Subjective Reasoning in Dynamic Games

Yossi Feinberg*†
Stanford University

Revised: May 2002

Abstract

A unified framework is developed for representation of, and reasoning about dynamic games. A game is described by the subjective knowledge of players at hypothetical situations – the epistemic game form. Subjective knowledge – termed confidence – allows us to replace objective reasoning about hypothetical events with reasoning about the confidence of hypothetical identities, i.e., the subjective reasoning of players in hypothetical situations. This leads to an endogenous definition for players' action sets. Applying subjective reasoning to the "Beer-Quiche" signaling game, the "Burning Money" game and others, provides a characterization of the dynamic reasoning by players that leads to the suggested solutions for these games. For perfect information games, rationality and common confidence of future rationality imply backward induction, although common confidence of rationality can logically contradict the definition of the game. **JEL Classification:** K9

* *Stanford Graduate School of Business, Stanford CA 94305-5015.* e-mail: yossi@gsb.stanford.edu

† This work has greatly benefited from extensive discussions and suggestions by Bob Wilson, Faruk Gul, and Chris Tyson. I would also like to thank Geir Asheim, Bob Aumann, Adam Brandenburger, Ehud Kalai, Piero La Mura, Marciano Siniscalchi, Yoav Shoham, Arnis Vilks and Muhamet Yildiz for helpful discussions.

1 Introduction

Modeling economic interactions as games has become a standard practice among economists. Inevitably, when we analyze the behavior of players in a given game we are forced to choose a solution concept from a plethora of possible alternatives. In many cases the analyst informally reasons about which solution "makes sense" in the given situation. Correspondingly, the technical definition of most solution concepts is accompanied by a verbal description of how the reasoning of the players will lead to the solution. There is, however, a gap between the mathematical definition of a solution, which essentially provides constraints on strategy profiles (sometimes with additional structure such as assessments – beliefs over information sets), and the intuitive and partial description of reasoning by the players about other players' actions, rationality and beliefs. This paper provides a single framework for both describing and analyzing dynamic games that bridges this gap. In this framework, the dynamic game form and the reasoning of players about behavior and rationality are expressed in the same language. This language allows us to make a precise connection between a solution concept and assumptions about the players' reasoning that are necessary or sufficient for the solution to obtain.

The cornerstone of our approach is the modelling of reasoning and knowledge in a dynamic game as seen by players at each and every decision point – their subjective view of the world. Putting forth this subjective representation of reasoning makes it easy to express hypothetical reasoning and clarifies what assumptions about the model and reasoning can be made available to the reasoning players. It helps us avoid the confusion between actual and hypothetical events when describing reasoning. This confusion, we find, is at the core of many of the difficulties in modeling reasoning in dynamic games and specifying which assumptions about reasoning can themselves be reasoned about by the modeled players.

In this paper we formulate the subjective framework and provide a number of applications of this representation of dynamic games and reasoning. Our first example is the centipede game. This game illustrates the basic principles of the subjective reasoning approach. We apply this approach and demonstrate how explicit subjective reasoning leads to equilibrium selection in our second application – the celebrated "Beer-Quiche" signaling game. We show how our intuition as modelers about the behavior of players can be translated to precise conditions on the reasoning of the players that imply this intuitive behavior. Our last example shows how the subjective framework captures delicate aspects of dynamic reasoning and behavior in a "burning money" game. This example demonstrates an additional aspect of the subjective approach. It allows us to express a player's reasoning about her own future actions. It captures the dynamics of reasoning by noting that once she burns money she has to reason about what she did and that she has to reason about *that* before she decides whether to burn money or not.

The subjective reasoning framework is suggested as a tool for expressing the reasoning of the players that leads to the behavior that we as modelers find to be "intuitive". It also allows us to study assumptions about rationality and reasoning by observing their implications on players' behavior. On one hand, we can rigorously analyze games such as the "Beer-Quiche" and the "Burning Money" games, or scrutinize the reasoning behind the divinity refinement in a signalling game (see Section 7.2). On the other hand, we can characterize the reasoning that corresponds to widely used solution concepts such as Nash and sequential equilibrium (see the companion paper [Feinberg 2001]).

The paper is structured as follows: Section 2 contains three applications of the subjective framework. Section 3 discusses the literature. In Section 4 the formal language of subjective reasoning is presented. In Section 5 we study the representation of a game via the subjective knowledge of players at the various decision points – the epistemic form. The characterization of the backward induction solution for perfect information games is given in Section 6. Section 7 contains the formal analysis of our main applications – the "Beer-Quiche" and "Burning Money" games. We also include an additional application for signalling games that relates to the divinity refinement.

2 Three Examples with Commentary

Both to illustrate what we will do in this paper and to show why we are doing it, consider the following three classic examples.

Consider the version of the centipede game depicted in Figure 1. The backward induction solution implies that Alice chooses to exit at her first decision point. An argument similar to the following is often invoked to justify this behavior: "If Alice expects that she will behave rationally at her second decision point and if she believes that Bob anticipates that she will, and if she anticipates that Bob will act rationally, then Alice must expect

Bob to exit and should therefore exit at her first decision point". But what if she stays? Bob must realize that under the above reasoning staying is not optimal for Alice, and hence that something must be wrong with respect to the assumptions made in this reasoning. Bob might then think that Alice will not act rationally at her second decision point as well, which would lead him to stay if he is rational. But then the rational thing for Alice to do at her first choice would indeed be to stay, since she would have been wrong to assume that if she stays then Bob will exit. We can go on to say that if staying is rational for Alice then Bob must know this, and again this seems to make staying the irrational choice for Alice – a paradox.

In this informal reasoning there are many notions that are not well defined. For one, the relationship between Alice's behavior (or rationality) at two different decision points is vague – we have not specified whether rationality is a property of a decision or of a decision maker. But even more importantly, while we have a collection of hypothetical statements; such as "what if Alice stays" – we are interested not only in these hypothetical statements, but also in the modeled agents' reasoning about the hypothesized situations, for example, "what Alice thinks that Bob will think if she stays". Moreover, Bob's reasoning has a specific context: We, and the agents, are interested in Bob's reasoning *after* Alice has stayed at her first decision point. Subjective reasoning is constructed on the premise that reasoning is done and is relevant *when and only when* a decision is to be made. Thus, in this example, we model three subjective view points: How Alice sees the game and reasons about behavior at her first decision point; how Bob views the situation when he (subjectively) knows that Alice has stayed and he is required to make a choice; and how Alice sees the game when she is at her second decision point. The subjective framework allows us to describe these three points of view rigorously, to analyze how in each one the decision maker can reason about behavior at other decision points, how they can reason about the reasoning made at another subjective view point, and so forth. Within this framework, intuitive verbal statements about reasoning can be made into precise logical statements.

We will use the term *confidence* for subjective knowledge. If we consider the statement or event "Bob is confident that the Alice stayed at her first decision point," the important point is that this event is actually part of the *definition* of the game. Bob at his hypothetical decision point, or, this hypothetical Bob, has no doubt that Alice has stayed. As noted above, this is the only point where Bob is required to make a choice, hence whatever rea-

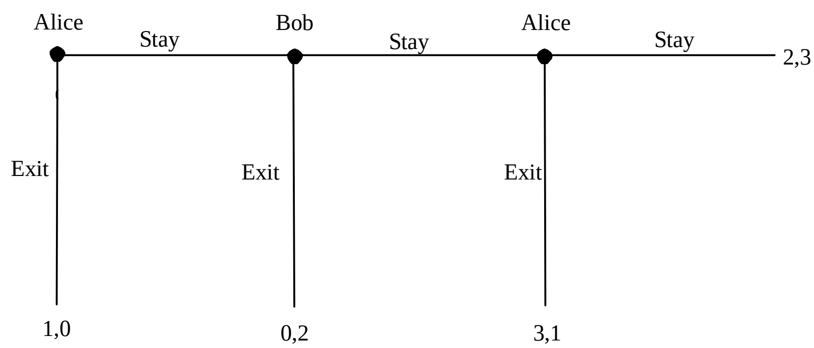

Figure 1: A centipede game

soning he does, or whatever reasoning is made about his reasoning, should assume that he knows that Alice stayed. Therefore, the event "Bob is confident that Alice stayed at her first decision point" is actually held in *common* confidence (everyone subjectively knows that everyone subjectively knows, and so on...). Here is where the subjectivity of knowledge plays a role. The fact that (the hypothetical) Bob acts upon the event "Alice stays at her first decision point" as if it is absolutely true does not imply that Alice (at her first decision point) agrees with Bob. In particular, Alice may choose not to stay. Alice can decide whether Bob – that Bob who is confident that she stayed – is hypothetical or actual.[1]

Using this formal expression of reasoning, we can now show that if Bob is rational and confident that Alice will be rational at the end, he will exit. If Alice at the beginning is rational and confident in the antecedent, she will exit. Since Bob is assumed to be confident she stayed, he could not be confident that Alice is rational and that she is confident that he is rational and that she is confident that he is confident that she will be rational at her second decision point – the combination of these would simply contradict the structure of the game.

This example demonstrates how describing the game using subjective knowledge combined with the description of reasoning about behavior can clarify this well known conundrum. The crux of this example is the use of the direct expression of reasoning at hypothetical decision points.

A more interesting relationship between behavior and reasoning is seen

[1]This hypothetical Bob will become "actual" if and only if he is called upon to choose.

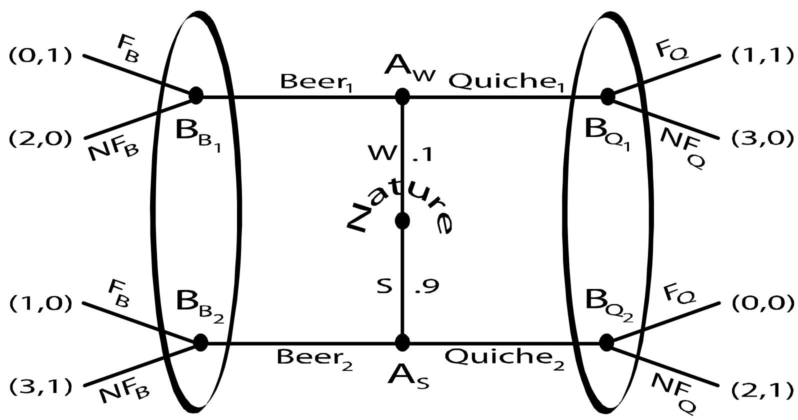

Figure 2: The "Beer – Quiche" game.

in our second example. Here subjective reasoning can be used to provide a justification for what we view as reasonable behavior in a given game. Consider the celebrated "Beer-Quiche" game depicted in Figure 2 from [Cho and Kreps 1987]. In this game, nature chooses either a weak Alice, A_W or a strong one, A_S, where the latter is much more likely to be chosen. Alice's type is her private information. Alice can then choose either *Beer* or *Quiche* and Bob observes her choice. Bob must then decide whether to fight Alice or not. Bob prefers to fight the weak Alice and not to fight the strong one. Alice prefers primarily not to be fought, but other things being equal A_w would rather have *Quiche* than *Beer* while A_S prefers *Beer* to *Quiche*. This game has two types of sequential equilibria, both are pooling equilibria – the two types of Alice make the same choices. In one type of equilibria, both types of Alice choose *Beer* and Bob fights, F, with a high probability when he observes *Quiche* and does not fight, NF, for sure when he observes *Beer*. The other kind of equilibria has both types of Alice choosing *Quiche* and Bob not fighting when observing *Quiche*, and fighting with high probability when observing *Beer*.

Intuitively, one might like to argue that if *Beer* is observed by Bob he would find it likely that the strong Alice chose *Beer*. Roughly speaking, if some belief would lead A_W to choose *Beer* the *same* belief would cause A_S to strictly prefer *Beer*. To get at this more intuitively appealing behavior, the intuitive criterion refines Nash equilibria roughly as follows: It requires that in a given equilibrium, the beliefs *off* the equilibrium path are such that types that conceivably could gain off the equilibrium path are deemed more likely compared to other types that would have nothing to gain by deviating. In the game above, the pooling equilibria with Alice choosing *Quiche* will fail the criterion since if Bob observes *Beer*, he will compare the equilibrium payoff for both types of Alice and conclude that only A_S could conceivably have thought that she will receive a higher payoff. This prompts him not to fight when observing *Beer* in contradiction with this equilibrium.

Much like in the centipede game, these verbal arguments raise many questions:

> How can Bob who observes *Beer*, denoted B_B, assume the equilibrium behavior in which he is not reached ?
>
> Why should he believe that Alice knows that she could get her equilibrium payoff if she chooses *Quiche* when he considers which type is likely to deviate, especially if the conclusion is that this equilibrium will not be played ?

There is a gap between *our* intuition that if *Beer* is chosen it is not reasonable that A_W alone would choose it and formal analysis of the players' reasoning and behavior.

The subjective framework bridges this gap. Consider the four subjective viewpoints in this game, A_W, A_S, B_B, B_Q , corresponding to the two types of Alice – weak and strong – and the two hypothetical Bobs at his decision points (having observed either *Beer* or *Quiche*)[2]. From the definition of the game we have that $C_{B_B}((W \text{ and } Beer_1)\text{or}(S \text{ and } Beer_2))$ holds[3], where the operator C_{B_B} stands for "B_B is confident of", i.e., Bob who observes *Beer* (B_B) is confident that either the weak (W) or the strong (S) Alice was realized by nature's move and whoever was realized chose *Beer* ($Beer_1$ or $Beer_2$ respectively). Consider the following statement: Both types of Alice have the same conjecture about Bob's behavior. This statement is not part

[2]We actually have six viewpoints, two for each observation of Bob. But the "two Bobs" that observe *Beer* (one as a choice by A_W and the other as a choice by A_S) cannot be distinguished from any subjective viewpoint, including their own. All possible viewpoints see them as identical and hence we identify them as one. The fact that there are two identical viewpoints will play a role when we claim that the game can be expressed by the subjective knowledge of the players at the various decision points.

[3]We actually have that this event is held in *common* confidence since we assume that the structure of the game is understood at every decision point.

of the definition of the game but is rather an assumption that will play a crucial role in deriving the intuitive behavior.

If both A_W and A_S have the same conjecture about what Bob will do, we can now state what would constitute rational behavior at the two decision points A_W and A_S. For every possible conjecture we consider the best responses summarized in the following table:

If B_B chooses:	F	F	NF	NF
and B_Q chooses:	F	NF	F	NF
then A_W's best response is:	Q	Q	B	Q
and A_S's best response is:	B	Q	B	B

We need to consider mixed conjectures as well, however the point is that whatever are the conjectures of A_W and A_S, as long as they are the same, we will never have A_W choosing *Beer* and A_S choosing *Quiche* as a best response. Within the language of subjective reasoning, shared conjectures and rationality of A_W and A_S *logically* imply that the event ($Beer_1$ and $Quiche_2$) does not hold[4]. If B_B is confident that both types of Alice have the same conjecture and is confident that they are rational we get that $C_{B_B}(\neg(Beer_1$ and $Quiche_2))$, i.e., B_B is confident that it is not true that both $Beer_1$ and $Quiche_2$ hold. If B_B is also rational we have the following best response table for B_B:

If A_W chooses:	Q	Q	B	B
and A_S chooses:	Q	B	Q	B
B_B's best response is:	–	NF	–	NF

Here B_B deems it impossible for both types to choose *Quiche* (the first column in the table above) by the definition of the game. This is part of B_B's subjective view of the world. When we add the condition that B_B finds both types of Alice to be rational, we have that B_B must find it impossible that only A_W would choose *Beer*[5]. Now, if $C_{A_S}C_{B_B}$(Alice's types have shared conjectures and are rational) and C_{A_S}(B_B is rational) then it is logically

[4]This event is not initially eliminated by the structure of the game from B_B's perspective. Bob might be confident that A_W is the only one choosing *Beer* and therefore assume that A_S is hypothetical which agrees with her choice of *Quiche* – a choice he is confident was not made.

[5]For B_B's mixed beliefs this will logically imply that he assigns no more than a probability of .1 to Alice's type being A_W.

implied that A_S is confident that B_B will not fight and similarly for A_W. We have that the best response for Alice can be confined given these assumptions to:

If B_B chooses:	F	F	NF	NF
and B_Q chooses:	F	NF	F	NF
A_W's best response is:	–	–	B	Q
A_S's best response is:	–	–	B	B

Hence, if A_S is also rational she will choose *Beer*. If we assume that B_Q is confident of all of the above then he will deduce *that* and will deduce that it must be A_W who chose *Quiche* (since he is confident that someone did); and if we assume his rationality as well, we know that he will choose F. If A_W is confident that B_Q is confident that A_S is confident that B_B will choose NF and that A_S is rational, then A_W will deduce that B_Q will chose F and hence she will choose *Beer* as well.

We have just concluded that common confidence that Alice's types have a shared conjecture, combined with R (rationality at all hypothetical decision points), CR (confidence of rationality *from* every subjective view point), CCR, $CCCR$ and $CCCCR$ imply the intuitive behavior on and off the actual play path[6]. Thus, once we demonstrate that these events and reasoning can be modeled explicitly, we will not only have the tools to express statements such as "When Bob observes *Beer* he finds it likely that A_S has been chosen by nature", but we will also be able to provide conditions on the decision makers' reasoning that imply such statements.

Our first example demonstrated how a dynamic game can be recast using subjective reasoning. The second example uses the subjective approach to obtain what a modeler may find to be intuitive behavior. This analysis of the "Beer-Quiche" game demonstrates how the subjective framework can assist in selecting a refinement by translating informal notions about reasoning into precise conditions about rationality. A second feature of this application is that it highlights two specific reasoning assumptions: maximal degrees of confidence in rationality and the joint conjectures assumption for types in a signalling game. In [Feinberg 2001] the first of these assumptions is developed into a solution concept.

[6]It is important to note that there cannot be confidence of all of the above from every subjective view point since B_Q would have to deduce that he is not reached, which contradicts his subjective knowledge.

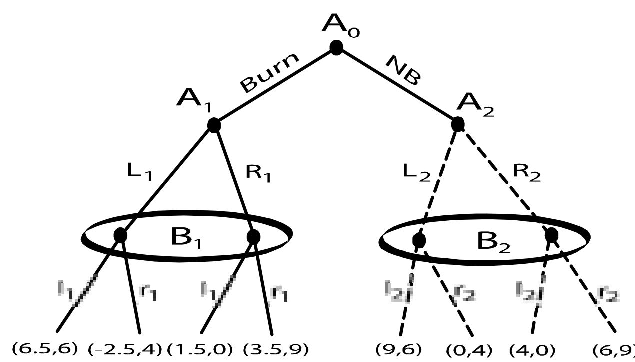

Figure 3: The "Burning Money" game.

But beyond the epistemic justification of behavior that we may find intuitive to begin with, our next application uncovers hidden aspects of reasoning that are paramount for deducing behavior. We show that our explicit formulation of reasoning reveals that a seemingly innocent refinement actually requires specific and demanding assumptions on reasoning. Here, the subjective approach helps us clarify why specific behavior is unintuitive.

Consider the "burning money" game depicted in Figure 3. Such games were studied by [Ben-Porath and Dekel 1992] who have shown that having the option of burning money could lead to a more favorable equilibrium payoff under iterated elimination of weakly dominated strategies. Here, if we consider the subgame denoted by the thick dashed line in Figure 3, we find multiple equilibria. In particular there is no immediate way to distinguish between (L_2, l_2) and (R_2, r_2). However, it has been argued that forward induction leads to the selection of the equilibrium (NB, L_2, l_2).

