{n,P}$, where $C_{n,T} \subset \{0,1\}^{\mathbf{T}} \times \{0,1\}^{\mathbf{T}} \times ...$, and $C_{n,P} \subset \{0,1\}^{\Delta(\mathbf{T})} \times \{0,1\}^{\Delta(\mathbf{T})}....$ The space of possible awareness architectures is $\mathbf{C} = C_1 \times C_2$. A normal form Bayesian game with type awareness is $\mathcal{U} = (\mathbf{N}, \mathbf{A}, \mathbf{T}, P, \mathbf{u}, \mathbf{C})$.

Recall that we defined for a finite sequence k of length x, (k, n) to be a sequence of length $x+1$, such that $(k,n)_i = k_i$ for each $i \leq x$ and $(k,n)_{i+1} = n$. Denote by $P(t_n)$ the marginal probability of t_n under the joint distribution P, i.e., $P(t_n) = \sum_{t_m \in \mathbf{T}_m} P(t_n, t_m)$. Let V_n be the set of all possible utility payoffs that player n can obtain, given types and actions of both players, i.e., $v \in V_n$ iff there exist $a \in \mathbf{A}$, and $t \in \mathbf{T}$ s.t. $v = u_n(a,t)$. Denote by $\Pr(a_m \mid s_m^{(n)}, P^{(n)}, t_n)$ the probability of action $a_m \in \mathbf{A}_m$, conditional on t_n, and given the conjectures $s_m^{(n)}, P^{(n)}$ (similarly for other orders of conjectures). Finally, denote by $\Pr(a_m, v \mid s_m^{(n)}, P^{(n)}, t_n)$ the joint probability of action $a_m \in \mathbf{A}_m$ and payoff $v \in V_n$, given $t_n, s_m^{(n)}, P^{(n)}$, and analogously for higher orders of conjectures.

Definition **3.** A *type-awareness equilibrium* (TAE) is a profile s and awareness architectures $(c_1, c_2) \in C_1 \times C_2$ such that

TAE1 $T_l^{(k,m)} \subset T_l^k$, and if $k = (n, ..., m)$, then $T^{(k,m)} = T^k$, $\forall l \in \mathbf{N}$, for all $k \in \mathbf{N}^x, \forall x < \infty$, $n, m \in \mathbf{N}$.

TAE2 If $P(t_n) > 0$ then $t_n \in T_n^{(n)}$ and $P^{(n)}(t_n) = P(t_n), \forall t_n \in T_n$.
If $P^{(k,n)}(t_m) > 0$ then $t_m \in T_m^{(k,n,m)}$ and $P^{(k,n,m)}(t_m) = P^{(k,n)}(t_m), \forall t_m \in T_m^{(k,n)}$, $\forall m, n \in \mathbf{N}, \forall k \in \mathbf{N}^x, \forall x < \infty$.

TAE3 $s^{(k,n)} : T^{(k,n)} \to \Delta(\mathbf{A}_n)$ and $s_m^{(k,n,m)} = s_m^{(k,n)}$.

TAE4 $\Pr(a_m, v \mid s_m^{(k,n)}, P^{(k,n)}, t_n) = \Pr(a_m, v \mid s_m^{(m)}, P^{(k)}, t_n)$, $\forall a_m \in \mathbf{A}_m$, $\forall m, n \in \mathbf{N}, m \neq n, \forall v \in V_n, \forall t_n \in T_n^{(k,n)} \forall k \in \mathbf{N}^x, \forall x \geq 0$.

TAE5 $s_n^{(k,n)} = \arg\max_{\hat{s}_n^{(k,n)} : T_n^{(k,n)} \to \Delta(\mathbf{A}_n)} E_{s_m^{(k,n)}} E_{P^{(k,n)} | t_n} \left[u(\hat{s}_n^{(k,n)}, s_m^{(k,n)}, t_n, t_m) \right], \forall t_n \in T_n^{(k,n)}$, $k \in \mathbf{N}^x, \forall x < \infty$, $n, m \in \mathbf{N}$.

TAE1 and TAE2 require consistency of equilibrium awareness architectures. TAE1 requires that a player cannot make conjectures about types that he is not aware of and that he is aware of his own conjectures. TAE2 requires that at each order, a player's conjecture about the distribution over types is consistent with the marginal distribution over his types under the distributional conjecture at the previous order. In particular, the marginal distribution over own types must be consistent with the true prior. Again, each player is aware of his own conjectures.

TAE3 and TAE4 require consistency of equilibrium strategy architectures. TAE3 requires that the strategy architecture is consistent with conjectures on type spaces, and that players are aware of their own strategies.

TAE4 requires that conditional on his own type, a player's strategy architecture, along with his awareness architecture, are consistent with the joint distribution over the actions of the other player and his own payoffs, at every order. To illustrate, suppose that at zero order, player 1 makes conjecture $P^{(1)}, T^{(1)}$, and $s^{(1)}$. This induces a (conjectured) probability distribution over 2's actions and 1's payoffs, conditional on his type. TAE4 requires that this be consistent with the distribution over 2's actions and 1's payoffs, conditional on 1's type, induced by 2's true strategy, and the true probability distribution over types. Similarly, at the first order, player 1 must conjecture that player 2's observation (induced by $P^{(1,2)}$, $s^{(1,2)}$, $T^{(1,2)}$), equals to what player 1 thought player 2's observation should be (as induced by $P^{(1)}$, $s^{(1)}$, $T^{(1)}$).[8]

TAE5 requires that players are best responding to the perceived randomization over the opponent's actions, and that they conjecture (at every order) that players are best replying.

In summary, in a TAE, players may have wrong conjectures about types, probability distribution over types, and the strategies that each is playing, as long as these are empirically consistent, and consistent with optimizing behavior.

How far is this model from the standard Harsanyi's formulation of a Bayesian game? If $\mathbf{C}_n = \{\mathbf{T} \times \mathbf{P} \times \mathbf{T} \times \mathbf{P}...\}, \forall n \in \mathbf{N}$, then $\mathbf{N}, \mathbf{T}, \mathbf{A}, \mathbf{u}, \mathbf{P}, \mathbf{C}$ is equivalent to the Bayesian game $\mathbf{N}, \mathbf{T}, \mathbf{A}, \mathbf{u}, \mathbf{P}$, and TAE is equivalent to Bayesian equilibrium.

[8] As we show below, in a private-value environment, TAE4 is equivalent to requiring that a player correctly conjectures the observed empirical distribution over the other player's actions.

Indeed, note that TAE1,TAE2, and TAE3 are then vacuously satisfied, TAE4 is the standard requirement that in equilibrium beliefs are correct, and TAE5 requires that agents play best replies. Our departure from the standard model is the departure from common knowledge of the type space and priors, which is embodied in TAE1-TAE3. We do not change the assumption of players' rationality, which is embodied in TAE4 and TAE5.

We now provide two examples to familiarize the reader with the definition of TAE. In the first example we illustrate how **C** affects the set of TAE. It is an example with private values, where an outcome of TAE can always be sustained as a Bayesian equilibrium outcome under the true prior, which we show in the next section.