The argument proceeds as follows: Burning money and choosing R_1 is dominated by NB (not burning) and choosing R_2. Hence if Bob (B_1) observes Alice burning money, he will deduce that L_1 was chosen and will choose l_1 since r_1 becomes dominated. Given that, NB together with R_2 becomes dominated by $Burn$ with L_1. This, in turn, implies that B_2 will choose l_2 and now $Burn$ with L_1 is eliminated as well.

In this argument we reason about Alice's actions and reasoning while conditioning back and forth about whether she burns money or not. Such arguments do not clearly separate Alice's reasoning when she decides *whether* to burn money, from her reasoning about playing the coordination game

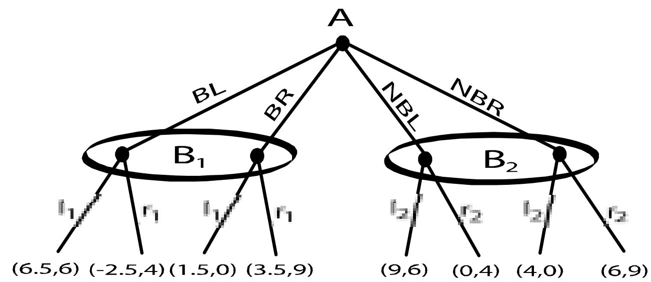

Figure 4: Burning money and acting simultaneously.

given her previous action. These deductions *do not* correspond to the exact description of the game in Figure 3 but rather fit more closely with reasoning in the game described by Figure 4, where the verbal reasoning above can be made into a precise logical argument[7]. If we are going to consider Bob's reaction to hypothetical situations we should notice that in Figure 3, Alice could face hypothetical situations as well. If we conclude that Alice would not burn money what would *she* think if she actually did? Would she just conclude that she acted irrationally, or might she form new conjectures about past and future behavior? Do we assume that a player always correctly anticipates their own actions? Can Alice change her mind, and if she doesn't, can we assume that Bob is confident of that? These questions become crucial if one wants to analyze the opportunity of burning money.

The subjective approach represents the reasoning from five view points in this game: A_0, A_1, A_2, B_1 and B_2. We first note that if Alice is allowed to change her mind after she alone takes an action then even common confidence of rationality does not narrow the possible outcomes of the game. For example, we might even have A_0 burning money and A_1, B_1 going right with the $(3.5, 9)$ payoff as the outcome of the game when a rational decision is made at every decision point and this fact is held in common confidence among all viewpoints. Being a little more specific, by changing her mind

[7]In fact, consider the game in Figure 4 where Alice *simultaneously* acts and decides whether to burn money or not. We find that the rationality of A implies $\neg BR$, hence CR implies that B_1 is confident that BL has been chosen. Hence together with the rationality of B_1 we have that l_1 will be chosen. CCR implies that A expects a payoff of 6.5 if she chooses BL. If B_2 is confident of that and of A's rationality we have that B_2 is confident in $\neg NBR$ which together with his rationality implies that he will choose l_2. We conclude that $R, CR, CCR, CCCR, CCCCR$ imply that (NBL, l_1, l_2) will be chosen.

we mean that A_1, A_2 can have a different conjecture about Bob's behavior than A_0 does and that A_0 can be wrong about what *she* will do at A_1 and A_2. It actually suffices that Bob will think that it is possible for Alice to change her mind, for the possibility of a failure of the coordination on the preferred equilibrium for Alice (without burning money), even with common confidence of rationality.

So what kind of assumptions about reasoning and rationality could lead to the best outcome for Alice? Are these assumptions reasonable? The answer is given with a reasoning procedure (made precise with the subjective framework). We find that is necessary for Alice to use *backward* reasoning about Bob's *forward* reasoning. Alice has to assume rationality about her own future self but *cannot* assume rationality about her past self and this must be accessible to Bob as he reasons forwards.

To see how subjective reasoning leads to this result, consider the following assumptions:

a. A_0, A_1 and A_2 share the same conjecture about Bob's behavior
b. A_0, A_1, A_2, B_1, B_2 are rational.
c. A_0 is confident of a and b.
d. B_1 is confident of a, b and c.
e. A_0 is confident of d.
f. B_2 is confident of a, b, c and e.
g. A_0 is confident of f.

Conditions (a) – (g) imply that Alice's preferred outcome will emerge. (b) together with (c) imply that A_0 will not choose B if her conjecture assigns a probability above $\frac{35}{198}$ to r_1. This follows from the fact that the *MaxMin* of the game following NB is $\frac{54}{11}$ hence whatever the conjecture of A_2 about B_2 is she expects at least that much since she is rational. But A_0 is confident that A_2 is rational and shares the same conjecture and so it follows that A_0 will only choose $Burn$ when she expects at least $\frac{54}{11}$ since A_0 is rational. Only when her conjecture assigns to r_1 a probability of at most $\frac{35}{198}$ could she expect such a payoff from choosing $Burn$.

Now (d) implies that the conclusion we just made can be made by B_1. In particular, since he is confident (from his subjective viewpoint) that A_0 chose $Burn$ he can *logically* deduce that her conjecture assigns no more than a probability of $\frac{35}{198}$ to him choosing r_1. Since he also finds A_1 to be rational and to share the same conjectures about his behavior as A_0, he will deduce that A_1 will choose L_1.

With (b) we deduce that B_1 will choose l_1 since he deduced that A_1

chooses L_1 (and he is confident that A_0 chose *Burn*). By (e) we have that A_0 will deduce that B_1 will choose L_1 and from (a) we have that A_1, A_2 will deduce the same.

Finally, (f) implies that B_2 is confident that A_0 expects 6.5 from choosing *Burn* (here B_2 needs to be confident that A_0 is confident that indeed A_1 is rational[8]). Since B_2 is confident that A_0 is rational and chose NB he is confident that she conjectures that he will choose l_2 with a probability of at least $\frac{13}{18}$ and by (a) so will A_2. B_2 expects L_2 to be chosen and (g) implies that it will.

A closer analysis of assumptions (c) – (g) is required (the need for (a) was noted above and (b) is naturally assumed). First, we note that no smaller number of reasoning steps (iterations of confidence of rationality) can lead to the same outcome. But, what is more important is the asymmetry of treatment of Alice's point of view. A_0 plays a special role. Yes, we can replace A_0 with A_1 or A_2 in (e) and (g), but then we would need to *add* confidence of A_1 (or A_2) in the rationality of B_1. However we cannot replace A_0 in (c) under any circumstances, since we cannot have A_1 be confident in A_0's rationality. Nor can we eliminate the condition that A_0 is confident of the rationality of A_1 and A_2.

We conclude that asymmetric assumptions are needed for the burning money game to generate the preferable outcome for Alice. Namely, Alice assumes rationality when reasoning about her own future reasoning *without* allowing Alice to assume rationality when reasoning about her own *past* self. It is the explicit reasoning at hypothetical points that uncovers this delicate yet substantial difference.

3 Literature

It is not a coincidence that each of these examples involves a *dynamic* game. Fortunately, normal form (static) games have been more amicable to the explicit description of reasoning and its relation to solution concepts. This is due to [Aumann and Brandenburger 1995] who provided conditions on the amount of knowledge of conjectures and of rationality that correspond to Nash equilibria. Previously, [Bernheim 1984] and [Pearce 1984] *defined* the

[8]Note that this is not implied by B_2's confidence in (d), since that only implies that A_0 is confident that B_1 is confident that A_0 is confident that A_1 is rational.

rationalizability solution concept as those strategy profiles that support common knowledge of rationality, and the seminal paper [Aumann 1987] demonstrated that common knowledge of rationality and a common prior characterize the correlated equilibria solution concept (see also [Brandenburger and Dekel 1987]).

For dynamic games most of the research done has focused on (generic) perfect information games. A wide range of formulations of reasoning has been suggested for the centipede type games. To name a few, the papers by [Aumann 1995], [Aumann 1998], [Reny 1992b], [Reny 1992a], [Stalnaker 1998], [Asheim Forthcoming], [Balkenborg and Winter 1997], [Clausing and Vilks 2000], [Clausing 2000], [Samet 1996] range from traditional common knowledge analysis to a framework where hypothetical inferences are explicitly modeled, through derivation of backward induction by consistency of preferences. See also [Brandenburger 2001] and the references therein for a more comprehensive list of results. Only a handful of results have been reported for more general dynamic games; most closely related are [Battigalli and Siniscalchi Forthcoming], [Brandenburger and Keisler 2002] and [Asheim and Perea 2000]. These contributions are based on expressing the beliefs of players using conditional probability systems (CPS) or lexicographic beliefs. Other than the obvious difference in formulation there are quite substantial differences between the CPS and the subjective framework. CPS capture hypothetical events as follows: Each player has an initial (ex-ante) belief about how the game is going to be played, beliefs that are updated according to the actual play. The crucial part of the definition is that if the play of the game surprises the player – an event with probability zero occurs – she will *revise* her beliefs according to her next order beliefs – beliefs conditional on that probability zero event. Hence, hypothetical events are reasoned about using explicit expression of conditionals on probability zero events[9]. In contrast, the subjective framework does not represent the reasoning from an ex-ante view point, beliefs are not constrained to be evolving or revised. Instead, beliefs are represented whenever there is a decision to be made based on the presumption that beliefs should only matter when a decision is made. In some sense only the ex-post beliefs are present. This difference explains why there is a more seamless transition to results pertaining to the normal

[9][Battigalli and Siniscalchi 1999] construct a universal type space for CPS, i.e., players can reason about the CPS used by other players, players can reason about other players reasoning about others' CPS, and so on.

(static) form of a game when CPS formulations are used, as is best seen in [Brandenburger and Keisler 2002]. On the other hand, it is more difficult to apply the revision of beliefs approach beyond perfect information games.

Another difference between the subjective framework and CPS is our explicit inclusion of the structure of the game into the scope of the decision makers' reasoning. Previous approaches to dynamic reasoning via the CPS have only used the structure of the game implicitly. However, an independent body of studies has considered the epistemic formulation of games separately from the formulation and analysis of reasoning in dynamic games. Among these are [Benthem 1999], [Bonanno 1992] and [Bonanno 1993]. In a recent exception [Clausing 2000] has provided a model that allows reasoning and structure of perfect information games to be described with the same language. This is done by introducing a conditional implication operator in addition to material implication as in the approach of [Samet 1996] to hypothetical events in perfect information games. In contrast, the subjective framework uses one standard implication operator, and one belief notion instead of both conditional and actual beliefs as in [Samet 1996]. Most importantly, the subjective representation of the game provides a considerable constraint on reasoning about rationality: Even for this class of perfect information games we show that common confidence of rationality *logically* contradicts the knowledge implied by the structure of the game, providing an explicit demonstration of the arguments suggested by [Reny 1992b].

While in [Battigalli and Siniscalchi Forthcoming] we find a characterization of forward induction, a companion paper [Battigalli and Siniscalchi 2000] characterizes the intuitive criterion using a notion of correct strong belief in rationality. This directly relates to the analysis of the "Beer-Quiche" game provided above. Indeed, their notion of a "best rationalization principle" has identical motivation as the "reasonable solution" concept defined and analyzed in the second part of this paper. Both try to find the extensive form counterpart to the notion of common knowledge of rationality in normal form games. One can say that [Battigalli and Siniscalchi Forthcoming] focus on the *common* or agreement on rationality, hence the *correct* part of correct strong belief, while our approach considers the *maximal* amount of rationality and confidence of rationality at all possible decision points in the game.

4 Subjective Epistemology

4.1 Syntax

We begin with a syntactic formulation of the language of subjective reasoning[10]. Consider a given finite alphabet $\{a, b, c, ...\}$ the elements of which are called *atomic formulas.* These formulas represent the fundamental statements – events – of interest, e.g. a statement such as "Alice chooses to exit at the first decision point in the centipede game". Next, we consider a set I called the *Identity Set.* This set of identities contains all possible manifestations of each player, i.e., an identity for each decision point (vertex) in the game tree. Each identity will posses it's own subjective view of the world. The pair $\{\{a, b, c, ...\}, I\}$ is called a *context.*

We first consider the following operators $\neg, \wedge, C_i, P_i^{\alpha}, u_i^r$ for all identities $i \in I$, rational α and real r. Finite applications of these operators generate all the statements (well formed formulas) in our language, i.e., all atomic formulas are statements (well formed formulas) and $\neg f$, $f \wedge g$, $C_i f$, $P_i^{\alpha} f$, $u_i^r f$ are statements for every staements f and g. The interpretations of such statements – $\neg f$, $f \wedge g$, $C_i f$, $P_i^{\alpha} f$, $u_i^r f$ – are "not f", "f and g", "i is confident of f", "i assigns to f a probability of at least α" and "i assigns a utility of at least r to f" respectively.

As usual, for a statement f in the language we define *mutual confidence of f to the degree m ($m > 0$)* as the formula $C^m = \bigwedge_{i \in I} C_i(C^{m-1} f)$ where $C_i^0 f = f$. The set $\{C^m f\}_{m=1,2,...}$ of formulas is called *common confidence of f.*

We now turn to the axiomatization of relationships among formulas. These are the rules that govern the logical deductions in our language. Here we have one major principle in mind: making sure that the views a subject has about the world are as close as possible to the views we have as modelers. Realizing the impossibility of perfect execution of this task, we emphasize the crucial points where our view of the world differs from the subject's views.

We begin with axioms of propositional calculus (see, for example, [Hughes and Cresswell 1996]):

[10] A semantic – state space – formulation can and will be provided. Although a state space approach is useful for some of our proofs, we find it easier to understand the properties of the subjective framework via syntax. Furthermore, the syntactic approach yields more transparency as to which epistemic conditions can be expressed in the syntax, and hence reasoned about by the players.

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 we define $f \vee g \equiv_{def} \neg(\neg f \wedge \neg g)$ and $f \Longrightarrow g \equiv_{def} \neg f \vee g$.

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

N Necessity: If f is a theorem then so is $C_i f$ for all i

To axiomatize the notion of confidence we use the following four axioms:

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

The interpretation of Axiom **K** is that our subjects are not only confident of theorems (as necessity implies); it is also a theorem that they can make deductions. Essentially they are assumed to be able to use the **MP** rule for statements that they are confident in.

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

Axiom **D** can be understood as requiring that no subject can be confident of both a proposition and its converse.

4 $C_i f \Longrightarrow C_i C_i f$

We add Axiom 4 because we will have all identities confident that their identity is confident of f if indeed it is[11][12].

[11]This does not mean that an identity knows who she actually is. For example, in the "Beer-Quiche" game we have B_{B_1} and B_{B_2}, the two identities that observe that Alice has chosen Beer. The first observes it after nature chose W and the second after nature chooses S. The crucial point is that these identities do not know what nature chose and hence do not know which exact identity they are when called upon to act. Axiom 4 states that whichever identity they are, if they are confident of a statement f they are confident that *that* identity is confident of f, in particular, if they cannot distinguish between the two identities they are confident that both identities are confident of f. In fact, we will use this observation to *define* an information set in a dynamic game.

[12]We also note that together with axiom 4, whatever we (as modelers) can logically deduce from i's confidence in a finite number of deductions,we can deduce that i is confident of the deduction it as well.

U $C_i(C_i f \Longrightarrow f)$

The axiom U states that i is confident that: whenever *she* is confident of an event it must hold. This is the crucial axiom that allows us to replace objective reasoning with a subjective framework: Instead of the classic objective knowledge axiom ($Kf \Longrightarrow f$) we have a subjective knowledge axiom. The difference is that if one identity is confident in f it does not imply that another identity cannot be confident in $\neg f$. In addition, each identity views herself as "being right". Hence, not only do we allow one identity to be confident in f and another to be confident in $\neg f$, we also allow the identities to be confident that this is indeed the case – confident that others might be confident that they are (subjectively) wrong. This is what allows for hypothetical contradicting viewpoints to coexist.

We now add beliefs and utilities to our framework. Starting with the knowledge-belief framework laid down in [Aumann 1999a] and [Aumann 1999b], we follow [Heifetz and Mongin 2001] in the axiomatization of the probability operators. Hence we assume the axiom scheme suggested by [Heifetz and Mongin 2001]. We note that an alternative approach via a richer algebraic syntax (see [Fagin, Halpern, Moses, and Vardy 1995]) is provided by [Fagin and Halpern 1994] and [Fagin, Halpern, and Megiddo 1990] as well as a more recent result by [Meier 2000].

We add the following two axioms about the interaction between belief and confidence.

$P_i^{\alpha}(f) \Longrightarrow C_i P_i^{\alpha}(f)$

$C_i f \Longrightarrow P_i^1(f)$

For our utility operators u_i^r the statement "i assigns a utility of at least r to f" should be understood in the context of von Neuman - Morgenstern utility functions. Given such a utility function i's expected utility conditioned on the event f is at least r. The following are the axioms regarding utilities and their relationship to beliefs and confidence.

$u_i^r(f) \Longrightarrow u_i^s(f)$ whenever $s < r$

$u_i^r(f \wedge g) \wedge u_i^s(f \wedge \neg g) \wedge P_i^{\alpha}(f \wedge g) \wedge P_i^{1-\alpha}(\neg(f \wedge g)) \wedge P_i^{\beta}(f \wedge \neg g) \wedge P_i^{1-\beta}(\neg(f \wedge \neg g)) \Longrightarrow u_i^{\alpha r + \beta s} f$ whenever $\alpha + \beta > 0$.

$\neg u_i^r(f \wedge g) \wedge \neg u_i^s(f \wedge \neg g) \wedge P_i^{\alpha}(f \wedge g) \wedge P_i^{1-\alpha}(\neg(f \wedge g)) \wedge P_i^{\beta}(f \wedge \neg g) \wedge P_i^{1-\beta}(\neg(f \wedge \neg g)) \Longrightarrow \neg u_i^{\alpha r + \beta s} f$ whenever $\alpha + \beta > 0$.

$C_i(f \wedge g) \Longrightarrow (u_i^r(f) \Longleftrightarrow u_i^r(g))$

$u_i^r(f) \Longleftrightarrow C_i u_i^r(f)$

We define a *(partial) description of the world* ω to be a set of statements that includes all the theorems and is closed under the system $KD4U$ together with the axioms for beliefs and utilities and the derivation rules, i.e., it includes all the statements that can be derived from the axioms and the statements in the set via finitely many applications of the derivation rules to statements in the set. We also require that this set of statements is *consistent*, i.e., $f \wedge \neg f \notin \omega$ (in particular no identity is confident in a contradiction $C_i(f \wedge \neg f) \notin \omega$). A description of the world. ω is called a *state of the world* if it is a *maximal* description of the world, i.e., $f \in \omega \iff \neg f \notin \omega$. Given a set of statements $\Im$, we define the *logical closure* of $\Im$ to be the (unique) minimal (set wise) partial description of the world that includes $\Im$, if no such description exists we have that $\Im$ must be logically inconsistent under our axiom scheme.

The following are the reasons for our use of a syntactic approach. First, in a syntactic framework it is easier to determine whether an event (or some properties in the model) can be expressed in the language, hence it is possible to determine whether the modeled players can reason about it[13]. There is no loss of information in the syntactic framework, i.e., one can always turn to the semantic – state space – framework. Finally, the syntactic formulation has an additional feature, it identifies a set of basic – atomic – events. More generally, events have a crude, yet intuitive, feature of complexity when expressed syntactically via the length of the statement, or the number of applications of operators, e.g. $C_iC_jC_if$ is a more elaborate statement than C_if. While we do not explicitly model the dynamics of deduction or measures of complexity, we can, and do, make use of this feature. We use atomic statements to define the notions of names of players and actions in a game, we also study conditions on reasoning about rationality by looking at adding higher and higher levels of mutual confidence of rationality and studying the implied behavior. This process of building reasoning about rationality from simple to higher levels of iterated reasoning is formalized in [Feinberg 2001] when introduction the reasonable solution concept.