EXAMPLE **2. Coordination-anticoordination game.** Define a normal-form game by $\mathbf{N} = \{1, 2\}, A_1 = \{up, down\}, A_2 = \{L, R\}, T_1 = \{+\}, T_2 = \{+, -\}, P(+, +) = \pi, P(+, -) = 1 - \pi, \pi \geq \frac{2}{3}$, and the payoffs are given by the following two payoff matrices:

$(t_1 = t_2 = +)$	L	R
up	$1, 1$	$0, 0$
down	$0, 0$	$2, 2$

$(t_1 = -t_2)$	L	R
up	$0, 1$	$2, 0$
down	$1, 0$	$0, 2$

In this game, depending on his type, player 2 either wants to coordinate with 1, or he wants to not coordinate.

Player 1's mixed strategy is $(\sigma, 1-\sigma)$, where $\sigma \in [0, 1]$ is the probability of playing *up*. Player 2's mixed strategy is $(\sigma_+, 1 - \sigma_+, \sigma_-, 1 - \sigma_-)$, where σ_t is the probability of playing L if $t_2 = t$. It can be easily verified that Bayesian equilibria of this game are $(\sigma = 0; \sigma_+ = 0; \sigma_- = 1)$, $(\sigma = 1; \sigma_+ = 1; \sigma_- = 0)$, $\left(\sigma = \frac{2}{3}; \sigma_+ = \frac{2}{3\pi}; \sigma_- = 0\right)$.

We now show how **C** affects TAE, and in particular, what **C** must include for existence.

First, if **C** includes the awareness architecture where **T** and **P** are common knowledge then all Bayesian equilibrium outcomes are supportable in TAE. Simply take all conjectures at every order to be equal to the true prior over the true type space, and each agent conjecturing the true strategy that the other is playing.

Second, suppose **C** does not include the awareness architecture where **T** and **P** are common knowledge. Then, neither of the two pure-strategy equilibria can be supportable if **T** and **P** are not common knowledge. For instance, consider the Bayesian equilibrium ($\sigma = 1; \sigma_+ = 1; \sigma_- = 0$). In a TAE, since 1 is playing *up*, he must be aware of both types of 2, and since for each of his two types, 2 has a distinct best reply, 1 has to correctly conjecture **P** as well.

If **C** only includes architectures where $T_2^{(1)} = \{-\}$ then no TAE exist. Indeed, under such conjecture and 1's randomization of *up* with probability $\frac{2}{3}$, in 1's view player 2 does not play a best reply by randomizing, so that TAE5 would be violated.

Finally, if **C** includes architecture where $T_2^{(1)} = \{+\}$, and $P^{(1)}(t_2 = +) = 1$, then only the outcome of the Bayesian equilibrium $\left(\sigma = \frac{2}{3}; \sigma_+ = \frac{2}{3\pi}; \sigma_- = 0\right)$ can be sustained as a TAE outcome. To see that, note that if 2 plays $\left(\sigma_+ = \frac{2}{3\pi}; \sigma_- = 0\right)$, then 1 observes L with probability $\pi\frac{2}{3\pi} = \frac{2}{3}$, and R with probability $\frac{1}{3}$. Given this, and $T_2^{(1)} = \{+\}$, 1 is indifferent between *up* and *down*, and playing $\sigma = \frac{2}{3}$ is a best reply that in 1's view makes 2 indifferent between his actions. In other words, $s_2^{(1)} = (\frac{2}{3}, \frac{1}{3})$, and satisfies TAE5. Similarly, we can construct higher-order conjectures for player 1. In turn, player 2 could either have correct conjectures about player 1's awareness architecture, or he could conjecture that player 1 is fully aware of 2's types. Thus, if **C** includes only this awareness architecture, a unique TAE exists, and is outcome equivalent to the mixed-strategy Bayesian equilibrium.

The next example is taken from Jackson and Kalai [1996]. In this example we illustrate how an outcome of a TAE, need not be sustainable in a Bayesian equilibrium under the true prior. However, in this example, that outcome will be sustainable in a Bayesian equilibrium under *some* prior.

EXAMPLE **3. Affirmative action, Example 5, Jackson and Kalai [1996].**

$\mathbf{N} = \{1, 2\}, A_1 = \{up, down\}, A_2 = \{L\}, T_1 = \{t_1\}, T_2 = \{t_2, t'_2\}$, $P(t_1, t_2) = \frac{1}{4}, P(t_1, t'_2) = \frac{3}{4}$, and the payoffs are given by the following two payoff matrices:

(t_1, t_2)	L
up	$1, 0$
down	$0, 0$

(t_1, t'_2)	L
up	$-2, 0$
down	$0, 0$

Let $T^{(k)} = \mathbf{T}, \forall k$, $P^{(1)}(t_1, t_2) = P^{(1)}(t_1, t'_2) = \frac{1}{2}$, $P^{(2)} = P$, and let $P^{(k)} = P^{(1)}, \forall k \in \mathbf{N}^x, x \geq 2$. Now let $s^{(k)} = (down, L)$. Then (c, s) thus defined constitutes a TAE, which is easy to check. It is clear that $(down, L)$ is not a Bayesian equilibrium under P. However, $(down, L)$ is a Bayesian equilibrium under $\tilde{P} = P^{(1)}$.

The general question that we explore is how different is a consistent situation where we do not require common knowledge assumptions from the standard Harsanyi formulation where we do. This difference may be either in terms of behavior or the outcome observed by an outsider.

4 TAE and Bayesian equilibrium

We define three notions of equivalence between TAE and Bayesian equilibrium. Let (s, c) be a TAE of $(\mathbf{N}, \mathbf{A}, \mathbf{T}, \mathbf{P}, \mathbf{u}, \mathbf{C})$ and let s^* be a Bayesian equilibrium of $(\mathbf{N}, \mathbf{A}, \mathbf{T}, \tilde{\mathbf{P}}, \mathbf{u})$. We say that (s, c) and s^* are *observationally equivalent* if $\Pr(a \mid s_1^{(1)}, s_2^{(2)}, P) = \Pr(a \mid s^*, \tilde{P}), \forall a \in \mathbf{A}$, that is, if the joint distribution over actions is the same in both equilibria.

We say that (s, c) and s^* are *weakly outcome equivalent* if $\Pr(a, v \mid s_1^{(1)}, s_2^{(2)}, P) = \Pr(a, v \mid s^*, \tilde{P}), \forall a \in \mathbf{A}, \forall v \in V_1 \times V_2$, that is, if the joint distribution over actions and payoffs is the same in both equilibria. Finally, we say that (s, c) and s^* are *strongly* *outcome equivalent* if $\Pr(a, v \mid s_1^{(1)}, s_2^{(2)}, P) = \Pr(a, v \mid s^*, \tilde{P}), \forall a \in \mathbf{A}, \forall v \in V_1 \times V_2$, and $\tilde{P} = P$.

Observational equivalence describes a situation where an outside observer knows the payoff structure and asks whether it is possible to justify observed behavior as Bayesian equilibrium behavior under some prior. Under weak outcome equivalence, the observer also knows the equilibrium payoffs that were attained by players. Under strong outcome equivalence, the observer knows the true distribution over the states of the world. Note that strong outcome equivalence implies weak outcome equivalence, which implies observational equivalence but the implications do not go in the other direction.

In order to obtain strong outcome equivalence, it is sufficient that a player can compute a best reply just by observing the distribution over own payoffs and opponent's actions. We first show that in two commonly used classes of environments this is the case. We say that a game is a private-value game if $u_n(a, t_n, t_m) = u_n(a, t_n, t'_m), \forall t_m, t'_m \in T_m, \forall t_n \in T_n, \forall a \in \mathbf{A}$. We say that a game is a strict common value game if $u_n(a, t_n, t_m) \neq u_n(a, t_n, t'_m), \forall t_m, t'_m \in T_m, \forall t_n \in T_n, \forall a \in \mathbf{A}$.

Proposition **3.** Let $(\mathbf{N}, \mathbf{A}, \mathbf{T}, \mathbf{P}, \mathbf{u}, \mathbf{C})$ be a type-awareness game. If it is either a private-value game or a strict common-value game then all its TAE are strongly outcome equivalent to Bayesian equilibria of $(\mathbf{N}, \mathbf{A}, \mathbf{T}, \mathbf{P}, \mathbf{u})$.

Proof. Take private values, and let (s, c) be a TAE. By TAE5, $s_1^{(1)}$ is a best reply to $s_2^{(1)}$ under the distribution $P^{(1)}$. TAE4 implies that

$$\Pr(a_2 \mid t_1, s_2^{(1)}, P^{(1)}) = \Pr(a_2 \mid t_1, s_2^{(2)}, P), \forall t_1 \in T_1, \forall a_2 \in A_2.$$

By private values, $u_1(a, t_1, t_2) = u_1(a, t_1), \forall (t_1, t_2) \in T, \forall a \in A$. Using these two observations, we have that

$$E_{s_2^{(1)},P^{(1)}} u_1(a, t_1, t_2) = E_{s_2^{(2)},P} u_1(a, t_1), \forall a_1 \in A_1.$$

Thus, $(s_1^{(1)}, s_2^{(2)})$ is a Bayesian equilibrium of $(\mathbf{N}, \mathbf{A}, \mathbf{T}, \mathbf{P}, \mathbf{u})$, which implies strong outcome equivalence.

Now consider strict common values. Take a $v_1 \in V_1$ (recall that V_1 is the set of utility payoffs for player 1), then $\exists a \in \mathbf{A}, (t_1, t_2) \in \mathbf{T}$, s.t. $v_1 = u_1(a, t_1, t_2)$ and by strict common values $u_1(a, t_1, t_2') \neq v_1, \forall t_2' \neq t_2$. Using this, we have that

$$\Pr(v_1, a_2 \mid t_1, s_2^{(1)}, P^{(1)}) = \Pr(t_2, a_2 \mid t_1, s_2^{(1)}, P^{(1)}), \tag{1}$$

so that

$$\Pr(t_2 \mid t_1, P^{(1)}) = \sum_{a_2 \in A_2} \Pr(t_2, a_2 \mid t_1, s_2^{(1)}, P^{(1)}).$$

Similarly,

$$\Pr(t_2 \mid t_1, P) = \sum_{a_2 \in A_2} \Pr(t_2, a_2 \mid t_1, s_2^{(2)}, P).$$

By TAE4 and (1),

$$\Pr(t_2, a_2 \mid t_1, s_2^{(1)}, P^{(1)}) = \Pr(t_2, a_2 \mid t_1, s_2^{(2)}, P), \forall t_2 \in T_2, \forall a_2 \in A_2,$$

which implies that $\Pr(t_2 \mid t_1, P^{(1)}) = \Pr(t_2 \mid t_1, P)$, i.e., 1 must correctly conjecture the type distribution, and similarly at higher orders. Thus, $(s_1^{(1)}, s_{(2)}^{(2)})$ is a Bayesian equilibrium. $\square$

Strict common values and private values provide a large class of economically interesting environments. To name just a few examples, public goods, market situations such as auctions and double auctions, many standard examples of oligopolistic competition, adverse selection models, and so on. Proposition 3 shows that in these environments, departing from the common knowledge requirements over the prior and type space does not alter the equilibrium predictions.

While strong outcome equivalence obtains in both of these environments, there is a difference. Namely, in the case of strict common values, in a TAE, the true prior needs to be common knowledge (modulo types with zero probabilities, of which the players do not need to be aware), and a TAE exists if and only if **C** includes such architecture. As shown in Example 2, in the case of private values the true prior need not be common knowledge. In the appendix A we provide two additional propositions for private-values. Proposition 5 provides conditions for existence of TAE, and Proposition 6 provides conditions under which in a TAE, players need to conjecture the true prior at every order.

Example 3 shows that strong outcome equivalence doesn't always obtain. We now provide a simple condition on the payoff structure of the game, under which strong outcome equivalence doesn't obtain.

Proposition 4. Suppose that for $m, n \in \mathbf{N}, m \neq n$, there exist $t_m \in T_m,\ t_n, t'_n \in T_n, a \in \mathbf{A}$ s.t. the following conditions hold:

1. $a_m \in \arg\max_{a'_m \in A_m} u_m(a_n, a'_m, t_n, t_m)$ and $a_m \notin \arg\max_{a'_m \in A_m} u_m(a_n, a'_m, t'_n, t_m)$,
2. $a_n \in \arg\max_{a'_n \in A_n} u_n(a'_n, a_m, t_n, t_m) \cap \arg\max_{a'_n \in A_n} u_n(a'_n, a_m, t'_n, t_m)$,
3. $u_m(a, t_n, t_m) = u_m(a, t'_n, t_m)$.

Then, under unrestricted **C**, there exists a **P**, and a TAE of $(\mathbf{N}, \mathbf{A}, \mathbf{T}, \mathbf{P}, \mathbf{u}, \mathbf{C})$, which is not strongly outcome equivalent to any Bayesian equilibrium of $(\mathbf{N}, \mathbf{A}, \mathbf{T}, \mathbf{P}, \mathbf{u})$.

Proof. Assume wlog that $n = 1, m = 2$, and let P be such that $P(t_1, t_2) + P(t'_1, t_2) = 1$ and $P(t_1, t_2)u_2(a_1, a_2, t_1, t_2) + P(t'_1, t_2)u_2(a_1, a_2, t'_1, t_2) < P(t_1, t_2)u_2(a_1, a'_2, t_1, t_2) + P(t'_1, t_2)u_2(a_1, a'_2, t'_1, t_2)$ (condition 1 implies that A_2 has at least two elements, and that it is possible to find such an a'_2).

Let $\tilde{P}$ be such that $\tilde{P}(t_1,t_2)+\tilde{P}(t'_1,t_2)=1$ and

$$\tilde{P}(t_1,t_2)u_2(a_1,a_2,t_1,t_2)+\tilde{P}(t'_1,t_2)u_2(a_1,a_2,t'_1,t_2)$$
$$\geq \tilde{P}(t_1,t_2)u_2(a_1,a'_2,t_1,t_2)+\tilde{P}(t'_1,t_2)u_2(a_1,a'_2,t'_1,t_2), \forall a'_2\in A_2, a'_2\neq a_2.$$

Now set $T^{(k)}=\{(t_1,t_2),(t'_1,t_2)\}$, $P^{(1)}=P, P^{(k)}=\tilde{P}, \forall k\neq(1)$, $s_1^{(k)}=a_1, s_2^{(k)}=a_1$, $\forall k$. It is immediate to verify that (s,c) is a TAE, and by construction, $(s_1^{(1)},s_2^{(2)})$ is not a Bayesian equilibrium under P, implying that strong outcome equivalence doesn't hold. □

The conditions in the above proposition build on the intuitions from Example 3. Condition 1 says that should exist two types of player n, such that for each of these player m has a different best reply. Condition 2 and 3 state that in order for m to not be able to discern the relative likelihoods of these two types, n must play the same action for both, and m must obtain the same payoffs. It easy to verify that in Example 3, these conditions are satisfied by taking $n=2$, $m=1$.