[13]This responds to the concern that, in some cases, results in models for describing reasoning may depend on events that are not accessible to the modeled players. One such example is the common prior assumption, which together with common knowledge of rationality implies correlated equilibrium.

4.2 Names, Actions and Information Sets

We now turn to the definition of the notions of names of players and actions in our language.

Definition 1 *An atomic formula n (a letter in the alphabet) is defined as a* name *in a description of the world ω if there is an identity i such that $C_i n \in \omega$ and such that for all identities j either $C_j n \in \omega$ or $C_j \neg n \in \omega$ and there is common confidence of these statements, i.e., when $C_j n \in \omega$ then $CC(C_j n) \subset \omega$ and when $C_j \neg n \in \omega$ then $CC(C_j \neg n) \subset \omega$. If n is a name at ω we say that it is i's name for all i such that $C_i n \in \omega$.*

The interpretation of n that justifies this definition is, for example, taking n to represent the statement "My name is Alice", this statement would be true for any possible manifestation of Alice, true or hypothetical, in the game, and it will be false for any other player. Moreover, since we do not consider games with uncertainties as to *who* the players are, we expect common confidence in which identity finds "My name is Alice" to be (subjectively) true for her[14].

Assume a given alphabet and an objective set of identities I, i.e., a context.

Definition 2 *An atomic event s is called an* action *for the identity i at a description ω if $CC(C_i s \iff s) \subset \omega$, i.e., there is common confidence that s holds if and only if i is confident that it does.*

For example, if an elementary statement s is interpreted as "Alice exits at her first decision point", our definition states that from every subjective viewpoint there is confidence that this statement holds if and only if Alice herself (at that decision point) is confident that she exits, from every subjective viewpoint there is confidence of *that*, and so on.

[14]This definition of a name is quite strong and definitely could be relaxed in various ways yielding game forms that are more general than extensive form games. Furthermore, weaker definitions of "names" maybe suitable to express other notions of membership in economic environments. For example, common confidence among those who obtain the name of who has the name and who doesn't without common confidence among everyone could fit nicely with the following: For a given market assume a name is "I participate in the market", all participant may know exactly who is in the market, and this can even be held in common confidence, but assuming that those who *do not* participate in the market are actually confident of who does, and that *this* is held in common confidence may be undesirable.

This definition states that s has the property that everyone agrees that it can only hold (from each subjective viewpoint) if i is confident that it does. In other words it is i's confidence that *determines* the (subjective) truthfulness of s for all reasoning viewpoints. Stated differently, if i is confident of s then s is a fact (at least there is agreement on this implication although there might be disagreement over the facts). In addition, the other direction of the definition implies that there is common confidence that i is confident of her own actions, this is in-line with a widely adopted postulate stating that players know their own actions.

Here the use of an atomic statement plays a crucial role, for example, $CC(C_i f \iff f)$ is always part of the description of the world when f is a theorem (since descriptions include all theorems) and we would not like to view theorems as actions.

In preparation for our definition of the epistemic form of a game we define the following two properties of actions.

Definition 3 *An action s of an identity i is said to be* precise *at ω, if for every other action s' of i we have $CC(s \Longrightarrow \neg s') \in \omega$.*

Definition 4 *The set of all actions S_i for i is said to be* proper *at ω if $CC(\bigvee_{s_i \in S_i} s_i) \in \omega$.*

We note that since the set of atomic events is finite the set S_i is finite and $CC(\bigvee_{s_i \in S_i} s_i)$ is a countable collection of statements in our language[15]. When an identity i has a precise and proper set of actions we denote this set by $A(i)$.

We can now define the notion of rationality.

Definition 5 *We say that i is* rational *at ω if for every two actions a, b of i, and every r we have that the statement $C_i(a \wedge \neg b) \wedge \neg u_i^r a \Longrightarrow \neg u_i^r b$ is in ω. This can also be written as $\forall r$, $\forall a, b \in A(i)$ we have $(\neg u_i^r a \wedge u_i^r b \Longrightarrow \neg a) \in \omega$.*

[15]In this paper mixed strategies of an identity i are viewed as uncertainty of other players as to i's pure actions. This coincides with previous models of reasoning about behavior such as [Aumann and Brandenburger 1995]. If one wants to allow a player to actively randomize between her actions, we suggest that this randomization be part of the explicit description of the world, i.e., be defined as an action. Such a distinction usually matters only for games with imperfect recall.

As before this definition describes a collection of statements, and we say that rationality holds at a description of the world if it includes this set of statements. We will omit reference to the description of the world ω when not required. In particular, we have the following definition.

Definition 6 *We say that j is confident that i is rational if j is confident in each and every formula in the set that defines i's rationality.*

Note that the action sets of each identity are assumed to be of common confidence by the definition of actions.

We denote by $R_i = \{\neg u_i^r a \wedge u_i^r b \Longrightarrow \neg a \mid a, b \in A(i), \text{ for all } r\}$ the collection of statements that define i's rationality. Using the set of statements from the second part of the definition of rationality yields a logically equivalent set of statements.

Definition 7 *We say that i and j are in the same information set if there is common confidence that they have identical subjective views of the world, i.e., if $CC(C_i f \iff C_j f) \subset \omega$ for all f.*

We interpret an information set as a collection of decision points where the player has identical information. This condition *logically* implies that if s is an action of i and j is in the same information set as i then s is also an action of j. However, it is important to note that if i and j are in the same information set it need not imply that if $C_i f \in \omega$ then $C_j f \in \omega$, but rather that no identity (including i and j) considers this to be possible.

5 The Epistemic Form of a Game

We now turn to the representation of a dynamic game via the subjective knowledge of hypothetical decision makers. A game is thought of as a collection of players, the decisions they may make, their knowledge and beliefs when making a decision, acts of nature and outcomes as well as the assumptions as to how these concepts relate to each other. The crux of the representation is that although the actual existence of a decision opportunity might depend on a specific previous action, we still want to capture hypothetical as well as actual decisions in our model. This is achieved by transferring the dependencies between decisions to knowledge of the decision makers, while preserving the dependencies between actions and outcomes.

We begin with identifying the properties that a description of the world should have if it is supposed to represent a dynamic game. We then use the centipede game as an example of how an epistemic form is generated. We continue with a general procedure of translating an extensive form game with perfect recall into an epistemic form. This constructive procedure is shown to satisfy the initial requirements on the representations that we stated. Most prominently, it is shown that the construction generates a logically consistent description of the world. We discuss to what extent the epistemic form is unique and its limitations with respect to imperfect recall.

When we construct the epistemic form of the game we would like the following properties to hold:

1. The language for the epistemic form of a dynamic game is given by letting the set of identities be the set of decision points in the game, and letting the set of atomic statements be the set of actions and names of players in the game.

2. The epistemic form of a given game is a set of statements Γ in this language.

3. An epistemic form Γ is a partial description of the game, i.e., it is logically consistent and closed under our axiom scheme.

4. The epistemic form is *epistemically* closed, i.e., for every identity $i \in I$ we have that $f \in \Gamma$ implies $C_i f \in \Gamma$.

5. The epistemic form Γ includes the description of names, actions and information sets as defined in the previous section. In particular, the set of actions is precise and proper.

6. The epistemic structure includes the description of the players' expected utility conditioned on any pure strategy profile.

7. The epistemic form includes a description of the dynamics of the game by explicitly stating what each identity subjectively knows about the play of the game leading to the corresponding decision point.

In addition to these properties of the epistemic form of a specific game, we would like to explore the following additional properties with respect to any formal procedure that recasts extensive form games via an epistemic form:

8. There is a procedure which translates every extensive form game to an epistemic form with the properties listed above.

9. The procedure generates a unique epistemic form for each extensive form game, i.e., the procedure in 8. is injective.

The procedure suggested below, that defines the epistemic form for an extensive form game, satisfies properties 1–7. It satisfies property 8 for games with perfect recall where at any decision point there are at least two actions to choose from. Finally, it satisfies property 9 up to a class of games with an interchangeable dynamic structure that preserves the subjective information in all decision points.

As an example, we consider the game in Figure 1 – the centipede game. We would like to construct the language and collection of statements that will represent this game in accordance with properties 1. – 7. above.

This game has three decision points. Hence, we define the set of identities to be $I = \{A_1, A_2, B\}$. There are six actions in the game and two players (two player names), our set of atomic statements contains eight elements denoted by $\{s_1, e_1, S, E, s_2, e_2, Alice, Bob\}$. Our interpretation of these statements is: $s_1 =$"Alice stays at her first decision point", $e_1 =$"Alice exits at her first decision point", $S =$"Bob stays at his decision point", $E =$"Bob exits at his decision point", $s_2 =$"Alice stays at her second decision point", $e_2 =$"Alice exits at her second decision point", $Alice =$"My name is Alice" and $Bob =$"My name is Bob". Note that these interpretations will be supported with the epistemic form of the game in accordance with our definitions of actions and names.

We define the epistemic form of the game be the logical and epistemic closure of the list of statements below. The language we used is also determined by the game since the set of identities is the set of decision points and the atomic statements are the names and actions. The logical closure is defined under the axiom scheme from Section 4.

Names:

$C_{A_1}Alice$, $C_{A_1}\neg Bob$, $C_{A_2}Alice$, $C_{A_2}\neg Bob$, $C_B Bob$, $C_B\neg Alice$

Actions:

$C_{A_1}s_1 \iff s_1$, $C_{A_1}e_1 \iff e_1$, $C_B S \iff S$, $C_B E \iff E$, $C_{A_2}s_2 \iff s_2$, $C_{A_2}e_2 \iff e_2$, $s_1 \iff \neg e_1$, $S \iff \neg E$, $s_2 \iff \neg e_2$

Dynamics:

$C_B s_1$, $C_{A_2}(s_1 \wedge S))$

Utilities:
$U^2_{A_1}(s_1 \wedge S \wedge s_2)$, $U^2_{A_2}(s_1 \wedge S \wedge s_2)$, $U^3_B(s_1 \wedge S \wedge s_2)$
$U^3_{A_1}(s_1 \wedge S \wedge e_2)$, $U^3_{A_2}(s_1 \wedge S \wedge e_2)$, $U^1_B(s_1 \wedge S \wedge e_2))$
$U^0_{A_1}(s_1 \wedge E \wedge s_2)$, $U^0_{A_2}(s_1 \wedge E \wedge s_2)$, $U^2_B(s_1 \wedge E \wedge s_2))$
$U^0_{A_1}(s_1 \wedge E \wedge e_2)$, $U^0_{A_2}(s_1 \wedge E \wedge e_2)$, $U^2_B(s_1 \wedge E \wedge e_2))$
$U^1_{A_1}(e_1 \wedge S \wedge s_2)$, $U^1_{A_2}(e_1 \wedge S \wedge s_2)$, $U^0_B(e_1 \wedge S \wedge s_2))$
$U^1_{A_1}(e_1 \wedge S \wedge e_2)$, $U^1_{A_2}(e_1 \wedge S \wedge e_2)$, $U^0_B(e_1 \wedge S \wedge e_2))$
$U^1_{A_1}(e_1 \wedge E \wedge s_2)$, $U^1_{A_2}(e_1 \wedge E \wedge s_2)$, $U^0_B(e_1 \wedge E \wedge s_2))$
$U^1_{A_1}(e_1 \wedge E \wedge e_2)$, $U^1_{A_2}(e_1 \wedge E \wedge e_2)$, $U^0_B(e_1 \wedge E \wedge e_2))$

Here and throughout the paper we use the definition $U^r_i f = \{u^r_i f\} \cup \{\neg u^{r+\varepsilon}_i f \mid \forall \varepsilon > 0\}$ for the set of statements that imply that i's expected utility given f is exactly r.

Recasting the centipede game as a set of statements was done in two parts. First, the language was determined by the number of decision points along the game tree (the set of identities) and the union of the set of actions and set of player names (the set of atomic statements). The second stage was constructing a particular partial description of the world such that each atomic statement is assigned as an identity action or a name. Finally, given the allocation of names and actions, utilities are assigned to outcomes and most importantly identities are assumed to be confident in the actions that lead to their decision point – the dynamics of the game are modeled. By requiring that the description be epistemically closed, we explicitly require that this description of the game be held in common confidence, i.e., there is common subjective knowledge of the game form.

5.1 Existence and Construction of the Epistemic Form

While it is easy to generalize the construction of the epistemic game form as it was demonstrated for the centipede game, there was an important assumption made in this example. We assumed that the logical and epistemic closure of the list of statements was a description of the game, i.e., that it was *consistent*. Otherwise, there would be no use in adding conditions about reasoning to the epistemic form since a contradiction logically implies *all* statements. In this section we present the general mapping from an extensive form game to an epistemic form – a language and a set of statements. The main result in this section shows that for games with perfect recall this construction always generates a (consistent) description of the world.

Consider a finite extensive form game $\Gamma = (N, H, P, \{G_i\}_{i \in N}, \{\tilde{u}_i\}_{i \in N})$ where N is a finite set of players, H is a finite set of finite sequences closed under elimination of the last member of a sequence – a set of histories – including the empty sequence, P is a surjective function from $H \setminus Z$ to N with Z being all maximal sequences in H, G_i is a partition of $P^{-1}(i)$ that satisfies $A(h) = A(h')$ with $A(h) = \{a | (h, a) \in H\}$ 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, i.e., the same partition member as h in $G_{P(h)}$. We denote by $s(h)$ the unique set of actions that comprises the history h. This representation follows the definition of extensive form games as appears in [Osborne and Rubinstein 1994] and is equivalent to the definition by Kuhn. Our first objective is to define the epistemic form corresponding to an extensive form game. However we will limit the description to a class of extensive form games that includes all games with perfect recall. We restrict our attention to extensive form games for which in every sequence $h \in H$ a choice a appears at most once, i.e., the elements of every sequence are distinct. This corresponds to the assumption that a player recalls *whether* she made a choice or not, although it allows the situation where the player forgets *which* choice she actually made. We call these *games with agent perfect recall.* For simplicity we first consider only games without nature moves.

For a given extensive form game $\Gamma = (N, H, P, \{G_i\}_{i \in N}, \{\tilde{u}_i\}_{i \in N})$ we define a context (the set of identities and atomic statements) $\{\mathcal{I}, \alpha\}$ by $\mathcal{I} = H \setminus Z$ and $\alpha = N \cup \bigcup_{h \in H \setminus Z} A(h)$. Given this context the *epistemic form* of the game Γ is the set of statements $\mathcal{L}$ that is the epistemic and logical closure of the following list of statements in the given language:

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

$$C_h a \Longleftrightarrow a \qquad \forall h \in H \setminus Z, \forall a \in A(h) \qquad \text{(h's actions)} \tag{2}$$

$$a \Longrightarrow \neg a' \qquad \forall h \in H \setminus Z, \forall a, a' \in A(h) \qquad \text{(actions are precise)} \tag{3}$$

$$\bigvee_{a \in A(h)} a \qquad \forall h \in H \setminus Z \qquad \text{(action sets are proper)} \tag{4}$$

$$C_{\bar{h}}(C_h f \Longleftrightarrow C_{h'} f) \qquad \forall \bar{h}, h \in H \setminus Z, \forall h' \in G(h) \qquad \text{(information sets)} \quad (5)$$

$$C_h(\bigvee_{h' \in G(h)} s(h')) \qquad \forall h \in H \setminus Z \qquad \text{(dynamic knowledge structure)} \quad (6)$$

$$U_h^r(s(h')) \qquad \forall h \in H \setminus Z, \tilde{u}_{P^{-1}(h)}(h') = r, \forall h' \in Z \qquad \text{(utilities)} \quad (7)$$

The description of the game – the epistemic form – is the list of formulas that can be derived from the collection of formulas above by logical deduction and by the identities' confidence operators. Beyond the attempt to capture notions such as the subjective knowledge at a decision point, confidence of actions and so on, the main question is whether this construction generates a *consistent* set of statements. Otherwise, anything can be logically derived from an inconsistent set of statements. In this sense the epistemic form is meaningful – or should we say it exists – only if it is a partial description of the game, i.e., does not lead to a logical contradiction.

Theorem 8 *The epistemic form of an extensive form game with agent perfect recall is logically consistent. In particular the epistemic form constitutes a partial description of the world.*

Proof. We need to show that the logical and epistemic closure $\mathcal{L}$ of the collection of statements above is indeed consistent. We show this by constructing a specific model for the game. A model is a set of states, a collection of relations that correspond to the operators in our language and a valuation function that assigns for every state and every statement in our language the value 1 or 0. The interpretation is that for a given statement in our language the set of states where the valuation is 1 corresponds to the event where the statement holds. To clarify how this proves consistency we first restrict the discussion to the language and axioms for the confidence operator alone, i.e., we temporarily discard probability assessments and utilities from our language. In this case, a model $\langle \Omega, (R_h)_{h \in H \setminus Z}, V \rangle$ is a set of possible worlds Ω , accessibility relations R_h – one relation for every confidence operator (dyadic relations where $\omega R_h \omega'$ is understood as: h finds ω' possible at the state ω), and a valuation function V that assigns 1 (true) or 0 (false) to every formula at every state of the world. It is required for a model that the valuation V satisfy the following rules:

$$V(\omega, \neg f) = 1 \iff V(\omega, f) = 0 \tag{8}$$

$$V(\omega, f \wedge g) = 1 \iff \{V(\omega, f) = 1 \text{ and } V(\omega, g) = 1\} \tag{9}$$

$$V(\omega, C_h f) = 1 \iff \{V(\omega', f) = 1 \qquad \forall \omega' \text{ such that } \omega R_h \omega'\} \tag{10}$$

The model $\langle \Omega, (R_h)_{h \in H \setminus Z}, V \rangle$ will demonstrate the consistency of a list of statements $\mathcal{L}$ (1–6 since we defer the discussion of utilities and probabilities) if we can show the following:

1. There exists a model with a state $\omega \in \Omega$ such that $V(\omega, f) = 1$ for every formula f in the list $\mathcal{L}$ and

2. Our axiom scheme is sound with respect to this model, i.e., that every theorem in our language is valid (has a valuation of 1) at every state in the model.

This follows from the fact that a list $\mathcal{L}$ is inconsistent only if there exists a finite list of formulae $f_1, ..., f_n \in \mathcal{L}$ such that $\neg(f_1 \wedge ... \wedge f_n)$ is a theorem in our language and in particular $V(\omega, \neg(f_1 \wedge ... \wedge f_n)) = 1$ for all $\omega \in \Omega$ in our model according to 2. which contradicts having 1. hold together with the properties 8–9 of V.

Still confining our language and the list $\mathcal{L}$ to statements that do not invoke probabilities or utilities, we note that the axiom system $KDU4$ is sound and complete[16] with respect to the class of standard models $\langle \Omega, (R_h)_{h \in I}, V \rangle$ that satisfy the following three conditions:

$$\forall h \forall \omega \exists \omega' \text{ we have } \omega R_h \omega' \qquad (D \text{ holds}) \tag{11}$$

$$R_h \text{ is transitive for all } h \qquad (4 \text{ holds}) \tag{12}$$

$$\forall h \forall \omega, \omega' \text{ we have that } \omega R_h \omega' \text{ implies } \omega' R_h \omega' \qquad (U \text{ holds}) \tag{13}$$

[16]Recall that completeness of an axiom scheme w.r.t. a class of models is that all valid statements (statements that hold at every state at every model in the class) are theorems of the language.