Still, in Example 3 one can easily verify that weak outcome equivalence obtains, and one might be tempted to believe that this is always the case. In the rest of this section we provide an example where we show that even observational equivalence doesn't always hold. That is, there exist games with TAE, which are not supportable as Bayesian equilibria under any prior. In such a game, an outside observer might observe a joint distribution over players' actions which could not be justified as a Bayesian equilibrium of that game under any prior. Given players' rationality, in such a case, the hypothesis of unawareness can be verified.

4.1 Failure of observational equivalence

The example is constructed on two principles. First, it satisfies the three conditions of Proposition 4 for both players. This allows to sustain an outcome in a TAE, in which each player is unable to discern relative likelihoods of two draws of types, call them (t_1, t_2), (t_1, t'_2) for player 1, and $(t_1, t_2), (t'_1, t_2)$ for player 2. The second principle is to construct payoffs in such a way that this TAE outcome cannot be sustained as a Bayesian equilibrium under any prior. Payoffs have to be such that player 1 must conjecture that (t_1, t_2) is sufficiently unlikely relative to (t_1, t'_2), while 2 must conjecture that (t_1, t_2) is sufficiently likely, relative to (t'_1, t_2).

Example **4.** Let each player have two types, $T_n = \{t_n, t'_n\}$, and two actions, and let the payoff structure for each draw of types be specified as follows.

(t_1, t_2)	L	R
u	$0, 6$	$0, 0$
d	$0, 0$	$-10, 0$

(t_1, t'_2)	L	R
u	$0, 0$	$0, 1$
d	$0, 0$	$4, 0$

(t'_1, t_2)	L	R
u	$0, -1$	$1, 0$
d	$0, 0$	$3, 0$

(t'_1, t'_2)	L	R
u	$0, 30$	$2, 0$
d	$0, 0$	$-50, 0$

Let the true prior distribution P be given by the following.

P	t_2	t'_2
t_1	$\frac{3}{10}$	$\frac{3}{10}$
t'_1	$\frac{3}{10}$	$\frac{1}{10}$

Now we will construct a TAE in which each type of player 1 plays u, and each type of player 2 plays R. Let's denote this strategy profile by $s^* = (s_1^{(1)}, s_2^{(2)})$. Then we will show that there does not exist a common prior $\tilde{P}$ under which this strategy profile would be a Bayesian equilibrium.

We start by considering the conditions on $P^{(1)}$ and $P^{(2)}$ that need to be satisfied to sustain s^* in a TAE. First, note that by TAE4, in order to sustain s^*, it has to be that $s^{(k)} = s^*$, for all $k \in N^x, x > 0$.

Notice that when player 2 plays R, and player 1 is of type t_1', player 1 must by TAE4 correctly conjecture the conditional distribution on player 2's types, i.e., $P^{(1)} |_{t_1'} (x) = P |_{t_1'} (x)$, $x \in \{t_2, t_2'\}$. This is so because player 1 obtains different payoffs for different types of player 2. Similarly, player 2 must correctly conjecture the conditional distribution on player 1's types when he is type t_2'.

By TAE5, in order for t_1 to play u when 2 plays R, $P^{(1)}$ must satisfy the following.

$$-10P^{(1)}(t_1, t_2) + 4P^{(1)}(t_1, t_2') \leq 0, \tag{2}$$

Similarly, for t_2 to play R when 1 plays u, $P^{(2)}$ must satisfy

$$6P^{(2)}(t_1, t_2) - 1P^{(2)}(t_1', t_2) \leq 0. \tag{3}$$

Also, by TAE2, $P^{(1)}(x, t_2) + P^{(1)}(x, t_2') = P(x, t_2) + P(x, t_2')$, $x \in \{t_1, t_1'\}$ and $P^{(2)}(t_1, x) + P^{(2)}(t_1', x) = P(t_1, x) + P(t_1', x)$, $x \in \{t_2, t_2'\}$. All these conditions are satisfied by the following $P^{(1)}$, $P^{(2)}$.

$P^{(1)}$	t_2	t_2'
t_1	$\frac{2}{10}$	$\frac{4}{10}$
t_1'	$\frac{3}{10}$	$\frac{1}{10}$

$P^{(2)}$	t_2	t_2'
t_1	0	$\frac{3}{10}$
t_1'	$\frac{6}{10}$	$\frac{1}{10}$

Now define the higher order conjectures on P inductively as follows. For $k \in N^x, x \geq 0$, given $P^{(k,1)}$, define $P^{(k,1,2)}$ such that $P^{(k,1,2)}(t_1, t_2) + P^{(k,1,2)}(t_1', t_2) = P^{(k,1)}(t_1, t_2) + P^{(k,1)}(t_1', t_2)$, and such that (3) is satisfied, and set $P^{(k,1,2)}(t_1, t_2') = P^{(k,1)}(t_1, t_2')$, $P^{(k,1,2)}(t_1', t_2') = P^{(k,1)}(t_1', t_2')$. In this way, $s_2^{(k,1)}$ and $P^{(k,1,2)}$ satisfy TAE5, and by construction TAE2 is satisfied by $P^{(k,1,2)}$. Analogously, given $P^{(k,2)}$, construct $P^{(k,2,1)}$, such that condition (3) holds (implying that TAE5 holds for $s_1^{(k,2)}$ and $P^{(k,2,1)}$) and TAE2 holds. Such construction is possible, and setting $c_n = (T, P^{(n)}, T, P^{(n,m)}, ...)$, $s^{(k)} = s^*$, $\forall k \in N^x, x > 0$, (s, c) is a TAE.

Now we show that there does not exist a common prior $\tilde{P}$, such that in the game $(\mathbf{N}, \mathbf{A}, \mathbf{u}, \mathbf{T}, \tilde{\mathbf{P}})$, s^* is a BE.

We first show that for no $\tilde{P}$ that puts mass 1 on a single type of one player s^* is a BE. We have to consider 10 different cases. The 4 cases where $\tilde{P}$ puts mass 1 on a single draw of types are trivial to check, and the other 6 cases are also simple. To illustrate the logic take for example a $\tilde{P}$ that puts mass 1 on $\{(t_1, t_2), (t_1', t_2)\}$. If player 1 is type t_1' and player 2 plays R, one would play d.

For every $\tilde{P}$ that puts 0 probability on exactly one draw of types, it is also easy to check in the same way that s^* is not supportable in a BE. The last case is when $\tilde{P}$ puts positive probability on all draws. Denote $p_1 = \tilde{P}(t_1, t_2)$, $p_2 = \tilde{P}(t_1, t_2')$, $p_3 = \tilde{P}(t_1', t_2)$, $p_4 = \tilde{P}(t_1', t_2')$, and write the incentive constraints,

$$0 \geq 4p_2 - 10p_1, \tag{4}$$

$$0 \geq 2p_3 - 52p_4, \tag{5}$$

$$0 \geq 6p_1 - p_3, \tag{6}$$

$$0 \geq 30p_4 - p_2. \tag{7}$$

Add (5) and $4\times$ (7), add (6) and $2\times$ (7), and add $5\times$ the first resulting inequality to $6\times$ the second one, to obtain $0 \geq 460p_4$, which is a contradiction.