This follows from standard proofs of soundness and completeness and can be found in [Chellas 1980] (Section 5.4). Note that the truth axiom corresponds to the condition $\forall h \forall \omega' \ \omega' R_h \omega'$ (reflexivity) so confidence of the truth axiom implies that every identity finds only reflexive states to be possible (from her subjective view point) yielding (13)[17].

Since the axiom scheme U has not been extensively studied, we briefly show why 13 is sufficient for the soundness of this class of models. Consider a model that satisfies 13 (in addition to 8 – 10). We need to show that $V(\omega, C_h(C_h f \implies f)) = 1$ for every state ω in the model, every h and every statement f. Given ω and f, we need to show $V(\omega, C_h(\neg(C_h f \wedge \neg f))) = 1$, by 10 we need to show that $V(\omega', \neg(C_h f \wedge \neg f)) = 1 \quad \forall \omega'$ such that $\omega R_h \omega'$. By 13 it suffices to show that $V(\omega', \neg(C_h f \wedge \neg f)) = 1$ holds for all ω' such that $\omega' R_h \omega'$. By 8 we need to show $V(\omega', C_h f \wedge \neg f) = 0$ holds for all ω' such that $\omega' R_h \omega'$, and by 9 it suffices to show that either $V(\omega', C_h f) = 0$ or $V(\omega', \neg f) = 0$ for ω' such that $\omega' R_h \omega'$. Assume that $V(\omega', \neg f) = 1$, i.e., $V(\omega', f) = 0$, then by 10 since $\omega' R_h \omega'$ we have that $V(\omega', C_h f) = 0$ and hence standard models that satisfy 13 are sound w.r.t. the axiom scheme U.

We now construct a model for the epistemic form of the game with respect to properties 1 – 6 using the language without probabilities and utilities.

Let $\Omega = N \times \prod_{h \in H \setminus Z} A(h)$, i.e., the state space is a product space where the first coordinate is the name and all other coordinates correspond to the action made by each identity. Note that we abuse the notation n and a by using the same notation in three different contexts, e.g. n is the atomic statement that will correspond to a player's name, it is also the player's name in the extensive form of the game and it is also a value of the first coordinate of a state. We use this notation because in all three interpretations n stands for a player's name and a stands for an action. They are naturally mapped to each other throughout and the interpretation should be obvious from the context. Let ω_h denote the h coordinate of ω and ω_0 denote ω's name coordinate. We define the accessibility relations R_h $(h \in H \setminus Z)$ for $\omega, \omega' \in \Omega$ denoted $\omega R_h \omega'$ whenever the following conditions hold

$$\omega'_0 = P(h) \tag{14}$$

$$\forall j \forall i \in G(j) \text{ we have } \omega'_i = \omega'_j \tag{15}$$

[17]In particular $KD4U$ is *strictly* between $KD4$ and $KD45$.

$$\forall \bar{h} \in G(h) \text{ we have } \omega'_{\bar{h}} = \omega_h \tag{16}$$

$$\exists \bar{h} \in G(h) \text{ such that whenever } (g,a) \text{ is a sub-history of } \bar{h} \text{ then } \omega'_g = a \tag{17}$$

In words, we say that ω' is accessible from ω according to R_h iff the statement "my name is $P(h)$" holds at ω', the same choice is made by all members of an information set and for the information set $G(h)$ the choice made by all members is the one made in ω by h, and h is reached in ω' (one of the histories that leads to the information set $G(h)$ occurs according to ω'.

Finally, we define V for the atomic formulae as follows:
for names we let[18]

$$V(\omega, n) = 1 \iff \omega_0 = n \tag{18}$$

and for actions we have

$$V(\omega, a) = 1 \iff \{\omega_h = a \qquad \forall h \text{ such that } a \in A(h)\} \tag{19}$$

Once defined for atomic formulae V uniquely extend to all formulae according to 8 – 10.

We first need to check that $\langle \Omega, (R_h)_{h \in H \setminus Z}, V \rangle$ is indeed a model for $KDU4$, i.e., that 11 – 13 are satisfied. Given h and ω 11 states that there exists a state ω' that h finds possible at ω. Consider the following state $\omega'_0 = P(h)$, $\forall \bar{h} \in G(h)$ let $\omega'_{\bar{h}} = \omega_h$,for each sub-history (g,a) of h let $\omega'_{g'} = a$ for every $g' \in G(g)$, and for every other h' such that $\omega'_{h'}$ has not been determined yet choose an arbitrary value in $A(h')$ for all $\bar{h} \in G(h)$. By our assumption that players remember whether they made a choice or not we have that no information set intersects the history h more than once and so the state ω' is well defined (the definition for sub-histories does not lead to a contradiction). To check for transitivity consider ω, ω', ω" such that $\omega R_h \omega'$ and $\omega' R_h \omega$". Conditions 14, 15, 17 are satisfied by ω" since it is accessible

[18]Here n on the left hand side stands for the atomic statement and on the right hand side for value of the first coordinate of a state.

from ω'. Since $h \in G(h)$ we have that by $\omega R_h \omega'$ and condition 16 $\omega'_h = \omega_h$ and so by $\omega' R_h \omega$" and condition 16 we have that $\omega"_{\bar{h}} = \omega_h$ for every $\bar{h} \in G(h)$ and so $\omega R_h \omega$" as required. Finally, when $\omega R_h \omega'$ then ω' satisfies conditions 14, 15, 17 but condition 15 implies 16 when replacing ω with ω' and we have $\omega' R_h \omega'$ as required.

We now need to verify that 1 – 6 are satisfied at every ω.

Since for every ω, h we have that $\omega R_h \omega'$ implies that $\omega'_0 = P(h)$ by 18 and 10 we have that $V(\omega, C_h n) = 1 \iff n = P(h)$.

For proving 2 we need to show that if $\omega R_{h'} \omega'$ for some ω, ω', h' then for every h we have that $\omega'_h = a \iff (\omega' R_h \tilde{\omega} \Longrightarrow \tilde{\omega}_h = a)$, this follows from 19 and the definition of R.

Both 3 and 4 follow from 19. To show that 5 holds we need to show that if h, h' are in the same information set and $\omega R_{\tilde{h}} \omega'$ for some ω and $\tilde{h}$ then $\omega' R_h \bar{\omega} \iff \omega' R_{h'} \bar{\omega}$ for all $\bar{\omega}$, according to 10. But the last condition follows from the definition of the relations R since at ω' the h and the h' coordinates are equal and in that case the same states are accessible from ω' for both h and for h'.

For 6 we need to show that for every h and every ω and ω' such that $\omega R_h \omega'$ there is an $h' \in G(h)$ such that $\omega'_g = a$ for every sub-history (g, a) of h'. However, this follows from 17.

Finally, we need to consider our full language with probability and utility operators and all the axioms. We need to find a model where the list $\mathcal{L}$ representing the epistemic form including 7 holds at a state of the world, where the model is sound with respect to the axiom scheme including the axioms for probability and utility operators. Given a context $\{\{a, b, c, ...\}, I\}$ we consider models $\langle \Omega, (R_h)_{h \in I}, (T_h)_{h \in I}, \left(\tilde{U}_h\right)_{h \in I}, V \rangle$ where Ω is a finite set, R_h are relations, $T_h : \Omega \to \Delta(\Omega)$ assigns a probability distribution for every identity h at every state ω, $\tilde{U}_h : \Omega \to \mathcal{R}$ are real valued functions and $V : \Omega \times \mathcal{F} \to \{0, 1\}$ is a valuation function assigning a truth value for every state ω and every statement $f \in \mathcal{F}$ in our language. We require that these objects satisfy the following conditions:

$$V(\omega,\neg f)=1 \iff V(\omega,f)=0 \tag{20}$$
$$V(\omega,f\wedge g)=1 \iff \{V(\omega,f)=1 \text{ and } V(\omega,g)=1\} \tag{21}$$
$$V(\omega,C_h f)=1 \iff \{V(\omega',f)=1 \quad \forall\omega' \text{ such that } \omega R_h \omega'\} \tag{22}$$
$$V(\omega,P_h^\alpha f)=1 \iff T_h(\omega)(\{\omega'\in\Omega|V(\omega',f)=1\})\geq\alpha \tag{23}$$
$$V(\omega,u_h^r f)=1 \iff E_{T_h(\omega)}(\tilde{U}_h|\{\omega'\in\Omega|V(\omega',f)=1\})\geq r \tag{24}$$
$$\forall h\forall\omega\exists\omega' \text{ such that } \omega R_h\omega' \tag{25}$$
$$R_h \text{ are transitive} \tag{26}$$
$$\omega R_h\omega' \text{ implies } \omega' R_h \omega' \tag{27}$$
$$\omega R_h\omega' \text{ implies } T_h(\omega)(\cdot)=T_h(\omega')(\cdot) \tag{28}$$
$$T_h(\omega)(\{\omega'\})>0 \text{ implies } \omega R_h\omega' \tag{29}$$

We need to show that our axiom scheme is sound w.r.t. the collection of such models, i.e., that if f is a theorem in our language then for all such models for every ω the valuation of f in ω is 1. With 20 – 22 and 25 – 27 we have the derivation rules and $KD4U$, since we follow [Heifetz and Mongin 2001] for the axiom scheme for probability assessments we have that this class is sound w.r.t. to the probability axioms since by 23 $\langle\Omega,(T_h)_{h\in I},V\rangle$ corresponds to a finite type space. Recall that [Heifetz and Mongin 2001] have shown that their axiom scheme is sound and complete w.r.t. the class of general (not necessarily finite) type spaces[19]. Property 28 assures that derivations based on the axiom $P_i^\alpha(f) \implies C_i P_i^\alpha(f)$ are valid in all such models since if $V(\omega,P_h^\alpha(f))=1$ and $\omega R_h\omega'$ then by 23 we have $T_h(\omega)(\{\bar\omega\in\Omega|V(\bar\omega,f)=1\})\geq\alpha$ and by 28 $T_h(\omega')(\{\bar\omega\in\Omega|V(\bar\omega,f)=1\})\geq\alpha$ finally by 23 we get $V(\omega',P_h^\alpha f)=1$ for all such ω' hence $V(\omega,C_h P_h^\alpha(f))=1$ as required. Similarly, $C_i f \implies P_i^1(f)$ is derived from 29 which states that the support of $T_h(\omega)(\cdot)$ is included in the set $\{\omega'\in\Omega|\omega R_h\omega'\}$. The requirement 24 implies that the expected utility axioms $u_i^r(f\wedge g)\wedge u_i^s(f\wedge\neg g)\wedge P_i^\alpha(f\wedge g)\wedge P_i^{1-\alpha}(\neg(f\wedge g))\wedge P_i^\beta(f\wedge\neg g)\wedge P_i^{1-\beta}(\neg(f\wedge\neg g)) \implies u_i^{\alpha r+\beta s}f$ whenever $\alpha+\beta>0$, $\neg u_i^r(f\wedge g)\wedge\neg u_i^s(f\wedge\neg g)\wedge P_i^\alpha(f\wedge g)\wedge P_i^{1-\alpha}(\neg(f\wedge g))\wedge P_i^\beta(f\wedge\neg g)\wedge P_i^{1-\beta}(\neg(f\wedge\neg g)) \implies \neg u_i^{\alpha r+\beta s}f$ whenever $\alpha+\beta>0$ and $C_i(f\wedge g)\implies(u_i^r(f)\iff u_i^r(g))$

[19] We could have considered general (infinite) type spaces as well. In that case additional requirements such as measurability for beliefs and boundedness for the utility function are needed. Since we only use models for proofs of consistency, we forgo this generalization for the sake of simplicity.

are satisfied. Note, that in 24 the expression $E_{T_h(\omega)}(\tilde{U}_h|\{\omega' \in \Omega|V(\omega', f) = 1\})$ is well defined when $T_h(\omega)(\{\omega' \in \Omega|V(\omega', f) = 1\}) > 0$, i.e., when we condition on an event such that h assigns it a positive probability at ω. However, we have not defined it for conditioning on probability zero events[20], here we assume an arbitrary choice of value for $E_{T_h(\omega)}(\tilde{U}_h|\{\omega' \in \Omega|V(\omega', f) = 1\}$ in such a case. Finally, by 28 and 24 we have that $V(\omega, u_h^r f) = 1$ leads to $V(\omega', u_h^r f) = 1$ whenever $\omega R_h \omega'$ hence $u_i^r(f) \Longleftrightarrow C_i u_i^r(f)$ as required.

Given a game Γ, we extend the model $\langle \Omega, (R_h)_{h \in H \setminus Z}, V \rangle$ above to the form $\langle \Omega, (R_h)_{h \in H \setminus Z}, (T_h)_{h \in H \setminus Z}, \left(\tilde{U}_h\right)_{h \in H \setminus Z}, V \rangle$ where $\Omega = N \times \prod_{h \in H \setminus Z} A(h)$ as before, the accessibility relations R_h ($h \in H \setminus Z$) are defined as before (satisfying 14–17). We define $T_h(\omega)(\omega') = \begin{cases} 1/n(\omega) & \omega R_h \omega' \\ 0 & \text{otherwise} \end{cases}$ where $n(\omega) = \#\{\omega' | \omega R_h \omega'\}$, i.e., equal probabilities to all states deemed possible at ω. Since ω contains a unique action for each identity there is a unique terminal node $z(\omega) \in Z$ in the game Γ such that all the actions in $z(\omega)$ are coordinates of ω, we define $\tilde{U}(\omega) = \tilde{u}_{P(h)}(z(\omega))$. Extending V according to 20–24 in addition to 18–19, it follows that $\langle \Omega, (R_h)_{h \in H \setminus Z}, (T_h)_{h \in H \setminus Z}, \left(\tilde{U}_h\right)_{h \in H \setminus Z}, V \rangle$ satisfies 20–29 and $V(\omega, f) = 1$ for every $f \in \mathcal{L}$ (the logical and epistemic closure of 1 – 7) and for every $\omega \in \Omega$ (hence the epistemic form is held in common confidence) and the proof is complete. $\blacksquare$

We now turn to games with nature moves. The main difference in the epistemic form that we construct is the addition of the restrictions on identities' beliefs that are generated by the distribution over moves of nature. An extensive form game with moves of nature can be represented as follows: $\Gamma = (N, H, P, \{\mu_h\}_{h \notin P^{-1}(N)}, \{G_i\}_{i \in N}, \{\tilde{u}_i\}_{i \in N})$ we let P be defined on a subset of $H \setminus Z$, each member of $H \setminus Z$ not in the domain of P is a nature move. For each nature move h we have a probability distribution μ_h over the set of acts of nature. Recall that $A(h)$ denotes the action set at decision point h, let $A(G) = A(h)$ for any $h \in G$. For extensive form games with nature moves the following augmentation to the epistemic form is required.

For the epistemic form we consider the language with identities $\mathcal{I} = P^{-1}(N)$ – all histories that are decision points for *players*. The atomic state-

[20]We can actually extend the definition to assume that $E_{T_h(\omega)}(\tilde{U}_h|\{\omega' \in \Omega|V(\omega', f) = 1\}$ can take multiple values when $T_h(\omega)(\{\omega' \in \Omega|V(\omega', f) = 1\}) = 0$ and modify 24 accordingly.

ments are $\alpha = N \cup \bigcup_{h \in H \setminus Z} A(h)$ – the names and all possible actions (by players and nature). In 1, 2, 5, 7 we replace $H \setminus Z$ with $P^{-1}(N)$. We do require that action sets be precise and proper for nature moves as well, hence 3, 4 remain intact. The main addition to the description of the game follows in the form of conditions on identities' beliefs given the prior distributions over nature moves. Let π denote a conjunction of statements that correspond to the actions in the pure action profile, i.e., $\pi = \bigwedge_{\text{all information sets } G} a_G$ for some choices $a_G \in A(G)$. For a given h and an action $a \in A(h)$ we denote by $\pi|_a$ the conjunction that is generated by replacing the member of $A(G(h))$ in π with a. The epistemic form of Γ is the logical and epistemic closure of the statements listed in the augmented 1, 2, 5, and 7, together with 3, 4 and with the addition of[21]:

For every pure action profile (conjunction) π, every identity $h \in \mathcal{I}$, every nature move $\bar{h}$ and acts of nature $a, b \in A(\bar{h})$ such that $\pi|_a \implies \bigvee_{h' \in G(h)} s(h')$ and $\pi|_b \implies \bigvee_{h' \in G(h)} s(h')$ we add

$$P_h^{\alpha}(\pi|_a) \iff P_h^{\beta}(\pi|_b) \qquad \text{whenever } \mu_{\bar{h}}(a)\beta = \mu_{\bar{h}}(b)\alpha \tag{30}$$

The collection of statements in 30 requires that if we take two pure strategy profiles – $\pi|_a$ and $\pi|_b$ – that differ only in an act of nature and such that both imply that the information set $G(h)$ is reached, then, at that information set, the player's beliefs about these two profiles preserve the proportion determined by the prior distribution over nature moves, i.e., conditioned on all else being equal, the identities' beliefs as to the nature moves are posteriors of the prior distribution of nature moves.

Now that the logical consistency of the epistemic form is established, we can formally ask whether various assumptions about rationality and reasoning are *consistent* with the epistemic form of the game, i.e., whether adding

[21]The proof of consistency of the epistemic form of extensive form games with nature moves and perfect agent recall is almost identical to that of games without nature moves. The main addition is a choice of identities' beliefs in the model that satisfies 30. To sketch an appropriate choice of belief distributions for identities, consider first a prior on all identities' choices that assigns equal probabilities to every action and is independently across identities and nature moves. Together with the distribution of nature moves, if we condition this prior distribution on the information set that corresponds to a given identity, we get a belief for this identity that is consistent with a model generated as in the proof of the previous theorem and preserves 30 as required.

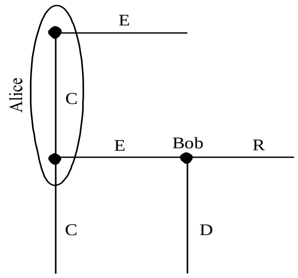

Figure 5: An extensive form game with no knowledge form.

the corresponding list of statements to the epistemic form results in a list that is consistent. In addition to consistency, we are interested in behavior. We wish to determine the behavior implied by adding such (consistent) conditions on rationality and reasoning.

Before we turn to the analysis of epistemic conditions, we note two limitations of the epistemic form: why we exclude games with agent imperfect recall and the extent to which the transition to the epistemic form of the game preserves the structure of a dynamic game.

We exclude games without agent perfect recall, because for some of these games the set of statements generated by $1-7$ is logically inconsistent. For example, in Appendix A we show that the game in figure 5 would have an inconsistent epistemic form.

Regarding the uniqueness of the epistemic form, the question we wish to answer is when do two different extensive form games produce logically equivalent epistemic forms.

We first note that the epistemic form is created with respect to a context, i.e., the set of identities and atomic statements already identify the number of decision points and the total number of actions and names. We also restrict attention to agent perfect recall games without nature moves, since there is little hope of uniqueness when it comes to nature moves[22]. It turns out that

[22]We limit the discussion to games without moves of nature, since when nature moves are allowed the probability distributions over these moves generate only restrictions on the identities beliefs, i.e., on posteriors conditioned on information sets. Clearly, these restrictions can sometimes be very weak and one should – generally – not expect uniqueness of the prior from given posteriors. As an example consider a game where nature moves first and all identities are informed of the outcome of nature's move,. Clearly there is no way to fully express the prior distribution of the nature move with the identities' posteriors.

the epistemic form identifies the extensive form almost uniquely, up to a very small class of equivalent extensive forms. The extensive forms that cannot be distinguished are exactly those that, from the point of view of the identities in the game, have the same dynamic structure.