Before concluding we remark that in the TAE from Example 4 each player conjectures that the other player holds a different conjecture about the probability over some types, at every order. Since both verify the distribution of actions in equilibrium, this is common knowledge, and therefore players agree to disagree. A subtle point here is that this is a consequence of the fact that in this TAE, there is no common prior that would support it as a Bayesian equilibrium.

5 Discussion

We discuss some relations between the material we presented so far, and some possible extensions. First, in the model of type awareness we allow for conjectures over types and probability distributions over these. One may wonder whether a model where the type space is common knowledge and conjectures are only over the distributions over types is rich enough. In principle, this is not the case, but as the following remark illustrates, it is the case in equilibrium.

Remark **1.** Consider two type awareness games, $(\mathbf{N}, \mathbf{A}, \mathbf{T}, \mathbf{u}, \mathbf{P}, \mathbf{C})$, where $\mathbf{C}$ is unrestricted, and $(\mathbf{N}, \mathbf{A}, \mathbf{T}, \mathbf{u}, \mathbf{P}, \bar{\mathbf{C}})$, where $\bar{C}_{n,T} = \{\mathbf{T}\} \times \{\mathbf{T}\} \times ...$, and $\bar{C}_{n,P} = \{0,1\}^{\Delta(\mathbf{T})} \times \{0,1\}^{\Delta(\mathbf{T})}...$ for both n. In other words, $\mathbf{T}$ is common knowledge in $(\mathbf{N}, \mathbf{A}, \mathbf{T}, \mathbf{u}, \mathbf{P}, \bar{\mathbf{C}})$. Then the set of TAE of $(\mathbf{N}, \mathbf{A}, \mathbf{T}, \mathbf{u}, \mathbf{P}, \mathbf{C})$ is equal to the set of TAE of $(\mathbf{N}, \mathbf{A}, \mathbf{T}, \mathbf{u}, \mathbf{P}, \bar{\mathbf{C}})$.

Proof. Whenever a player is not aware of some type, this is in equilibrium equivalent to putting probability 0 on that type. This shows that $\{TAE \text{ of } (\mathbf{N}, \mathbf{A}, \mathbf{T}, \mathbf{u}, \mathbf{P}, \mathbf{C})\} \subset \{TAE \text{ of } (\mathbf{N}, \mathbf{A}, \mathbf{T}, \mathbf{u}, \mathbf{P}, \bar{\mathbf{C}})\}$. The other inclusion is trivial. □

We also remark that if in a TAE a player puts probability 0 to some type at every order, this is equivalent to him being unaware of that type.

Second, we presented two models of awareness equilibrium. In the first one, awareness is about the action sets, in the second one, the action sets are common knowledge, and awareness is about the type space. The richness of the latter framework suggests that we could model action awareness equilibria of a given complete information game, as type-awareness equilibria of some carefully constructed game with incomplete information. It may be possible to construct the appropriate game using the intuitions from Example 4. The difficulty is to construct a game with incomplete information that has TAE which are not "too dependent" on the prior. This is a question that we want to explore further.

Finally, in the present work, we explored a situation where agents can verify the joint distribution over actions of the opponent and own payoffs. But there are many interesting situations, where players cannot verify actual actions taken by the other player but just some statistic of those. An example of such a situation is moral hazard, and more generally, models of hidden actions. On the one hand, our model is flexible enough to extend to such environments - one would have to modify the definition of TAE slightly. On the other hand, the TAE considered here will always be equilibria under those circumstances as well (just consider a situation where a player happens to *conjecture* the true distribution over the opponents actions). Nevertheless, this may be useful for possible applications of our model. It is also a necessary step for writing down a model where agents are unaware of other agents-if an agent is aware of some other agent's actions he needs to be aware of that agent.

Appendix A–Action Awareness Equilibrium

In this Appendix we first provide a formal definition of ℓ^*-AAE. Second, we construct an example to illustrate the existence of non-generic games in which, for very $l < K$, every outcome can be supported in an l^*-AAE. Third, we prove Theorem 1. Fourth, we provide and prove an analogous result to Theorem 1 for ℓ^*-AAE (Theorem 2). Finally, we provide an example which illustrates some other feature of ℓ^*-AAE.

Definition **4.** Fix an $\ell \geq 1$ and let $a^* = (\underline{a}^*, \bar{a}^*)$ be a pure strategy Nash equilibrium of the game $\Gamma = (\mathbf{N}, \mathbf{A}, \mathbf{u})$. An ℓ^*-Action-awareness equilibrium, ℓ^*-AAE, is an ℓ-AAE where $a^* \in A^{(n)}, n = 1, 2$.

An ℓ^*-AAE, is an ℓ-AAE which is an equilibrium even if some Nash-equilibrium profile is in the awareness sets of both players.

In the following example we construct a non-generic and non-trivial game where there is a unique Nash equilibrium, but for very $l < K$ every outcome can be supported in an l^*-AAE.

Example **5.** For each K there exists a $\Gamma = (\mathbf{N}, \mathbf{A}, \mathbf{u}), |A_n| = K, \forall n \in \{1, 2\}$, such that the following holds. Γ has a unique pure-strategy Nash Equilibrium, let $(\underline{1}, \bar{1})$ be the unique NE. Then for each $\ell, 2 \leq \ell < k$, every outcome is sustainable as an l^*-AAE.

To see this, consider the following game. To define u, take first the matrix for the row player, $u_1(\underline{p}, \bar{q}), 1 \leq p, q \leq K$. Let $u_1(\underline{1}, \bar{q}) = 1, u_1(\underline{q}, \bar{1}) = 0, q = 1, ..., K$. For each column $p = 2, ..., K$, assign a 1 in precisely one unassigned location in such a way that the assigned 1's don't lie in only one row. This can obviously be done. Let $u_1(\underline{p}, \bar{q}) = 0$ for all the other locations. Take player 2 and do exactly the same, but also take care so that $(u_1(\underline{p}, \bar{q}), u_2(\underline{p}, \bar{q})) \neq (1, 1)$ for $(\underline{p}, \bar{q}) \neq (\underline{1}, \bar{1})$. Since the 1s assigned to columns of player 1 are not in the same row, such assignment is possible (reader can easily verify that). See Figure 2 for an example of such a game.

$_1\backslash^2$	$\bar{1}$	$\bar{2}$	$\bar{3}$
$\underline{1}$	$1,1$	$1,0$	$1,0$
$\underline{2}$	$0,1$	$1,0$	$0,1$
$\underline{3}$	$0,1$	$0,1$	$1,0$

Figure 2

Now we have to show that u has the desired properties. Clearly, the profile $s = (\underline{1}, \bar{1})$ is a pure strategy NE of Γ. To show that this is the unique pure-strategy NE of Γ observe that for every $(p, q) \in \{1, ..., K\}^2, (p, q) \neq (1, 1)$, at least one player gets a 0. Suppose (wlog) it is the row player 1. Then, by construction there is another column q' such that $u_1(\underline{p}, \bar{q}') = 1$ so that 1 would want to deviate.