Definition 9 *An extensive form game* $\Gamma = (N, H, P, \{G_i\}_{i\in N}, \{\tilde{u}_i\}_{i\in N})$ *has* epistemically equivalent *information sets if there exist two distinct information sets* $G, \bar{G}$ *such that* $\bigvee_{h\in G} s(h) \iff \bigvee_{h'\in\bar{G}} s(h')$ *is a tautology.*

Epistemically equivalent information sets are quite common. For example, in the extensive form of a normal form game we have that every pair of information sets is epistemically equivalent. However, changing the order of moves changes the epistemic form of these games. This follows from the observation that the epistemic form captures the size of an information set through the number of identities associated with it, as well as the number of actions available to members of the information set. Together with epistemic equivalence these characterize the degree to which the epistemic form captures an extensive form as is stated in the following proposition.

Proposition 10 *An extensive form game (without moves of nature) has a unique epistemic form as long as there are no two epistemically identical information sets with exactly the same number of decision points and the same number of actions for each decision point.*

Proof. See Appendix A. ■

As a general example, consider the games depicted in Figures 6 and 7 with identical (generic) payoffs associated with each history. These games have two distinct extensive forms since Bob immediately follows Alice's choice of L in 6, while Carol does in 7. However, it is easily verifiable that both games have the same epistemic form. Intuitively, there is no distinguishing feature between these two information sets that is observable from the point of view of a decision point in these information sets. For example, in both games Bob cannot distinguish between $(L \wedge a) \vee (L \wedge b)$ and L since they are logically equivalent, as are the number of identities that he has and the actions available to him in both games. However, in the game depicted in Figure 8 the players know that Bob follows Alice and that Carol follows Bob from the number of identities they each have. Hence, we have a different epistemic form for this game and – if one wishes – the analysis of behavior can depend on the (known) order of moves.

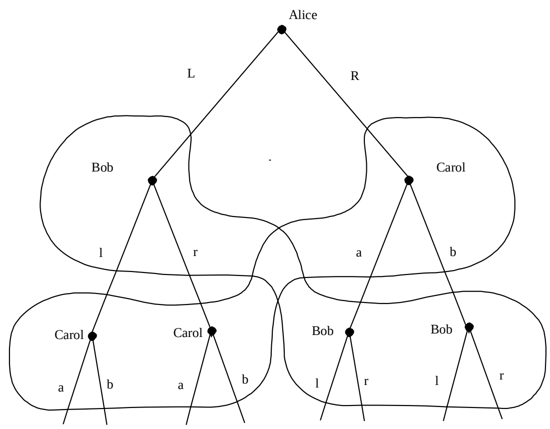

Figure 6: A game

6 Common Confidence, Rationality and the Centipede Game.

6.1 There is no Common Confidence of Rationality in the Centipede Game

Given the epistemic form of the centipede game detailed above, it is quite simple to show that there can be no common confidence of rationality in this game. This follows the spirit of the arguments suggested by [Reny 1992b].

Proposition 11 *Common confidence of rationality is inconsistent with the epistemic form of the centipede game.*

Proof. Recall that an identity is rational if for every two actions a, b of that identity, for all r we have $\neg u_i^r a \wedge u_i^r b \Longrightarrow \neg a$. Assume that we add the list of statements that correspond to common confidence of rationality. In particular A_2 is rational. Since $C_{A_2}(s_1 \wedge S)$ we have that $C_{A_2}((s_1 \wedge S \wedge e_2) \Longleftrightarrow (e_2))$. Since $U_{A_2}^3(s_1 \wedge S \wedge e_2)$ we have that $U_{A_2}^3(e_2)$ – note that these are sets of statements. Similarly we have that $U_{A_2}^2(s_2)$. Since s_2, e_2 are both A_2's choices we have that rationality implies that $C_{A_2} e_2$. We actually use the statements $u_{A_2}^{2.5}(e_2)$, $\neg u_{A_2}^{2.5}(s_2)$ which follow from the description of the game and $\neg u_{A_2}^{2.5}(s_2) \wedge u_{A_2}^{2.5}(e_2) \Longrightarrow \neg e_2$ which is one of the statements describing A_2's rationality.

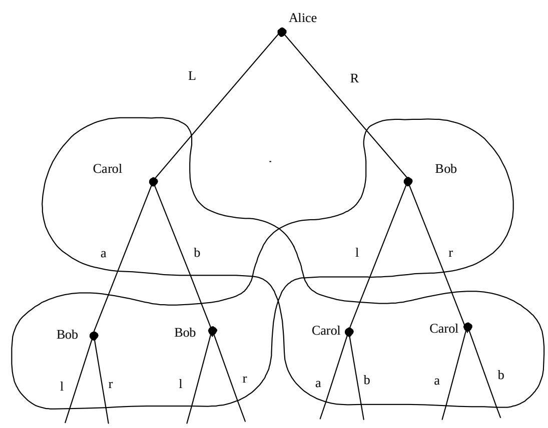

Figure 7: A different game

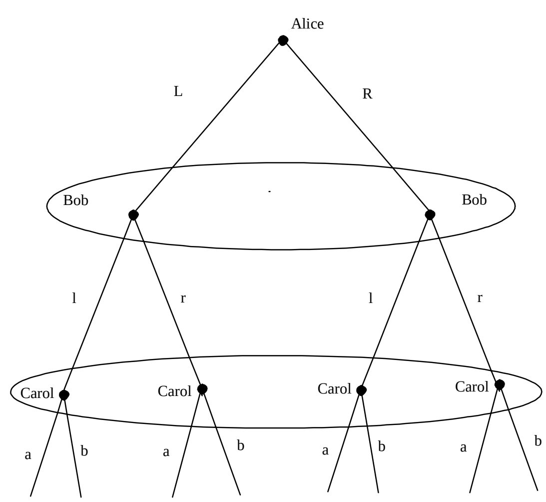

Figure 8: This game has a different epistemic form

All of our assumptions and deductions are accessible to the identities in our model. We only used the assumption that A_2 is rational so far. If B is confident that A_2 is rational we could follow all the above deduction prefixed by C_B. Doing so is quite tedious so we omit it, however, we note that the use of a finite number of statements in our deduction is crucial since C_B can only be applied to finite conjunctions of statements. We get $C_B C_{A_2} e_2$.

Since $C_B s_1$ and e_2 is one of A_2's choices, we have that $C_B(s_1 \wedge S \wedge e_2 \iff S)$ and $C_B(s_1 \wedge E \wedge e_2 \iff E)$ (note that $s_1 \wedge E \wedge e_2$ is not an empty event from Bob's view point. e_2 is the event that the *hypothetical* identity will exit, the fact that identity will not exist does not make this event false). We now have that $U_B^1(S)$ and $U_B^2(E)$ and Bob's rationality implies that $C_B E$.

Finally, this analysis can be repeated by A_2 as long as she is confident of our assumptions, and these are that Bob is confident that she is rational and that Bob himself is rational. Since these assumptions are part of the common confidence of rationality we assumed we will have A_2 deducing what we just did, i.e., $C_{A_2} C_B E$. But, $C_{A_2}(C_B E \implies E)$ so $C_{A_2} E$ and this contradicts the epistemic form stating that $C_{A_2} S$ and $C_{A_2}(S \iff \neg E)$ and the proof is complete. ■

6.2 A Subjective Condition on Rationality Implying Backward Induction

The following condition on rationality implies backward induction in the centipede game:

Let φ correspond to the sentence "every decision that has not been made yet will be rational". Then common confidence of φ is logically consistent with the epistemic form of the game and implies that *all* identities exit in the centipede game. Note that φ is not exactly a statement in our language since the set of decisions that have not been made yet depends on the subjective viewpoint of the reasoning identity. We need to formulate this condition as stating that there is common confidence that each identity i is confident that all the identities that have not made a decision (from i's subjective perspective) are rational. These identities that i is considering are exactly those that, in description of the game, i is not informed of their action.

Formally, the condition is expressed as the set of statements $CC(C_i R_j)$ $\forall i \neq j$ such that j does not precede i, we name this set of statements by *com-*

mon confidence of hypothetical rationality or CHR[23]. We have the following proposition:

Proposition 12 *Common confidence of hypothetical rationality is consistent with a game of perfect information and implies that every identity (real and hypothetical) follows the backward induction solution.*

Proof. Theorem 6 in [Feinberg 2001] shows that any reasonable solution satisfies CHR, but by definition the reasonable solution is consistent with the epistemic form of the game, hence CHR must be consistent with the game. See the proof of Proposition 13 for showing that CHR implies BI (backward induction). ■

This result is typical for the epistemic conditions that are studied in [Feinberg 2001]. We also want the weakest conditions that are strong enough to imply the relevant solution, but we also want them to be the strongest conditions that are weak enough not to contradict the description of the game – the epistemic form. This suggests that the following result may also be of interest since it provides weaker epistemic conditions that still imply backward induction.

The collection of statements $C_{i_0}C_{i_1}C_{i_2}...C_{i_{n-1}}R_{i_n}$ $\forall n, i_0, ..., i_n$ such that i_k follows i_{k-1} in the game, will be called *iterated future confidence of rationality* or FCR for short.

Proposition 13 *Rationality of all identities together with iterated future confidence of rationality –* FCR *– is consistent with the epistemic form of a game with perfect information and logically implies BI.*

Proof. See Appendix A. ■

7 Applications of Subjective Reasoning

7.1 The "Beer-Quiche" Game

We now return to the analysis of the "Beer-Quiche" game. First we present the epistemic form of the "Beer-Quiche" game. Our first result demonstrates

[23]We abuse the notion $CC(C_iR_j)$ to imply that there is common confidence in each and every one of the formulas C_if where $f \in R_j$.

that reasoning about rationality alone does not pose *any* restriction on behavior in this game. We then show that when we add the condition of common confidence that both of Alice's types have the same conjecture about Bob's behavior things change dramatically and iterative reasoning about rationality leads to the intuitive behavior.

Consider the language generated by the set of identities: $I = \{A_W,\ A_S,\ B_{B_1},\ B_{B_2},\ B_{Q_1},\ B_{Q_2}\}$ and atomic statements $\{W,\ S,\ Beer_1,\ Quiche_1,\ Beer_2,\ Quiche_2,\ F_B,\ NF_B,\ F_Q,\ NF_Q,\ Alice,\ Bob\}$ and the axiom scheme and derivation rules for the subjective framework. The epistemic form of the "Beer-Quiche" game is given by the logical and epistemic closure of the following list of statements:

Names: $C_{A_W}Alice$, $C_{A_S}Alice$, $C_{A_W}\neg Bob$, $C_{A_S}\neg Bob$, $C_{B_{B_1}}Bob, C_{B_{B_2}}Bob$, $C_{B_{Q_1}}Bob$, $C_{B_{Q_2}}Bob$, $C_{B_{B_1}}\neg Alice$, $C_{B_{B_2}}\neg Alice$, $C_{B_{Q_1}}\neg Alice$, $C_{B_{Q_2}}\neg Alice$

Actions:

$C_{A_W}Beer_1 \iff Beer_1, C_{A_W}Quiche_1 \iff Quiche_1, Beer_1 \iff \neg Quiche_1$, $C_{A_S}Beer_2 \iff Beer_2$, $C_{A_S}Quiche_2 \iff Quiche_2$, $Beer_2 \iff \neg Quiche_2$

$C_{B_{B_1}}F_B \iff F_B, C_{B_{B_1}}NF_B \iff NF_B, F_B \iff \neg NF_B, C_{B_{B_2}}F_B \iff F_B$, $C_{B_{B_2}}NF_B \iff NF_B$

$C_{B_{Q_1}}F_Q \iff F_Q, C_{B_{Q_1}}NF_Q \iff NF_Q, F_Q \iff \neg NF_Q, C_{B_{Q_2}}F_Q \iff F_Q$, $C_{B_{Q_2}}NF_Q \iff NF_Q$

Information sets:

$C_i(C_{B_{B_1}}f \iff C_{B_{B_2}}f)$, $C_i(C_{B_{Q_1}}f \iff C_{B_{Q_2}}f)$ for all statements f for every identity i.

Dynamics:

$C_{A_W}W$, $C_{A_S}S$, $C_{B_{B_1}}((W \wedge Beer_1) \vee (S \wedge Beer_2))$, $C_{B_{B_2}}((W \wedge Beer_1) \vee (S \wedge Beer_2))$, $C_{B_{Q_1}}((W \wedge Quiche_1) \vee (S \wedge Quiche_2))$, $C_{B_{Q_2}}((W \wedge Quiche_1) \vee (S \wedge Quiche_2))$

Nature moves:

$P^{\alpha}_{B_{B_1}}(W \wedge Beer_1 \wedge Beer_2) \implies P^{9\alpha}_{B_{B_1}}(S \wedge Beer_1 \wedge Beer_2)$, $P^{\alpha}_{B_{B_2}}(W \wedge Beer_1 \wedge Beer_2) \implies P^{9\alpha}_{B_{B_2}}(S \wedge Beer_1 \wedge Beer_2)$

$P^{\alpha}_{B_{B_1}}(S \wedge Beer_1 \wedge Beer_2) \implies P^{\frac{1}{9}\alpha}_{B_{B_1}}(W \wedge Beer_1 \wedge Beer_2)$, $P^{\alpha}_{B_{B_2}}(S \wedge Beer_1 \wedge Beer_2) \implies P^{\frac{1}{9}\alpha}_{B_{B_2}}(W \wedge Beer_1 \wedge Beer_2)$

$P^{\alpha}_{B_{Q_1}}(W \wedge Quiche_1 \wedge Quiche_2) \implies P^{9\alpha}_{B_{Q_1}}(S \wedge Quiche_1 \wedge Quiche_2)$, $P^{\alpha}_{B_{Q_2}}(W \wedge Quiche_1 \wedge Quiche_2) \implies P^{9\alpha}_{B_{Q_2}}(S \wedge Quiche_1 \wedge Quiche_2)$

$P^{\alpha}_{B_{Q_1}}(S \wedge Quiche_1 \wedge Quiche_2) \implies P^{\frac{1}{9}\alpha}_{B_{Q_1}}(W \wedge Quiche_1 \wedge Quiche_2)$, $P^{\alpha}_{B_{Q_2}}(S \wedge$

$Quiche_1 \wedge Quiche_2) \Longrightarrow P_{B_{Q_2}}^{\frac{1}{9}\alpha}(W \wedge Quiche_1 \wedge Quiche_2)$

Utilities:

$U_{A_W}^0(W \wedge Beer_1 \wedge F_B \wedge Quiche_2 \wedge F_Q)$, $U_{A_W}^0(W \wedge Beer_1 \wedge F_B \wedge Quiche_2 \wedge NF_Q)$, $U_{A_W}^0(W \wedge Beer_1 \wedge F_B \wedge Beer_2 \wedge F_Q), ...$[24]

Proposition 14 *Common confidence of rationality is logically consistent with the epistemic form of the "Beer-Quiche" game. Furthermore, it is consistent with* all *pure strategy profiles.*

Proof. See Appendix B. ■

Proposition 14 formalizes our intuition that without any further restrictions on reasoning any action of any given identity is a best response to some conjecture of that identity as to the behavior at other decision points. Maximal conditions on rationality and confidence of rationality in the form of common confidence of rationality are consistent with the epistemic form and have no implication on restricting the players' behavior. In contrast, we now add the assumption that A_W and A_S have the same conjecture about Bob's behavior. This – usually implicit – assumptions restricts the extent of reasoning about rationality that the epistemic form can sustain. Together with the maximal iterations of confidence in rationality that is consistent with the game we show that the intuitive behavior emerges.

We first define the additional epistemic condition formally. We say that there are *identical conjectures for Alice* if for every identities i, j of Alice and every action a of a player other than Alice, we have $P_i^\alpha a \iff P_j^\alpha a$. As always, common confidence of identical conjectures for Alice is defined as common confidence in each and every statements that defines identical conjectures for Alice. We say that *maximal assumptions on rationality* have been added to a description of the world if we sequentially added the maximal set of statements corresponding to rationality of all identities ($R = \bigwedge_{i \in I} R_i$), confidence of rationality for all identities ($\bigwedge_{i \in I} C_i R$), confidence of confidence of rationality ($C^2 R$) and so on, such that the new set of statements is consistent. For example, if we add maximal assumptions on rationality to an epistemic

[24]For each of the 6 identities we have 32 possible pure strategy profiles to consider (including nature's choice). There are 8 possible play paths that determine a payoff which take the values $0, 1, 2$ or 3. Here we stated 3 out of the 32 profiles, all 3 have the same play path $W, Beer_1$ and F_B determining the payoff of 0.

form, we add $R = \bigwedge_{i \in I} R_i$, $\bigwedge_{i \in I} C_i R$, $C^2 R$, $C^3 R, ...$ as long as the addition does not lead to a logical contradiction[25].

Proposition 15 *Common confidence that both types of Alice have identical conjectures about Bob's behavior and maximal assumptions on rationality and confidence of rationality, imply the intuitive criterion behavior for the "Beer-Quiche game.*

Proof. See Appendix B. ■

We find it interesting that the (intuitive) reasoning leading to the intuitive behavior does not invoke Nash equilibrium notions for this game.

Corollary 16 *Common confidence of identical conjectures for Alice and common confidence of rationality logically contradict the epistemic form of the "Beer-Quiche" game.*

Proof. Follows directly from the proof of Proposition 15. ■

7.2 Further Refinements in Signalling Games

In addition to the application of subjective reasoning to the "Beer-Quiche" game, we wish to consider other refinements for signalling games. Refinements such as Divinity and Universal Divinity due to [Banks and Sobel 1987] and criteria such as $D1$ (see [Fudenberg and Tirole 1991]) seem to have an even bigger gap between the formal definition and an intuitive justification[26]. Beyond results that relate these refinements to the notion of stability, the question is whether such refinements can be derived from the players' reasoning about behavior. We confront this question by considering the signalling game depicted in Figure 9. In this game the strategy profile (Out_1, Out_2, M) (both types of Alice choose Out and Bob chooses M) is an equilibrium which satisfies the intuitive criterion. However, this equilibrium fails criterion $D1$ (and universal divinity). This follows from the fact that for every distribution μ over the members of Bob's information set, if a distribution λ over $\{U, M, D\}$ is a best response of Bob to μ, then we have that if A_1 (weakly) prefers In_1 to Out_1 given λ then A_2 (strictly) prefers In_2 to Out_2. Hence

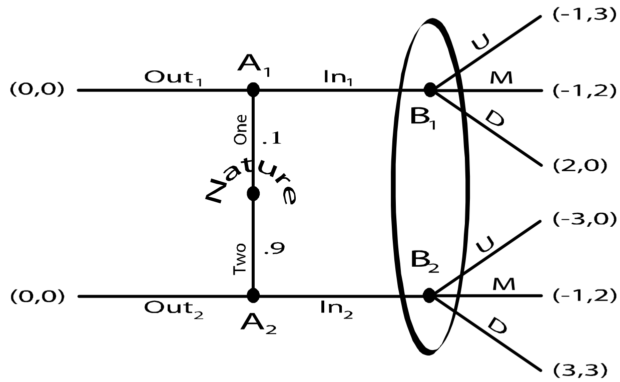

Figure 9: Another signalling game.

these refinements imply that Bob will assume that A_2 has "deviated" and will choose D. This leaves us with the equilibrium (In_1, In_2, D).