To show that every outcome is an ℓ^*-AAE, for all $\ell < K$, observe first that if an outcome is an ℓ^*-AAE, $\ell > 2$ then it must be an $(\ell-1)^*$-AAE (reduce the supporting awareness set of each player by one action). Thus, it is enough to show the claim for $\ell = K - 1$. So take an outcome $(\underline{p}, \bar{q}) \in \{1, ..., k\}^2, (p, q) \neq (1, 1)$, and suppose that $u_1(\underline{p}, \bar{q}) = 0$ (if it is 1, then there is no deviation for player 1 anyway). This is not column 1, since there player 1 gets 1. By construction there are $K - 2$ other rows in column p such that player 1 gets 0 in those rows, and taking those $K - 2$ rows and row p also includes row 1. Similarly for player 2, so that we have constructed the awareness sets which include action 1 for both players, and no player has a profitable deviation from the profile $(\underline{p}, \bar{q})$.

Proof. (Theorem 1) By Lemma 1 we can focus on ℓ-AAE with the property that agents are aware of the same actions and make correct conjectures. Let player 1 be the row player. Given an ordered set S, Denote by $S_{(r)}$ the r-th order statistic of S.

Step 1. A profile $(\underline{p}, \bar{q})$ is supportable as an ℓ-AAE if and only if

$$u_1(\underline{p}, \bar{q}) \geq \{u_1(\underline{1}, \bar{q}), ..., u_1(\underline{K}, \bar{q})\}_{(\ell)} \text{ and } u_2(\underline{p}, \bar{q}) \geq \{u_1(\underline{p}, \bar{1}), ..., u_2(\underline{p}, \bar{K})\}_{(\ell)}. \quad (8)$$

For illustration, suppose first that $\ell = 2$. Under genericity, the claim is that a strategy profile $(\underline{p}, \bar{q})$ is then sustainable as an 2-AAE if and only if

$$u_1(\underline{p}, \bar{q}) > \min_{p' \in \{1,...,K\}} u_1(\underline{p}', \bar{q}) \text{ and}$$

$$u_2(\underline{p}, \bar{q}) > \min_{q' \in \{1,...,K\}} u_2(\underline{p}, \bar{q}').$$

To see the only if part, suppose that $u_1(\underline{p}, \bar{q}) = \min_{p' \in \{1,...,K\}} u_1(\underline{p}', \bar{q})$. By genericity of Γ it is therefore $u_1(\underline{p}, \bar{q}) < u_1(\underline{p}', \bar{q}), \forall p' \neq p$. This implies that regardless of what other row p' comprises $A^{(1)}$, player 1 will at the profile $(\underline{p}, \bar{q})$ deviate to $\underline{p}'$.

To see the if part suppose that a profile $(\underline{p}, \bar{q})$ satisfies the above condition. There exist a $p' \neq p$ and a $q' \neq q$ such that $u_1(\underline{p}, \bar{q}) > u_1(\underline{p}', \bar{q})$ and $u_1(\underline{p}, \bar{q}) > u_1(\underline{p}, \bar{q}')$. Let $A^{(1)} = A^{(2)} = \{\underline{p}, \underline{p}', \bar{q}, \bar{q}'\}$, and $(\underline{p}, \bar{q})$ is an 2-AAE outcome supported by such awareness structure. Similarly, we prove the claim for general ℓ. Note that we do not need genericity in this step. End of Step 1.

By genericity of Γ, there exists a strict ordering of 1's payoffs in each column, and a strict ordering of 2's payoffs in each row.

Step 2. $e_\ell(\Gamma) \geq K^2 - 2(\ell - 1)K$.

Fix an $\ell \in \{1, ..., K\}$. By Step 1, we will minimize the number of outcomes that can be supported under ℓ-AAE by "optimally" assigning the $\ell - 1$ lowest payoffs to player 1 in each column and $\ell - 1$ lowest payoffs to player 2 in each row. An allocation which minimizes the number of outcomes supportable as ℓ-AAE is one where all these payoffs are allocated to different profiles. Since there are K columns, $\ell - 1$ worse payoffs to 1 in each column, K rows, and $\ell - 1$ worse payoffs to 2 in each row, there are in total at most $2K(\ell - 1)$ action profiles that can be eliminated. This gives the desired lower bound on $e_\ell(\Gamma)$.

Step 3. $e_\ell(\Gamma) \le K^2 - (\ell - 1)K$.

Fix $\ell \in \{1, ..., K\}$. By Step 1, we will maximize the number of outcomes by allocating the $\ell - 1$ lowest elements of each row and each column in way which takes least space in the game matrix. That is achieved for instance by having every outcome which is the worst payoff in a given row for the column player to also be the worst payoff in the given column for the row player. Since there are K rows and columns and there are by genericity $\ell - 1$ strictly worst payoff in each, we can thus eliminate at least $K(\ell - 1)$ outcomes, which gives the desired upper bound on $e_\ell(\Gamma)$. □

We now provide an analogous result to Theorem 1, for the ℓ^*-AAE. We denote by $floor[x]$ the largest integer that is smaller than $x \in I\!R$, and by $mod[y, r]$ the leftover from integer division of an integer y with integer r.

THEOREM **2.** Let Γ be a generic $K \times K$ game. Denote by $e_N(\Gamma)$ the number of pure strategy Nash Equilibria of Γ and by $e_{\ell^*}(\Gamma)$ the number of ℓ^*-AAE of Γ. Then $e_N(\Gamma) \le e_{\ell^*}(\Gamma) \le floor[\frac{K-1}{K-\ell+1}](K - \ell + 1)^2 + (mod[K - 1, K - \ell + 1])^2 + 1$.

Proof. The lower bound is a consequence of the following simple Lemma.

LEMMA **2.** $e_N(\Gamma) = e_{\ell^*}(\Gamma)$ if and only if the following condition holds. For every profile $(\underline{p}, \bar{q})$ and every Nash Equilibrium profile $(\underline{p}^*, \bar{q}^*)$, either $u_1(\underline{p}, \bar{q}) \le u_1(\underline{p}^*, \bar{q})$ or $u_2(\underline{p}, \bar{q}) \le u_2(\underline{p}, \bar{q}^*)$.

Proof. The if part is obvious: regardless of what Nash Equilibrium is taken along with a strategy profile $(\underline{p}, \bar{q})$, one of the players has incentives to deviate (also by genericity) to the Nash Equilibrium strategy.

To see the only if part, take a profile $(\underline{p}, \bar{q})$ and suppose there exists a Nash Equilibrium $(\underline{p}^*, \bar{q}^*) \neq (\underline{p}, \bar{q})$ such that the above condition does not hold. Take $A^{(1)} = A^{(2)} = \{\underline{p}, \underline{p}^*, \bar{q}, \bar{q}^*\}$ and it is clear that $(\underline{p}, \bar{q})$ is an ℓ^*-AAE profile for $\ell = 2$. □

The upper bound is constructed via a "geometric" argument. Fix an $\ell, 2 \le \ell < K$, and we show by induction on ℓ and K that $e_{\ell^*}(\Gamma) \leq floor[\frac{K-1}{K-\ell+1}](K-\ell+1)^2 + (mod[K-1, K-\ell+1])^2 + 1$. Consider first a Γ, such that $e_N(\Gamma) = 1$, and assume without loss of generality that $(\underline{1}, \bar{1})$ is the Nash Equilibrium profile.