We begin our study of the reasoning behind this refinement by observing that common confidence that Alice's types have identical conjectures and common confidence of rationality do not contradict the epistemic form of this game. Furthermore, these assumptions are all logically consistent with the strategy (Out_1, Out_2, M).

Proposition 17 *The epistemic form of the game in Figure 9 is logically consistent with common confidence of identical conjectures of Alice, common confidence of rationality and the actions Out_1, Out_2 and M*[27]

Proof. See Appendix C. ■

This result is not surprising given our intuitive explanation of the "Beer-Quiche" game. In the current game the mere assumptions that A_1 and A_2 have identical conjectures does not exclude any behavior of Alice's types from being rational. Refinements such as $D1$ are based on comparisons of best response regions as criteria for establishing Bob's conjectures about Alice's types. The objective of these refinements for this game is for Bob to consider In_2 to be at least as likely as In_1 even "off" the equilibrium.

[25]This is an instance of the reasonable solution concept presented in [Feinberg 2001].

[26]See [Fudenberg and Tirole 1991] p. 452.

[27]We actually show that *all* pure strategy profiles are consistent with these conditions.

It turns out that adding an assumption about subjective reasoning about higher order uncertainty does eliminate (Out_1, Out_2, M) from being consistent with the game, CC(identical conjectures) and CCR, while retaining (In_1, In_2, D). The assumption needed is that each player is confident that the other player has a specific conjecture as to the first player's behavior. There is no second order uncertainty about conjectures.

Definition 18 *We say that* Bob is certain about Alice's conjectures *if there exists a single probability distribution α defined on Bob's actions such that Bob is confident that Alice's beliefs follow α. We can write this as* $P^{\beta}_{Bob}(\bigwedge_{a\in A(Bob)} P^{\alpha(a)}_{Alice} a) \Longrightarrow C_{Bob}(\bigwedge_{a\in A(Bob)} P^{\alpha(a)}_{Alice} a)$ *for every distribution α and $\beta > 0$.*

However, since we are in a subjective framework we need to note two things. First, even if Bob is certain that Alice has a specific conjecture this does not mean that this indeed is her conjecture. Moreover, even if Alice is confident that Bob is certain of her conjectures, she may still be uncertain as to what Bob thinks these conjectures are. The fact that she knows that he is certain about her conjectures does not imply that she knows which conjectures he attributes to her.

Definition 19 *We say that* Alice and Bob have mutual certainty about conjectures *if Alice is certain about Bob's conjectures and Bob is certain about Alice's conjectures.*

We find that common confidence that Alice and Bob have mutual certainty about conjectures eliminates (Out_1, Out_2, M).

Theorem 20 *The epistemic form of the game together with CC(identical conjectures), CCR, and CC(mutual certainty about conjectures) are logically consistent and imply (In_1, In_2, D).*

Proof. See Appendix C. ■

As in the "Beer-Quiche" game, we did not invoke conditions that constrain the discussion to a Nash equilibrium. However, the condition used here to *refine* the intuitve criterion might seem similar to the condition of mutual knowledge of conjectures (and rationality) that appears in [Aumann and Brandenburger 1995] and characterizes Nash equilibrium in a two-person normal form game. This similarity is misleading since the main thrust of mutual knowledge of conjectures is that knowledge implies truth, hence what you think (know) are the conjectures of someone else actually *are* their conjecture. Furthermore, if you know that they know your conjectures you have no doubt what they think your conjectures are. As noted above, these two features, in general, do not hold when we consider confidence instead of knowledge.

7.3 Burning Money – Reasoning about Reasoning about Oneself

We conclude the applications of the subjective framework with the analysis of the "Burning Money" game depicted in figure 3. Consider the language generated by the set of identities: $I = \{A_0,\ A_1,\ A_2,\ B_{1_1},\ B_{1_2},\ B_{2_1},\ B_{2_2}\}$. We have two identities for each of Bob's information sets. The set of atomic statements is $\{Burn,\ NB,\ L_1,\ R_1,\ L_2,\ R_2,\ l_1, r_1, l_2, r_2,\ Alice,\ Bob\}$. We will use the notation B_1 (B_2) whenever a statement holds for both B_{1_1} and B_{1_2} (B_{2_1} and B_{2_2}). The epistemic form is the logical and epistemic closure of the following list of statements:

Names: $C_{A_0}Alice$, $C_{A_1}Alice$, $C_{A_2}Alice$, $C_{A_0}\neg Bob$, $C_{A_1}\neg Bob$, $C_{A_2}\neg Bob$, $C_{B_1}Bob$, $C_{B_2}Bob$, $C_{B_1}\neg Alice$, $C_{B_2}\neg Alice$

Actions:

$C_{A_0}Burn \iff Burn$, $C_{A_0}NB \iff NB$, $Burn \iff \neg NB$, $C_{A_1}L_1 \iff L_1$, $C_{A_1}R_1 \iff R_1$, $L_1 \iff \neg R_1$, $C_{A_2}L_2 \iff L_2$, $C_{A_2}R_2 \iff R_2$, $L_2 \iff \neg R_2$

$C_{B_1}l_1 \iff l_1$, $C_{B_1}r_1 \iff r_1$, $l_1 \iff \neg r_1$, $C_{B_2}l_2 \iff l_2$, $C_{B_2}r_2 \iff r_2$, $l_2 \iff \neg r_2$

Information sets:

$C_i(C_{B_{1_1}}f \iff C_{B_{1_2}}f)$, $C_i(C_{B_{2_1}}f \iff C_{B_{2_2}}f)$ for all statements f for every identity i.

Dynamics:

$C_{A_1}Burn$, $C_{A_2}NB$, $C_{B_1}Burn$, $C_{B_2}NB$

Utilities:

$U^{6.5}_{A_0}(Burn \wedge L_1 \wedge l_1 \wedge L_2 \wedge l_2)$, $U^{6.5}_{A_1}(Burn \wedge L_1 \wedge l_1 \wedge L_2 \wedge l_2)$, $U^{6.5}_{A_2}(Burn \wedge L_1 \wedge l_1 \wedge L_2 \wedge l_2)$, $U^{6}_{B_1}(Burn \wedge L_1 \wedge l_1 \wedge L_2 \wedge l_2)$, $U^{6}_{B_2}(Burn \wedge L_1 \wedge l_1 \wedge L_2 \wedge l_2)$, ...[28]

[28]For each of the 7 identities we have 32 possible pure strategy profiles to consider. There are 8 possible play paths that determine a payoff. Here we stated the payoff to

As we discussed in Section 2, there are quite unique conditions on reasoning that would enforce the outcome preferable by Alice – the payoff generated by the choice of (NB, L_2, l_2). We begin by stating various conditions that do not uniquely imply this outcome or are inconsistent with the structure of the game. This also relates the notions studied above of identical conjectures and maximal assumptions of rationality to this game.

Theorem 21 *In the "Burning Money" game the following results hold:*

1. Common confidence of rationality is consistent with the form of the game but does not uniquely *imply* $(NB, L_2.l_2)$

2. Common confidence of identical conjectures of Alice's types and common confidence of rationality are inconsistent with the epistemic form of the game.

3. Common confidence of identical conjectures of Alice's types and maximal assumptions of rationality do not necessarily imply that $(NB, L_2.l_2)$ *is the outcome of the game.*

Proof. See Appendix D. ■

The reason we do not get the outcome $(NB, L_2.l_2)$ from maximal reasoning about rationality is that under the assumption of common confidence of identical conjectures for Alice, we have that confidence of confidence of rationality already implies that A_0 is indifferent between *Burn* and NB. While this is covered in the proof of Theorem 21, the argument is fairly straightforward. If A_0 is confident that A_1 and A_2 are rational and have identical conjectures, then A_0's rationality implies that *Burn* is chosen only if the expected payoff from a best response by A_1 is *at least* the expected payoff from a best response by A_2 and vice versa. If A_1 is confident that A_0 is confident that both A_1 and A_2 are rational, and if A_1 is confident that A_0 is rational, then since A_1 is confident of *Burn* we have that the expected payoff from a best response to the conjectures is not higher at NB. If A_2 is confident of confidence in rationality and is also confident of A_0's rationality, we have the opposite inequality and the expected payoff from *Burn* and NB is equal. The source of this constraint is that A_2 and, more importantly, A_1, both assume that A_0 is rational even though they each observe a different action taken by A_0 – hence the indifference outcome.

each identity for a single strategy profile. Note that from the structure of the game some identities will consider this profile to be subjectively impossible. However, they can still assign expected payoff to such an event.

The following result demonstrates how relaxing this condition provides epistemic conditions that lead to $(NB, L_2.l_2)$ as the unique outcome of the game, as was demonstrated in Section 2.

Proposition 22 *The epistemic form of the game and conditions (1)–(8) below uniquely imply the outcome* $(NB, L_2.l_2)$ *in the "Burning Money" game*[29].

1. *Common confidence of identical conjectures of Alice*
2. *Rationality of all identities*
3. *Confidence of* A_0 *that* A_1 *and* A_2 *are rational*
4. *Confidence of Alice's identities in* Bob's *identities rationality*
5. *Confidence of Bob's identities in (2) –* all *identities' rationality, and in (3) and (4).*
6. *Confidence of Alice's identities in (5)*
7. *Confidence of Bob's identities in (6)*
8. *Confidence of Alice's identities in (7)*

Proof. See Appendix D. ■

The conditions that lead to the outcome $(NB, L_2.l_2)$ according to Proposition 22 are mostly symmetric, with the exception of condition (3) and its influence on higher orders of confidence in rationality. The asymmetry arises since we do not allow A_1 and A_2 to be confident in A_0's rationality. Relaxing this requirement would lead to a contradiction. This suggests the nature of restrictions on reasoning that are implicit when the games in Figures 3 and 4 are treated as identical.

[29]The eight conditions presented here imply the conditions used in presenting the result in Section 2. We provide the conditions in this form to allow for as much symmetry as possible between the reasoning attributed to players. In Section 2 we took only the statements directly needed for the proof of the result out of conditions (1) – (8). While some conditions can be replaced with weaker ones, it is important to note that no strict subset of these conditions can uniquely imply $(NB, L_2.l_2)$.

References

Asheim, G. B. (Forthcoming): "On the Epistemic Foundations for Backward Induction," *Mathematical Social Sciences*.

Asheim, G. B., and A. Perea (2000): "Lexicographic Probabilities and Rationalizability in Extensive Games," *preprint*.

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

Aumann, R. J. (1987): "Correlated equilibrium as an expression of Bayesian rationality," *Econometrica*, 55(1), 1–18.

——— (1995): "Backward induction and common knowledge of rationality," *Games Econom. Behav.*, 8(1), 6–19, Nobel Symposium on Game Theory (Björkborn, 1993).

——— (1998): "On the Centipede Game," *Games and Economic Behavior*, 23, 97–105.

——— (1999a): "Interactive epistemology. I. Knowledge," *Internat. J. Game Theory*, 28(3), 263–300.

——— (1999b): "Interactive epistemology. II. Probability," *Internat. J. Game Theory*, 28(3), 301–314.

Balkenborg, D., and E. Winter (1997): "A necessary and sufficient epistemic condition for playing backward induction," *J. Math. Econom.*, 27(3), 325–345.

Banks, J. S., and J. Sobel (1987): "Equilibrium selection in signaling games," *Econometrica*, 55(3), 647–661.

Battigalli, P., and M. Siniscalchi (1999): "Hierarchies of conditional beliefs and interactive epistemology in dynamic games," *J. Econom. Theory*, 88(1), 188–230.

——— (2000): "Epistemic Characterizations of the (Iterated) Intuitive Criterion," *Preprint*.

——— (Forthcoming): "Interactive Beliefs and Forward Induction," *Journal of Economic Theory.*

Ben-Porath, E., and E. Dekel (1992): "Signaling future actions and the potential for sacrifice," *J. Econom. Theory*, 57(1), 36–51.

Benthem, J. V. (1999): "When are Two Games the Same," *Unpublished manuscript.*

Bernheim, B. D. (1984): "Rationalizable strategic behavior," *Econometrica*, 52(4), 1007–1028.

Bonanno, G. (1992): "Set-theoretic equivalence of extensive-form games," *Internat. J. Game Theory*, 20(4), 429–447.

Bonanno, G. (1993): "The logical representation of extensive games," *Internat. J. Game Theory*, 22(2), 153–169.

Brandenburger, A. (2001): "The Power of Paradox," *Preprint*, (6).

Brandenburger, A., and E. Dekel (1987): "Rationalizability and correlated equilibria," *Econometrica*, 55(6), 1391–1402.

Brandenburger, A., and H. J. Keisler (2002): "Epistemic Conditions for Iterated Admissibility," *Preprint.*

Chellas, B. F. (1980): *Modal logic.* Cambridge University Press, Cambridge-New York, An introduction.

Cho, I.-K., and D. M. Kreps (1987): "Signaling games and stable equilibria," *Quart. J. Econom.*, 102(2), 179–221.

Clausing, T. (2000): "Belief Revision in Games of Perfect Information," *University of Magdeburg Working Paper.*

Clausing, T., and A. Vilks (2000): "Backward Induction in General Belief Structures with and Without Strategies," *Unpublished manuscript.*

Fagin, R., and J. Y. Halpern (1994): "Reasoning about knowledge and probability," *J. Assoc. Comput. Mach.*, 41(2), 340–367.

Fagin, R., J. Y. Halpern, and N. Megiddo (1990): "A logic for reasoning about probabilities," *Inform. and Comput.*, 87(1-2), 78–128.

Fagin, R., J. Y. Halpern, Y. Moses, and M. Y. Vardy (1995): *Reasoning About Knowledge*. MIT Press, MA.

Feinberg, Y. (2001): "Epistemic Characterizations of Equilibria and the Reasonable Solution," *preprint*.

Fudenberg, D., and J. Tirole (1991): *Game theory*. MIT Press, Cambridge, MA.

Heifetz, A., and P. Mongin (2001): "Probability Logic for Type Spaces," *Games Econom. Behav.*, 35, 31–53.

Hughes, G. E., and M. J. Cresswell (1996): *A new introduction to modal logic*. Routledge, London.

Meier, M. (2000): "An Infinitary Modal Logic for Type Spaces," *preprint*.

Osborne, M. J., and A. Rubinstein (1994): *A course in game theory*. MIT Press, Cambridge, MA.

Pearce, D. G. (1984): "Rationalizable strategic behavior and the problem of perfection," *Econometrica*, 52(4), 1029–1050.

Reny, P. J. (1992a): "Backward induction, normal form perfection and explicable equilibria," *Econometrica*, 60(3), 627–649.

——— (1992b): "Rationality in Extensive-Form Games," *Journal of Economic Perspectives*, 6(4), 103–118.

Samet, D. (1996): "Hypothetical knowledge and games with perfect information," *Games Econom. Behav.*, 17(2), 230–251.

Stalnaker, R. (1998): "Belief revision in games: forward and backward induction," *Math. Social Sci.*, 36(1), 31–56, Logic & foundation of the theory of games and decisions, 2 (Turin, 1996).

8 Appendix A

Lemma 23 *The game in Figure 5 has no epistemic form, i.e., the epistemic form constructed from 1 – 7 for this game, yields a logically inconsistent set of statements within our axiom scheme.*

Proof. An epistemic form game for this extensive form game should have three identities: A_1, A_2, B , Alice's decision making point at the first intersection, Alice's decision making point at the second intersection and Bob's decision making point. Since A_1, A_2 are at the same information set we have that $C_B(C_{A_1} f \iff C_{A_2} f)$ for all f is part of the description of the game. However, by the extensive form game Bob must be confident that Alice continues at the first intersection and exits at the second intersection. Consider the atomic formula corresponding to Alice's choice to continue at the first intersection and denote this formula by a. From the definition of choice we have that $C_B(C_{A_1} a \iff a)$ and since $C_B a$ we have that $C_B C_{A_1} a$ and so $C_B C_{A_2} a$. But from the definition of an information set we have that Alice in the second intersection will choose to continue if and only if she is confident that she makes the same choice at the first intersection. So if we denote by b the atomic statement that Alice continues at the second intersection, then $C_{A_2}(a \iff b)$ is in common confidence. So we have that $C_B C_{A_2}(a \iff b)$ and with $C_B C_{A_2} a$ above we conclude that $C_B C_{A_2} b$. Since b is a choice for Alice at the second intersection we have that $C_B b$ a contradiction to extensive form game where Bob is confident that Alice does not continue at the second intersection. ■

It should be noted that the determining feature of the game in Figure 5, is that it has a decision point right after contradicting actions in a single information set. Here we see that, within the subjective framework, mixed behavior strategies are interpreted using uncertainty of others as to the pure action chosen. If we want to use explicit randomizations by players we have to introduce the choice of randomization devices explicitly in our game, or change the subjective framework. Since, as long as we retain that an identity is confident of the *pure* action she chooses, and that there is common confidence that identities have identical views of the world at a given information set, the game in 5 becomes logically inconsistent.

Proof of Proposition 10. We need to show that if $\Gamma = (N, H, P, \{G_i\}_{i \in N}, \{\tilde{u}_i\}_{i \in N})$ and $\Gamma' = (N', H', P', \{G'_i\}_{i \in N'}, \{\tilde{u}'_i\}_{i \in N'})$ are two extensive form games such that the epistemic and logical closures of the formulae generated by 1 – 7 coincide, then Γ is identical to Γ' up to a permutation of information sets within every collection of identical information sets. Following the result of [Bonanno 1992] it suffices to show that the epistemic form uniquely determines the role of the atomic formulae as actions or names and that if all players know the dynamics of the game they extensive form is uniquely determined. Denote this logical and epistemic closure by $\mathcal{L}$. We first note that the syntax implies $N \cup \bigcup_{h \in H \setminus Z} A(h)$ and $N' \cup \bigcup_{h' \in H' \setminus Z'} A'(h')$ are sets of the same size (all atomic formulae). Additionally, the syntax implies that the number of histories in $H \setminus Z$ equals that of $H' \setminus Z'$ (the number of identities). Next we need to show that the atomic formulae maintain their types, i.e., names remain names and actions stay actions. We note that for names every identity is confident of the formula or its negation by 1. However, for actions there is always at least one identity that in the description of the game it is not specified whether she is confident of the formulae or confident in its negation. So if a is an action in Γ (specifically an action of h) then neither $C_h a$ nor $C_h \neg a$ are in $\mathcal{L}$ (as can be seen if one constructs the model presented in the proof of proposition 8). Thus, the set of actions are exactly the atomic formulae for which there is an identity h such that neither $C_h a$ nor $C_h \neg a$ are part of the game's description. In particular N and N' are sets of the same size. Next we show that if two identities $h, \bar{h}$ in Γ are in the same information set then the corresponding identities in Γ' are in the same information set. Let a be h's action. Then $CC(C_h a \iff a)$ implies that we have $CC(C_h a \vee C_h \neg a) \subset \mathcal{L}$ but $C_h a \notin \mathcal{L}$ and $C_h \neg a \notin \mathcal{L}$. Since $\bar{h}$ follows $CC(C_h f \iff C_{\bar{h}} f)$ for all f we have that $CC(C_h a \vee C_h \neg a) \subset \mathcal{L}$ but if $\bar{h}$ is not in the same information set this can only imply that $\bar{h}$ has subjective knowledge about the action a from the dynamics of the game Γ' , i.e., $C_{\bar{h}} a \in \mathcal{L}$ or $C_{\bar{h}} \neg a \in \mathcal{L}$ contradicting our assumption.