Suppose first that $\ell = 2$. Then we can for every K do the following. By genericity of Γ all the outcomes in the row 1 and column 1 cannot be sustained as ℓ^*-AAE. Also, without loss of generality, we make a construction where as many ℓ^*-AAE profiles as possible are concentrated in the lower right hand corner of the game bi-matrix. Consider a profile $(\underline{K}, \bar{K})$. This profile can be supported if in row K there is 1 outcome which is worse for player 1, and in column K there is 1 worse outcome for player 2. Moreover, $(\underline{1}, \bar{K})$ and $(\underline{K}, \bar{1})$ have to be worse for the corresponding player (since $\{\underline{1}, \bar{1}\} \subset A^{(i)}$ by definition of ℓ^*-AAE). By the same logic, all other outcomes in K-th row and K-th column can be sustained. Similarly, in all the rows $K-1, ..., 2$ the first outcome cannot be sustained but all the others can. The same applies to the columns.

Now let $2 < \ell < K$. Exactly as before, the outcomes in rows $\ell, ..., K$ and columns $\ell, ..., K$ are sustainable. If $2\ell - K - 1 \leq 1$ then all the outcomes in rows $2, ..., \ell - 1$ and columns $2, ..., \ell - 1$ can also be sustained as ℓ^*-AAE by making them higher than $\ell - 1$ outcomes in the succeeding rows and columns. In the first row and column only the Nash Equilibrium is sustainable.

If $2\ell - K - 1 > 1$, then consider the game Γ' obtained by taking the first $\ell - 1$ rows and $\ell - 1$ columns of Γ and let $\ell' = 2\ell - K - 1$. Now, the outcomes of Γ' that are sustainable as ℓ^*-AAE of Γ must be sustainable as $(\ell')^*$-AAE of Γ', so that $e_{\ell^*}(\Gamma) \leq (K - \ell + 1)^2 + e_{(\ell')^*}(\Gamma')$. The claim now follows from induction.

Note that the assumption that Γ has a unique Nash Equilibrium was made only for convenience, since if there are more Nash Equilibria, we can first re-arrange the players' actions so that all of those lie on the diagonal. □

The upper bound as stated in Theorem 2 is independent of the number of Nash

Equilibria. However, if a generic game has a unique Nash Equilibrium this imposes additional structure on the game, and the upper bound may never be attained. We illustrate this with an example of potential games[9]. Potential games are a very natural class to consider since a subgame of a potential game is also a potential game, and every potential game has at least one pure-strategy Nash Equilibrium. Many commonly studied games are potential games, e.g. prisoners' dilemma, congestion games, or Cournot games with quasi-linear demand.

EXAMPLE **6.** Let Γ be the following 4×4 potential game with $e_N(\Gamma) = 2$, given by the following matrix P.

P	$\bar{1}$	$\bar{2}$	$\bar{3}$	$\bar{4}$
$\underline{1}$	10	0	2	6
$\underline{2}$	1	3	5	2
$\underline{3}$	2	4	6	3
$\underline{4}$	5	8	7	9

Figure 4

Let $\ell = 2$, so that by Theorem 2 the upper bound on $e_{\ell^*}(\Gamma) = 10$. Clearly, the 9 right lower corner outcomes of Γ along with the left upper corner Nash Equilibrium constitute the set of 2^*-AAE of Γ, so that the upper bound is tight in this case. Also note that it is easy to extend the example to general potential games with at least 2 Nash Equilibria and different ℓ.

Consider now the potential game Γ with a unique pure-strategy Nash Equilibrium, given by the matrix $\tilde{P}$ below. The unique Nash Equilibrium of Γ is the profile $(\underline{1}, \bar{1})$.

[9]A game Γ is an *ordinal potential game* if there exists a potential function $P : \mathbf{A} \to I\!R$ which represents Γ in the following way: $u_1(\underline{p}, \bar{q}) - u_1(\underline{p}', \bar{q}) > 0 \iff P(\underline{p}, \bar{q}) - P(\underline{p}', \bar{q}) > 0$, and $u_2(\underline{p}, \bar{q}) - u_2(\underline{p}, \bar{q}') > 0 \iff P(\underline{p}, \bar{q}) - P(\underline{p}, \bar{q}') > 0, \forall \underline{p}, \underline{p}' \in \mathbf{A}_1, \forall \underline{q}, \underline{q}' \in \mathbf{A}_2$. A profile $(\underline{p}, \bar{q})$ is a pure-strategy Nash Equilibrium of Γ if and only if $P(\underline{p}, \bar{q}) \geq \max\{P(\underline{p}', \bar{q}); p' = 1, ..., K\} \cup \{P(\underline{p}, \bar{q}'); q' = 1, ..., K\}$. In particular the maximum of all elements of matrix P is a pure-strategy Nash Equilibrium of Γ. See Monderer and Shapley [1996].

P	$\bar{1}$	$\bar{2}$	$\bar{3}$	$\bar{4}$
$\underline{1}$	10	9	2	6
$\underline{2}$	1	3	5	2
$\underline{3}$	2	4	6	3
$\underline{4}$	9	8	7	4

Figure 5

Now observe that the path of best replies from each profile $(\underline{p}, \bar{q})$ eventually ends up in $(\underline{1}, \bar{1})$. This is in fact a property of potential games with a unique Nash Equilibrium. At some point such path enters either column or row 1, suppose that the path enters column 1 in row p (in $\tilde{P}$, $p = 3$). But this implies that $P(\underline{p}, \bar{1}) > P(\underline{p}, \bar{q}), q > 1$, so that no element in row p can be sustainable as an ℓ^*-AAE outcome, which means that the upper bound may never be attained in a potential game with a unique Nash Equilibrium. Nonetheless, the additional structure imposed by uniqueness of Nash Equilibrium eliminates only one additional row (or column), so that in a large game this effect is negligible.

Appendix B–Additional Results for Private-Value Games.

In this appendix we only consider private-value environments. The next proposition provides conditions for the existence of TAE.

Proposition **5.** Consider a private-value environment. Let $C_{n,P} = \{0,1\}^{\Delta(\mathbf{T})} \times \{0,1\}^{\Delta(\mathbf{T})}$.... Then a TAE exists if and only if there exists a $\bar{T} = \bar{T}_1 \times \bar{T}_2$ such that the following two properties hold.

1. There exists a Bayesian equilibrium s^* such that
   $\cup_{t_n \in T_n} supp\,(s_n^*(t_n)) = \cup_{t_n \in \bar{T}_n} supp\,(s_n^*(t_n))$.

2. There exists a $c_n \in C_n$ with the property that $(c_{n,T})_{(k,n)} = T_n$ and $(c_{n,T})_{(k,m)} = \bar{T}_m$, for all $k \in \mathbf{N}^x, x \geq 0$, $n, m \in \mathbf{N}$.

Proof. Suppose that both conditions hold, and let $\bar{T}$ be as specified in the proposition. Take player 1. We will show that if $\bar{T}_2 = T_2 \setminus \hat{t}$, for some $\hat{t}$, then there exists a TAE. The general case then follows immediately by induction.