As noted above, following [Bonanno 1992] set theoretic form of a dynamic game we need to verify that if all identities subjectively know the dynamics of the game then it is uniquely determined. Consider the special cases depicted in Figures 6 – 8. The epistemic form of the first two games is given by the epistemic and logical closure of the following:

$I = \{A, B_1, B_2, B_3, C_1, C_2, C_3\}$ (seven decision making points) and the alphabet has nine elements $\{L, R, l, r, a, b, Alice, Bob, Carol\}$.

Names:

$CC(C_A Alice \wedge C_{B_1} \neg Alice \wedge C_{B_2} \neg Alice \wedge C_{B_3} \neg Alice \wedge C_{C_1} \neg Alice \wedge C_{C_2} \neg Alice \wedge C_{C_3} \neg Alice)$

$CC(C_A \neg Bob \wedge C_{B_1} Bob \wedge C_{B_2} Bob \wedge C_{B_3} Bob \wedge C_{C_1} \neg Bob \wedge C_{C_2} \neg Bob \wedge C_{C_3} \neg Bob)$

$CC(C_A \neg Carol \wedge C_{B_1} \neg Carol \wedge C_{B_2} \neg Carol \wedge C_{B_3} \neg Carol \wedge C_{C_1} Carol \wedge C_{C_2} Carol \wedge C_{C_3} Carol)$

Actions:

$CC(C_A L \iff L)$

$CC(C_A R \iff R)$
$CC(L \iff \neg R)$
$\vdots$
$\vdots$
Dynamics:
No additional formula is added to describe the dynamics.
Information Sets:
$CC(C_{B_1} f \iff C_{B_2} f \iff C_{B_3} f)$
$CC(C_{C_1} f \iff C_{C_2} f \iff C_{C_3} f)$

We have omitted the description of preferences assuming they are the same for all three games. For the game in Figure 8 the epistemic form is given by:

$I = \{A, B_1, B_2, C_1, C_2, C_3, C_4\}$ (seven decision making points) and the alphabet has nine elements $\{L, R, l, r, a, b, Alice, Bob, Carol\}$.

Names:
$CC(C_A Alice \wedge C_{B_1} \neg Alice \wedge C_{B_2} \neg Alice \wedge C_{C_1} \neg Alice \wedge C_{C_2} \neg Alice \wedge C_{C_3} \neg Alice \wedge C_{C_4} \neg Alice)$
$CC(C_A \neg Bob \wedge C_{B_1} Bob \wedge C_{B_2} Bob \wedge C_{C_1} \neg Bob \wedge C_{C_2} \neg Bob \wedge C_{C_3} \neg Bob \wedge C_{C_4} \neg Bob)$
$CC(C_A \neg Carol \wedge C_{B_1} \neg Carol \wedge C_{B_2} \neg Carol \wedge C_{C_1} Carol \wedge C_{C_2} Carol \wedge C_{C_3} Carol \wedge C_{C_4} Carol)$
Actions:
$CC(C_A L \iff L)$
$CC(C_A R \iff R)$
$CC(L \iff \neg R)$
$\vdots$
$\vdots$
Dynamics:
Once again no additional formula is added to describe the dynamics.
Information Sets:
$CC(C_{B_1} f \iff C_{B_2} f)$
$CC(C_{C_1} f \iff C_{C_2} f \iff C_{C_3} f \iff C_{C_4} f)$

The difference between the two description is the identification of the size of each information set. This is available from the epistemic form as shown above. Hence the epistemic form provides a description of an extensive form game that is finer than the set theoretic description. ■

Proof of Proposition 13. Since CHR implies FCR we have that FCR is consistent with the description of the game. The proof that FCR implies backward induction follows from backward induction...

Consider a game $\Gamma = (N, H, P, \{G_i\}_{i \in N}, \{\tilde{u}_i\}_{i \in N})$ with perfect information, i.e., with $G(h) = \{h\}$ for all $h \in H \setminus Z$. For every maximal $h \in H \setminus Z$ with respect to sub-histories we have that R_h implies that h will not choose a dominated action $a \in A(h)$, i.e., $R_h \to \neg b$ for all $b \in A(h)$ such that $\tilde{u}_{P^{-1}(h)}(h, b) < \tilde{u}_{P^{-1}(h)}(h, a)$ for some $a \in A(h)$. This follows from the fact that h is confident that the history h indeed occurred (by the definition of the epistemic form of the game), i.e., h's expected utility from the choice b is $\tilde{u}_{P^{-1}(h)}(h, b)$ and we have that the epistemic form of the game implies that the set of statements $U^r_{P^{-1}(h)} b$ holds with $r = \tilde{u}_{P^{-1}(h)}(h, b)$. Similarly we have that $U^s_{P^{-1}(h)} a$ holds with $s = \tilde{u}_{P^{-1}(h)}(h, a)$ and hence we have that the rationality of h implies that h will not choose a dominated action *given* that she is reached. The induction argument is based on the observation that the previous argument we made is fully accessible to all past identities. Consider an identity h in the game and assume that we have shown the following:

For every future identity $h' = (h, l)$ for some non-empty continuation of h we have that the collection of formulae $C_{i_0} C_{i_1} C_{i_2} ... C_{i_{n-1}} R_{i_n}$ such that i_k follows i_{k-1} and $i_0 = h'$ together with $R_{h'}$ imply that h' will follow the backward induction solution.

It is important to note that the *implication* is our induction assumption[30]. Since h is confident in each of the formulae $C_{i_0} C_{i_1} C_{i_2} ... C_{i_{n-1}} R_{i_n}$ such that i_k follows i_{k-1} and $i_0 = h'$ for every continuation h' we have by the induction assumption that this logically implies that the continuations follow the backward induction solution, hence h can logically deduce (just like we did) that all the continuations follow the backward induction solution[31]. Now, the rationality of h implies that he follows the backward induction solution (just like the first stage of the proof) since his beliefs about future play follow the backward induction solution. ■

Note that there is no need to show that the set of statements considered

[30] It is important to note that only a finite number of statements out of U^r_i are actually needed for each implication, and the same holds for statements from the sets R_i, CR_i etc.

[31] In the case where indifference of one identity may cause the expected utility of another identity not to be predetermined, neither we (as modelers) nor a third identity reasoning about such future situation can fully determine the future backward induction play. We see all possible plays as part of the backward induction solution for these general games with perfect information.

is consistent with the description of the game since it is a subset of the collection considered in Proposition 12.

9 Appendix B

Proof of Proposition 14. It suffices to construct a model for the "Beer-Quiche" game in which identities are rational at every state of the world (hence there is common confidence of rationality) and for every pure strategy profile there is a state where this profile holds.

Consider the set $\Omega = \{Alice, Bob\} \times \{W, S\} \times \{Beer_1, Quiche_1\} \times \{Beer_2, Quiche_2\} \times \{F_B, NF_B\} \times \{F_Q, NF_Q\}$ and define the following relations:

$$(., ., Beer_1, ., ., .) R_{A_W} (Alice, W, Beer_1, ., NF_B, F_Q)$$

$$(., ., Quiche_1, ., ., .) R_{A_W} (Alice, W, Quiche_1, ., F_B, F_Q)$$

$$(., ., Quiche_1, ., ., .) R_{A_W} (Alice, W, Quiche_1, ., ., NF_Q)$$

$$(., ., ., Beer_2, ., .) R_{A_S} (Alice, S, ., Beer_2, ., F_Q)$$

$$(., ., ., Beer_2, ., .) R_{A_S} (Alice, S, ., Beer_2, NF_B, NF_Q)$$

$$(., ., ., Quiche_2, ., .) R_{A_S} (Alice, S, ., Quiche_2, F_B, NF_Q)$$

$$(., ., ., ., F_B, .) R_{B_B} (Bob, W, Beer_1, Quiche_2, F_B, .)$$

$$(., ., ., ., NF_B, .) R_{B_B} (Bob, ., ., Beer_2, NF_B, .)$$

except when the right hand side is $(Bob, W, Quiche_1, Beer_2, NF_B, .)$

$$(., ., ., ., ., F_Q) R_{B_Q} (Bob, W, Quiche_1, Beer_2, ., F_Q)$$

$$(., ., ., ., ., NF_Q) R_{B_Q} (Bob, ., ., Quiche_2, ., NF_Q)$$

except when the right hand side is $(Bob, W, Beer_1, Quiche_2, ., NF_Q)$

We assume that B_{B_1}, B_{B_2} and B_{Q_1}, B_{Q_2} have identical relations denoted R_{B_B} and R_{B_Q} respectively (this will satisfy the epistemic conditions for information sets) and that (.) indicates an arbitrary value of the corresponding coordinate in ω. By definition, for all i and every ω there exists an ω' such that $\omega R_i \omega'$ and it is easily verified that for all $i, \omega, \omega', \omega''$ we have $\omega R_i \omega'$ and $\omega' R_i \omega''$ imply $\omega R_i \omega''$ as well as $\omega R_i \omega'$ implies $\omega' R_i \omega'$ hence $KD4U$ holds. The epistemic form of the game holds at every ω since all identities know their choices, the dynamics are preserved e.g., $\cdot R_{A_S} \omega$ implies that ω's first coordinate is S, or $\cdot R_{B_Q} \omega$ implies that either ω has W and $Quiche_1$ or it has S and $Quiche_2$ in the corresponding coordinates. We now need to add beliefs and utilities. Utilities are added as usual for every ω according to the payoff in the unique play path implied by ω. The choice of beliefs is quite arbitrary as long as they satisfy the axioms and respect the structure of the game. Satisfying the axioms implies that the beliefs of i at ω have to be supported on the range of $\omega R_i \cdot = \{\omega' | \omega R_i \omega'\}$ and that for states with identical ranges i has identical beliefs. Respecting the structure of the game implies that the conditions on beliefs with respect to the actions of nature are satisfied, e.g., B_Q finds the a state with $W \wedge Quiche_1 \wedge Quiche_2$ nine times more likely than $S \wedge Quiche_1 \wedge Quiche_2$. The latter property implies that all identities are rational at every state ω in such a model. Let us verify this condition for B_{Q_1} at $\omega = (Bob, S, Beer_1, Beer_2, NF_B, NF_Q)$ at this state we have that B_Q (both identities) is not fighting, hence she finds the following states possible from her subjective view point: $\omega_1 = (Bob, S, Beer_1, Quiche_2, F_B, NF_Q)$, $\omega_2 = (Bob, S, Beer_1, Quiche_2, NF_B, NF_Q)$, $\omega_3 = (Bob, S, Quiche_1, Quiche_2, F_B, NF_Q)$, $\omega_4 = (Bob, S, Quiche_1, Quiche_2, NF_B, NF_Q)$, $\omega_5 = (Bob, W, Quiche_1, Quiche_2, F_B, NF_Q)$, $\omega_6 = (Bob, W, Quiche_1, Quiche_2, NF_B, NF_Q)$ whatever her beliefs at ω are they are supported on these six states and satisfy that the probability of the set $\{\omega_3, \omega_4\}$ is nine times more likely than that of $\{\omega_5, \omega_6\}$, given such beliefs B_Q's unique best response is NF_Q which implies that B_Q is rational at ω. Following the same analysis for every identity i at every ω we find that they are all rational at every ω and hence we have common confidence of rationality. Since we have a state for every possible strategy profile the proof is complete. ■

Proof of Proposition 15. We have to show that common confidence of identical conjectures for Alice and maximal consistent assumptions about rationality and higher orders of confidence of rationality imply the intuitive behavior. Our proof follows the exact reasoning made in Section 2.

Consider the epistemic form of the game as given in Section 7.1. Common conjectures for Alice can be represented by the following statements:

$$\begin{aligned} P^{\alpha}_{A_W} F_B &\iff P^{\alpha}_{A_S} F_B \quad \forall \alpha \\ P^{\alpha}_{A_W} NF_B &\iff P^{\alpha}_{A_S} NF_B \quad \forall \alpha \\ P^{\alpha}_{A_W} F_Q &\iff P^{\alpha}_{A_S} F_Q \quad \forall \alpha \\ P^{\alpha}_{A_W} NF_Q &\iff P^{\alpha}_{A_S} NF_Q \quad \forall \alpha \end{aligned} \tag{31}$$

Now consider adding common confidence of each of the statements in 31 to the epistemic form of the game. Given this extended list of statements we now begin adding statements that correspond to the rationality and confidence in rationality of identities. As we add these statements we consider their logical implications on behavior.

We first add the set of statements R which denotes the set of statements corresponding to $R = \bigcup_{i \in I} R_i$ – the rationality of all identities. The rationality of A_W implies that if $Beer_1$ holds we must have that $P^{\alpha}_{A_W} NF_Q \implies P^{\frac{1}{2}+\alpha}_{A_W} NF_B$, i.e., the following collection of statements are logically implied by adding rationality to the description of the game.

$$Beer_1 \implies (P^{\alpha}_{A_W} NF_Q \implies P^{\frac{1}{2}+\alpha}_{A_W} NF_B) \quad \forall \alpha \tag{32}$$

Note that this in particular implies that $Beer_1 \implies \neg P^{\beta}_{A_W} NF_Q$ for $\beta > \frac{1}{2}$. From 31 and 32 we have

$$Beer_1 \implies (P^{\alpha}_{A_S} NF_Q \implies P^{\frac{1}{2}+\alpha}_{A_S} NF_B) \quad \forall \alpha \tag{33}$$

but since $P^{0}_{A_S} NF_Q$ is a theorem we have that $Beer_1 \implies P^{\frac{1}{2}}_{A_S} NF_B$ as well as $Beer_1 \implies \neg P^{\beta}_{A_S} NF_Q$ for $\beta > \frac{1}{2}$. We conclude that

$$Beer_1 \implies u^{1.5}_{A_S} Beer_2 \wedge \neg u^{1.5}_{A_S} Quiche_2 \tag{34}$$

since the expected utility to A_S is at least 2 from choosing $Beer_2$ by $Beer_1 \implies P^{\frac{1}{2}}_{A_S} NF_B$ and the expected utility is no more than 1 when A_S chooses $Quiche_2$ according to $Beer_1 \implies \neg P^{\beta}_{A_S} NF_Q$ for $\beta > \frac{1}{2}$. From 34 and R_{A_S} (the rationality of A_S) we have that

$$Beer_1 \implies Beer_2 \tag{35}$$

We have just deduced that the statements in the epistemic form of the game together with common confidence of identical conjectures for Alice and the rationality of Alice imply that Alice's strong type chooses beer whenever the weak type does.

Now assume that we add the statements that correspond to CR, i.e., that all identities are confident in every identity's rationality. Since we added common confidence in Alice's identical conjectures we have that B_B (representing both B_{B_1} and B_{B_2}) is confident in each and every formula in 31. Since he is confident in the rationality of both A_W and A_S we can apply axiom K and deduce that he is confident in the statements in 32, 33 and 34 and hence in 35. This follows from the fact that we can use a finite number of statements from the list in every stage of the deduction[32]. We conclude that

$$C_{B_B}(Beer_1 \Longrightarrow Beer_2) \tag{36}$$

By the epistemic form of the game we have

$$C_{B_B}((W \wedge Beer_1) \vee (S \wedge Beer_2)) \tag{37}$$

which together with 36 imply

$$C_{B_B}(Beer_2) \tag{38}$$

By the description of nature moves in the epistemic form we have

$$P^{\alpha}_{B_B}(W \wedge Beer_1 \wedge Beer_2) \Longrightarrow P^{9\alpha}_{B_B}(S \wedge Beer_1 \wedge Beer_2) \quad \forall \alpha \tag{39}$$

38 implies that $P^{\alpha}_{B_B}(W \wedge Beer_1 \wedge Beer_2) \iff P^{\alpha}_{B_B}(W \wedge Beer_1)$ and $P^{9\alpha}_{B_B}(S \wedge Beer_1 \wedge Beer_2) \Longrightarrow P^{9\alpha}_{B_B}(S \wedge Beer_2)$ since $S \wedge Beer_1 \wedge Beer_2 \Longrightarrow S \wedge Beer_2$ and so we have $P^{\alpha}_{B_B}(W \wedge Beer_1) \Longrightarrow P^{9\alpha}_{B_B}(S \wedge Beer_2)$ which together with 37 implies

$$P^{.9}_{B_B}(S \wedge Beer_2) \tag{40}$$

the rationality of B_B and 40 imply that B_B chooses NF_B.

We now consider adding C^2R. Since A_S is now confident that B_B is confident that she is rational, CR implies that she is confident that B_B is rational and common confidence implies that she is confident that B_B

[32] In particular we can only consider a finite, yet dense enough, collection of probabilities α, e.g. all the multiples of $\frac{1}{100}$ will suffice.

is confident that Alice has identical conjectures, we can show that A_S is confident in 36 – 40 and hence is confident of NF_B. By A_S's rationality (R) we have that this implies $Beer_2$. From C^3R in addition to C^2R, CR and common confidence of Alice's identical conjectures, we can deduce that B_Q is confident that A_S can make the previous deduction. From CR he is also confident that she is rational and hence deduces that she chooses $Beer_2$. We have that $C_{B_Q}Beer_2$. By the epistemic form this implies $C_{B_Q}Quiche_1$ and hence the rationality of B_Q now implies F_{B_Q}. If we now add C^4R we have that A_W can make this deduction and hence $C_{A_W}(NF_{B_B} \wedge F_{B_Q})$ hence her rationality implies $Beer_1$.

We conclude that common confidence of identical conjectures for Alice, R, CR, C^2R, C^3R and C^4R logically imply $Beer_1$, $Beer_2$, NF_{B_B} and F_{B_Q} which is exactly the intuitive behavior (on *and* off the equilibrium path). Obviously we cannot add C^5R to the previous list of statements since this will imply that B_Q can deduce that A_W chooses $Beer_1$ in addition to $Beer_2$ being chosen which contradicts the epistemic form.

Finally, showing that common confidence of identical conjectures for Alice and R, CR, C^2R, C^3R and C^4R are consistent together with the epistemic form can be done by providing a model for this collection of statements. We leave this construction to the reader. ∎

10 Appendix C

Proof of Proposition 17. We first note that if Alice's conjecture (both A_1 and A_2 having identical conjectures) about Bob's behavior is that (U, M, D) are chosen with probabilities $\mu_1 = (\frac{3}{5}, 0, \frac{2}{5})$ respectively, then the rationality of A_1 and A_2 implies that In_1 and Out_2 are chosen. If Alice's conjecture is $\mu_2 = (0, \frac{5}{7}, \frac{2}{7})$ then rationality implies Out_1 and In_2. Assume that Bob is confident that A_1 and A_2 have the same conjecture and that they are rational then we have the following:

If Bob is confident that Alice's conjecture is μ_1 then Bob's rationality implies U.

If Bob is confident that Alice's conjecture is μ_2 then Bob's rationality implies D.

If Bob assign equal probability to Alice's conjecture being either μ_1 or μ_2 then Bob's rationality implies M.

We have just concluded that each of his pure strategies is a best response for a specific conjecture that Bob has about Alice's behavior as derived from his conjecture about her conjectures about his behavior and Bob's confidence in Alice's rationality. This allows for Alice to holds any conjecture – in particular μ_1 and μ_2 – as it is supported on rational responses by Bob. Furthermore, any combination of choices by A_1 and A_2 is rational given a corresponding conjecture about Bob's behavior.