Take an action $\hat{a} \in supp\left(s_2^*(\hat{t})\right)$. Since $\bar{T}$ satisfies Condition 1, there exists a type $\bar{t} \in \bar{T}$, s.t. $\hat{a} \in supp\,(s_2^*(\bar{t}))$. Let $\alpha(\hat{a} \mid \hat{t})$ denote the probability that s_2^* assigns to $\hat{a}$ given type $\hat{t}$. Let $\bar{P}(\bar{t}) = P(\bar{t}) + \alpha(\hat{a} \mid \hat{t})P(\hat{t})$, and allocate all the new mass of type $\bar{t}$ to the play of action a. That is, define the new mixed strategy of $\bar{t}$ as follows. Let

$$\bar{\alpha}(\hat{a} \mid \bar{t}) = \frac{\alpha(\hat{a} \mid \bar{t})P(\bar{t}) + \alpha(\hat{a} \mid \hat{t})P(\hat{t})}{P(\bar{t}) + \alpha(\hat{a} \mid \hat{t})P(\hat{t})},$$

and for every other action $\bar{a} \in supp\,(s_2^*(\bar{t}))$, let

$$\bar{\alpha}(\bar{a} \mid \bar{t}) = \frac{\alpha(\bar{a} \mid \bar{t})P(\bar{t})}{P(\bar{t}) + \alpha(\bar{a} \mid \hat{t})P(\hat{t})}.$$

By construction, the aggregate empirical distribution over actions of player 2 is unchanged by this transformation. By taking all the actions of type $\hat{t}$, and repeating this procedure, we allocate all the mass of $\hat{t}$ to other types of player 2, without affecting the empirical distribution over actions of player 2. By private values, player 1 only cares about the distribution over 2's actions. This proves sufficiency of conditions 1 and 2.

The converse follows from Proposition 3. □

We say that a TAE outcome *requires correct architectures* if there exists a unique TAE (s,c), where $s_n^{(k)} = s_n^*, (c_{n,T})_{(k,m)} = T_m, (c_{n,P})_{(k)} = P$, for all $k \in \mathbf{N}^x, x \geq 0$, $n, m \in \mathbf{N}$, where s^* is a Bayesian equilibrium sustaining the given outcome. We already saw that for the case of strict common values every TAE requires correct architectures. The next proposition provides a characterization of TAE that require correct architectures in private-value environments.

Proposition 6. Let $C_{n,P} = \{0,1\}^{\Delta(\mathbf{T})} \times \{0,1\}^{\Delta(\mathbf{T})}...$, and let s^* be a Bayesian equilibrium. Then the outcome of s^* requires correct architectures if and only if

$$supp\,(s_n^*(t_n)) \cap supp\,(s_n^*(\bar{t}_n)) = \emptyset, \forall t_n, \bar{t}_n \in T_n, t_n \neq \bar{t}_n. \qquad (9)$$

Proof. If condition (9) holds, then the claim follows: unless both players have correct architectures, TAE4 cannot be satisfied.

For the converse, if (9) doesn't hold, then we can construct a TAE which does not require correct architectures. In particular, a player's conjecture about the distribution of types need not be correct. Suppose therefore that (9) doesn't hold, so there exists two types $\bar{t}, \hat{t} \in T_2$, s.t. $supp(s_2^*(\bar{t})) \cap supp(s^*(\hat{t})) = \{\bar{a}\}$. We can assume wlog that the intersection of the supports is a single action, since the same construction can be made if the intersection is larger. Now, let player 1 conjecture the correct type space of player 2, but 1's conjecture about the probability distribution be given by

$$\bar{P}(\bar{t}) = P(\bar{t}) + \epsilon, \bar{P}(\hat{t}) = P(\hat{t}) - \epsilon,$$

where as before, $P(t)$ gives the true marginal distribution of type t, and ϵ is some positive number. Define 1's conjecture on 2's strategy, $s_2^{(1)}(\bar{t})$, by defining mixing probabilities $\bar{\alpha}(. \mid \bar{t})$ and $\bar{\alpha}(. \mid \hat{t})$. First let

$$\frac{P(\bar{t})}{\bar{P}(\bar{t})}\alpha(a \mid \bar{t}) = \bar{\alpha}(a \mid \bar{t}), \forall a \in supp(s_2^*(\bar{t})) \setminus \{\bar{a}\},$$

$$\frac{P(\hat{t})}{\bar{P}(\hat{t})}\alpha(a \mid \hat{t}) = \bar{\alpha}(a \mid \hat{t}), \forall a \in supp(s_2^*(\hat{t})) \setminus \{\bar{a}\}.$$

Since mixing probabilities have to sum to 1 for each type, we have

$$\bar{\alpha}(\bar{a} \mid \bar{t}) = 1 - \frac{P(\bar{t})}{\bar{P}(\bar{t})} \left(\sum_{a \in supp(s_2^*(\bar{t})) \setminus \{\bar{a}\}} \alpha(a \mid \bar{t}) \right), \text{ and}$$

$$\bar{\alpha}(\bar{a} \mid \hat{t}) = 1 - \frac{P(\hat{t})}{\bar{P}(\hat{t})} \left(\sum_{a \in supp(s_2^*(\hat{t})) \setminus \{\bar{a}\}} \alpha(a \mid \hat{t}) \right).$$

Now we make ϵ small enough so that $\bar{\alpha}(a \mid t) \in (0, 1)$, for all a and t. ¿From the last two equalities, we can easily verify that

$$P(\bar{t})\alpha(\bar{a} \mid \bar{t}) + P(\hat{t})\alpha(\bar{a} \mid \hat{t}) = \bar{P}(\bar{t})\bar{\alpha}(\bar{a} \mid \bar{t}) + \bar{P}(\hat{t})\bar{\alpha}(\bar{a} \mid \hat{t}).$$

Thus, the observed probability of playing each action is the same under P and α and under $\bar{P}$ and $\bar{\alpha}$. Clearly, the other requirements of TAE are also satisfied. □

References

[1] Battigalli P. and P. Guatoli[1988]: "Conjectural Equilibrium", mimeo.

[2] Dekel B., B. Lipman, and A. Rustichini[1998]: "Standard state-space models preclude unawareness", *Econometrica*, **66**, 159-173.

[3] Fagin R. and J. Y. Halpern[1988]: "Belief, awareness and limited reasoning", *Artificial Intelligence*, **34**, 39-76.

[4] Feinberg Y.[2004]: "Subjective reasoning-games with unawareness". Stanford Research Paper 1875.

[5] Feinberg Y.[2005]: "Games with Incomplete Awareness". Stanford Research Paper 1894.

[6] Fudenberg D. and D. K. Levine[1993]: "Self-Confirming Equilibrium", *Econometrica*, **61**, 523-545.

[7] Halpern J. Y.[2001]: "Alternative semantics for unawareness", *Games and Economic Behavior*, **37**, 321-339.

[8] Heifetz A., M. Meier, and B. Schipper[2005]: "Interactive Unawareness", Journal of Economic Theory, forthcoming.

[9] Jackson M. O. and E. Kalai[1998]: "Social Learning in Recurring Games", *Games and Economic Behavior*, **21**, 102-134.

[10] Li Jing[2006]: "Information structures with unawareness", mimeo. University of Pennsylvania.

[11] Martens, J.F. and S. Zamir[1985]: "Formulation of Bayesian analysis for games with incomplete information", *International Journal of Game Theory*, **14**.

[12] Modica S. and A Rustichini[1999]: "Unawareness and partitional information structures", *Games and Economic Behavior*, **27**, 265-298.

[13] Rubinstein A. and A. Wolinsky[1994]: "Rationalizable Conjectural Equilibrium Between Nash and Rationalizability", *Games and Economic Behavior*, **6**, 299-311.