Consider the following model. Let the state space be $\Omega = \{Alice, Bob\} \times \{One, Two\} \times \{In_1, Out_1\} \times \{In_2, Out_2\} \times \{U, M, D\}$ and the relations:

$$(.,.,In_1, In_2,.)R_{A_1}(Alice, One, In_1, In_2,.)$$

$$(.,.,In_1, Out_2,.)R_{A_1}(Alice, One, In_1, Out_2,.)$$

$$(.,.,Out_1, In_2,.)R_{A_1}(Alice, One, Out_1, In_2,.)$$

$$(.,.,Out_1, Out_2,.)R_{A_1}(Alice, One, Out_1, Out_2,.)$$

$$(.,.,In_1, In_2,.)R_{A_2}(Alice, Two, In_1, In_2,.)$$

$$(.,.,In_1, Out_2,.)R_{A_2}(Alice, Two, In_1, Out_2,.)$$

$$(.,.,Out_1, In_2,.)R_{A_2}(Alice, Two, Out_1, In_2,.)$$

$$(.,.,Out_1, Out_2,.)R_{A_2}(Alice, Two, Out_1, Out_2,.)$$

$$(.,.,.,.,U)R_B(Bob,.,In_1, In_2, U)$$

$$(.,.,.,.,U)R_B(Bob, One, In_1, Out_2, U)$$

$$(.,.,.,.,U)R_B(Bob, Two, Out_1, In_2, U)$$

$$(.,.,.,.,.,M)R_B(Bob,.,In_1,In_2,M)$$

$$(.,.,.,.,.,M)R_B(Bob,One,In_1,Out_2,M)$$

$$(.,.,.,.,.,M)R_B(Bob,Two,Out_1,In_2,M)$$

$$(.,.,.,.,.,D)R_B(Bob,.,In_1,In_2,D)$$

$$(.,.,.,.,.,D)R_B(Bob,One,In_1,Out_2,D)$$

$$(.,.,.,.,.,D)R_B(Bob,Two,Out_1,In_2,D)$$

where we identify B_1 and B_2 for simplicity. We assign the following probability distributions:

$$T_{A_i}((.,.,In_1,In_2,.))(\{(Alice,i,In_1,In_2,D)\}) = 1$$

$$T_{A_i}((.,.,In_1,Out_2,.))(\{(Alice,i,In_1,In_2,U)\}) = \frac{3}{5}$$

$$T_{A_i}((.,.,In_1,Out_2,.))(\{(Alice,i,In_1,In_2,D)\}) = \frac{2}{5}$$

$$T_{A_i}((.,.,Out_1,Out_2,.))(\{(Alice,i,Out_1,Out_2,M)\}) = 1$$

$$T_{A_i}((.,.,Out_1,In_2,.))(\{(Alice,i,In_1,In_2,M)\}) = \frac{5}{7}$$

$$T_{A_i}((.,.,Out_1,In_2,.))(\{(Alice,i,In_1,In_2,D)\}) = \frac{2}{7}$$

$$T_B((.,.,.,.,U))(\{(Bob,One,In_1,Out_2,U)\}) = 1$$

$$T_B((.,.,.,.,M))(\{(Bob, One, In_1, Out_2, M)\}) = \frac{1}{2}$$

$$T_B((.,.,.,.,M))(\{(Bob, Two, Out_1, In_2, M)\}) = \frac{1}{2}$$

$$T_B((.,.,.,.,D))(\{(Bob, Two, Out_1, In_2, D)\}) = 1$$

where i stands for 1 (respectively One) and 2 (respectively Two) since both types have the same conjectures about Bob's behavior. It is easy to check that this is a model for the epistemic form of the game[33]. By definition Alice has identical conjectures at every state hence we have that common confidence that Alice has identical conjectures holds at every state. It also follows from the previous discussion that all identities are rational at every state of the world, hence we have CCR at every state. Since every pure strategy profile has a state in which it holds we have that all strategies are consistent with CCR, common confidence of identical conjectures and the epistemic form, and in particular so is the strategy profile (Out_1, Out_2, M) as required. ■

Proof of Theorem 20. Will prove that the epistemic form, CCR, CC(identical conjectures) and CC(mutual certainty about conjectures) imply (In_1, In_2, D) we leave the proof of consistency to the reader. Consider Alice's possible conjectures about Bob. Let α be a probability distribution that Alice might assign to Bob's actions. We also denote the statement defining these probabilities by $\alpha = P_A^{\alpha(U)}U \wedge P_A^{\alpha(M)}M \wedge P_A^{\alpha(D)}D$ and hence $C_B\alpha$ denotes the statement that correspond to Bob being confident that $\alpha(U), \alpha(M)$, $\alpha(D)$ are Alice's beliefs about his behavior. Given Bob's conjecture about the behavior of Alice – A_1 and A_2 – we will consider the derived probabilities that Bob assigns to the events One and Two and denote these by β and $1-\beta$ respectively. let $\beta = P_B^{\beta}(One) \wedge P_B^{1-\beta}(Two)$ be the corresponding event. We note that if Alice is confident that Bob has a specific conjecture about here behavior then there exists a probability β as above such that $C_A\beta$ (where A stands for both A_1 and A_2). Throughout we assume common confidence of identical conjectures for Alice. For every conjecture α of Alice as to Bob's behavior we have the following

[33]Note, for example, that the distribution over nature moves is vacuously preserved since $T_B((.,.,.,.,a))(\{(Bob,.,In_1,In_2,a)\}) = 0$ for every action $a \in \{U, M, D\}$.

$$R_{A_1} \text{ when } \alpha_U + \alpha_M < 2\alpha_D \text{ implies } In_1 \tag{41}$$

$$R_{A_1} \text{ when } \alpha_U + \alpha_M > 2\alpha_D \text{ implies } Out_1 \tag{42}$$

$$R_{A_2} \text{ when } 3\alpha_U + \alpha_M < 3\alpha_D \text{ implies } In_2 \tag{43}$$

$$R_{A_2} \text{ when } 3\alpha_U + \alpha_M > 3\alpha_D \text{ implies } Out_2 \tag{44}$$

If Bob has no uncertainties about Alice's conjecture we have the following:

$$C_B R_{A_1} \text{ and } C_B\alpha \text{ when } \alpha_U + \alpha_M < 2\alpha_D \text{ imply } C_B In_1 \tag{45}$$

$$C_B R_{A_1} \text{ and } C_B\alpha \text{ when } \alpha_U + \alpha_M > 2\alpha_D \text{ imply } C_B Out_1 \tag{46}$$

$$C_B R_{A_2} \text{ and } C_B\alpha \text{ when } 3\alpha_U + \alpha_M < 3\alpha_D \text{ imply } C_B In_2 \tag{47}$$

$$C_B R_{A_2} \text{ and } C_B\alpha \text{ when } 3\alpha_U + \alpha_M > 3\alpha_D \text{ imply } C_B Out_2 \tag{48}$$

since the epistemic form implies $C_B((One \wedge In_1) \vee (Two \wedge In_2))$ we have

$$C_B R_{A_1}, C_B R_{A_2} \text{and } C_B\alpha \text{ imply } \alpha_U + \alpha_M \leq 2\alpha_D \text{ or } 3\alpha_U + \alpha_M \leq 3\alpha_D \tag{49}$$

Let β be the probability distribution that Bob assigns to *One* and *Two* as induced from a conjecture about Alice's behavior, we have that

$$R_B \text{ and } \beta \geq \frac{1}{2} \text{ imply } \neg D \tag{50}$$

$$R_B \text{ and } \beta \leq \frac{1}{2} \text{ imply } \neg U \tag{51}$$

from 50 and 51 we have for all β

$$C_A\beta \text{ and } C_A R_B \text{ imply } C_A\neg D \vee C_A\neg U \tag{52}$$

since Bob is confident that Alice has no uncertainties about his conjectures we have that

$$C_B C_A R_B \text{ implies } C_B(C_A \neg D \vee C_A \neg U) \tag{53}$$

from 53 we have

$$C_B\alpha \text{ and } C_B C_A R_B \text{ imply that either } \alpha(D) = 0 \text{ or } \alpha(U) = 0$$

however $\alpha(D) = 0$ and the rationality of Alice implies Out_1 and Out_2. Hence $C_B R$ implies that $\alpha(D) > 0$ and we have that Bob being confident that Alice is certain about his conjecture yields the following

$$\text{if } C_B\alpha,\ C_B R_A \text{ and } C_B C_A R_B \text{ then } \alpha(U) = 0 \tag{54}$$

we also have that

$$P_A^1 \neg U \text{ and } R_A \text{ imply that } Out_2 \Longrightarrow Out_1 \tag{55}$$

since the rationality of A_2 implies that if $\alpha(U) = 0$ we must have $\alpha(M) \geq 3\alpha(D)$ or $\alpha(M) \geq \frac{3}{4}$ which implies Out_1 by the identical conjecture assumption and the rationality of A_2. From 54 and 55 we have that CC(mutual certainty about conjectures), CCR, CC(Alice has identical conjectures) imply

$$C_B(Out_2 \Longrightarrow Out_1) \tag{56}$$

56 and the epistemic form imply $C_B(In_2)$ which implies $P_B^{.9}(Two \wedge In_2)$ from the definition of nature moves. Together with R_B we now have that Bob will choose D. The confidence of Alice in each of $C_B\alpha$, $C_B R_A$ and $C_B C_A R_B$ as well as her confidence that Bob is confident that she is certain about his conjectures imply $C_A D$. Together with her rationality we now have $C_A D$ implies In_1 and In_2 as required. ■

11 Appendix D

Proof of Theorem 21. Consider the epistemic form of the "Burning Money" game as given in Section 7.3. For the proof of result (1) it suffices to construct a model for the epistemic form where identities are rational at every state and there exists a state such that (NB, L_2, l_2) does not hold. Let $\Omega = \{Alice, Bob\} \times \{NB, Burn\} \times \{L_1, R_1\} \times \{L_2, R_2\} \times \{l_1, r_1\} \times \{l_2, r_2\}$ we refer to B_1 (resp. B_2) as representing B_{1_1} and B_{1_2} (resp. B_{2_1} and B_{2_2}). Consider the relations:

$$(., NB, ., ., ., .)R_{A_0}(Alice, NB, L_1, L_2, l_1, l_2) \tag{57}$$

$$(., Burn, ., ., ., .)R_{A_0}(Alice, Burn, L_1, R_2, l_1, r_2) \tag{58}$$

$$(., ., L_1, ., ., .)R_{A_1}(Alice, Burn, L_1, L_2, l_1, l_2) \tag{59}$$

$$(., ., R_1, ., ., .)R_{A_1}(Alice, Burn, R_1, L_2, r_1, l_2) \tag{60}$$

$$(., ., ., L_2, ., .)R_{A_2}(Alice, NB, L_1, L_2, l_1, l_2) \tag{61}$$

$$(., ., ., R_2, ., .)R_{A_2}(Alice, NB, L_1, R_2, l_1, r_2) \tag{62}$$

$$(., ., ., ., l_1, .)R_{B_1}(Bob, Burn, L_1, L_2, l_1, l_2) \tag{63}$$

$$(., ., ., ., r_1, .)R_{B_1}(Bob, Burn, R_1, L_2, r_1, l_2) \tag{64}$$

$$(., ., ., , ., ., l_2)R_{B_2}(Bob, NB, L_1, L_2, l_1, l_2) \tag{65}$$

$$(., ., ., , ., ., r_2)R_{B_2}(Bob, NB, L_1, R_2, l_1, r_2) \tag{66}$$

since at every state in Ω all identities consider only one state to be possible we have that their probability distributions are trivially defined. Similarly, the extension with utilities follows from assigning the utility of outcomes corresponding to the states the identities find possible. We define $V : \Omega \times \mathcal{F} \to \{0, 1\}$ by setting $V(\omega, f) = 1$ if and only if the corresponding coordinate of ω if f for every atomic statement $f \in \{Alice, Bob, NB, Burn, L_1, R_1, L_2, R_2, l_1, r_1, l_2, r_2\}$, we then extend V to the set of all statements $\mathcal{F}$ in the language using 8 – 10. We notice that 25 – 27 are satisfied, i.e., the relations are non-empty, transitive and if ω is a state that is viewed as possible (from some state) then at ω the state ω itself is viewed as possible. From the definition of the relations we have that at every state each identity is confident of their own action in that state. Hence, we have common confidence that every identity os confident of an action iff the action holds. Similarly, all other statements in the epistemic form of the "burning money" game holds[34].

We need to show that there is common confidence of rationality. This follows from showing that at every state every identity is rational. Consider A_0, 57 implies that whenever NB holds in a state ω we have that A_0 is confident that L_1, L_2, l_1 and l_2 will be chosen, hence NB is a best response. For a state where $Burn$ holds, she is confident that L_1, R_2, l_1, r_2 occurs according to 58, which implies that $Burn$ is the rational thing to choose. Since at every ω either $Burn$ or NB hold we find that A_0 is rational at every ω. IN the same manner 59 – 60 imply the rationality of A_1 at every state, 61 – 62 the rationality of A_2 and 63 – 66 imply the rationality of Bob's identities. Since every strategy profile holds at some state of the world in this model we find that common confidence of rationality is consistent with the epistemic form and does not restrict behavior in this game. In particular, common confidence of rationality need not imply (NB, L_2, l_2).

In order to prove (2) we must show that the epistemic form is inconsistent with common confidence of rationality and common confidence that A_0, A_1 and A_2 have identical conjectures. Assume by way of contradiction that these assumptions hold. Let α_1 denote the probability that A_0 assigns to l_1 and α_2 the probability she assigns to l_2. By common confidence of joint conjectures these are the probabilities that A_1 and A_2 assign as well. Since A_0, A_1 and A_2 are confident and agree in their conjectures, and since A_0 is confident that A_1 is rational and A_1 *is* rational we have

$$u^q_{A_0} Burn \iff u^q_{A_1} Burn \text{ for all } q \tag{67}$$

similarly we have that A_0's confidence in A_2's rationality together with A_2's rationality imply

$$u^q_{A_0} NB \iff u^q_{A_2} NB \text{ for all } q \tag{68}$$

Since A_1 and A_2 are confident that A_0 is confident that they are rational, they are confident in 67 and 68 as well. This implies that

$$C_{A_1}(u^q_{A_0} Burn) \iff u^q_{A_0} Burn \text{ for all } q \tag{69}$$

[34]For example, the dynamic structure holds since A_1 and B_1 are confident that B holds, while A_2 and B_2 are confident that NB holds.

$$C_{A_2}(u^q_{A_0}NB) \iff u^q_{A_0}NB \text{ for all } q \tag{70}$$

Since A_1 and A_2 are confident in each others rationality and are confident in each other's conjectures we also can deduce

$$C_{A_2}(u^q_{A_0}Burn) \iff u^q_{A_0}Burn \text{ for all } q \tag{71}$$

$$C_{A_1}(u^q_{A_0}NB) \iff u^q_{A_0}NB \text{ for all } q \tag{72}$$

A_1 confidence in A_0's rationality implies

$$C_{A_1}(u^q_{A_0}NB \Longrightarrow u^q_{A_0}Burn) \text{ for all } q \tag{73}$$

since A_1 is confident that A_0 chooses $Burn$. Similarly we get

$$C_{A_2}(u^q_{A_0}Burn \Longrightarrow u^q_{A_0}NB) \text{ for all } q \tag{74}$$

Since B_1 is confident of all the assumptions we used leading to 67 – 74 we have that he is confident in each one of them. Since 69, 72 and 73 imply $u^q_{A_0}NB \Longrightarrow u^q_{A_0}Burn$ we have

$$C_{B_1}(u^q_{A_0}NB \Longrightarrow u^q_{A_0}Burn) \text{ for all } q \tag{75}$$

and since B_1 is confident in 70, 71 and 74 we have

$$C_{B_1}(u^q_{A_0}NB \iff u^q_{A_0}Burn) \text{ for all } q \tag{76}$$

hence

$$C_{B_1}(u^q_{A_2}NB \iff u^q_{A_1}Burn) \text{ for all } q \tag{77}$$

We also have $C_{B_1}u^4_{A_2}NB$ since $u^{4\alpha_2+6(1-\alpha_2)}_{A_2}R_2$ from the epistemic form (no matter what α_2 is) and B_1 is confident that A_2 is rational. From 77 we get

$$C_{B_1}u^4_{A_1}Burn \tag{78}$$

the epistemic form implies that

$$\neg u^4_{A_1}R_1 \tag{79}$$

since B_1 is confident that A_1 is rational and that 79 holds, together with 78 we have

$$C_{B_1}L_1 \tag{80}$$

hence the rationality of B_1 implies

$$l_1 \tag{81}$$

each of the other identities is confident that B_1 is confident in all of the assumptions that led to 80 and they are confident in his rationality, hence they are all confident in 81 and we have

$$L_1 \text{ by } A_1\text{'s rationality and } U^{6.5}_{A_1}Burn \tag{82}$$

Applying the deduction leading to 77 for B_2 we get

$$C_{B_2}(u^q_{A_2}NB \iff u^q_{A_1}Burn) \text{ for all } q \tag{83}$$

confidence of B_2 in 82 and 83 imply

$$C_{B_2}u^{6.5}_{A_2}NB$$

hence $C_{B_2}\neg R_2$ and the rationality of B_2 implies l_2 which together with confidence of A_2 in the confidence of B_2 in the assumptions leading to the deductions for B_2 imply that $C_{A_2}l_2$ and the rationality of A_2 implies $U^9_{A_2}NB$ contradicting 82 and 77 and the desired contradiction is achieved.

For (3) we denote by $\mathcal{R}$ the set of statements that define the rationality of all identity. We begin by extracting the maximal consistent level of confidence in rationality from the proof of (2). Let α_1 and α_2 be the identical conjecture that A_0, A_1 and A_2 have about the actions l_1 and l_2 respectively. We assume common confidence of identical conjectures for Alice. We repeat the main steps in the proof of (2). Adding $C_{A_0}\mathcal{R}$ implies that $U^s_{A_0}Burn$ for $s = Max\{6.5\alpha_1 - 2.5(1-\alpha_1), 1.5\alpha_1 + 3.5(1-\alpha_1)\}$ (from confidence of A_0 in the rationality of A_1) and $U^t_{A_0}NB$ for $t = Max\{9\alpha_2, 4\alpha_2 + 6(1-\alpha_2)\}$. $C_{A_1}\mathcal{R}$ and $C_{A_1}C_{A_0}\mathcal{R}$ implies $s \geq t$ and together with $C_{A_2}\mathcal{R}$ and $C_{A_2}C_{A_0}\mathcal{R}$ we get $s = t$. $C_{B_1}\mathcal{R}$ and $C_{B_1}C_{A_0}\mathcal{R}$ and implies $C_{B_1}L_1$ hence his rationality implies l_1. $C_{A_0}\mathcal{R}$, $C_{A_0}C_{B_1}\mathcal{R}$ and $C_{A_0}C_{B_1}C_{A_0}\mathcal{R}$ implies $\alpha_1 = 1$ and $s = 6.5$. Since $t = 6.5$ we have that $\alpha_2 = \frac{6.5}{9}$. From the rationality of A_2 we now get L_2. Furthermore, $C_{B_2}\mathcal{R}$, $C_{B_2}C_{A_0}\mathcal{R}$, $C_{B_2}C_{A_0}C_{B_1}\mathcal{R}$ and $C_{B_2}C_{A_0}C_{B_1}C_{A_0}\mathcal{R}$ imply that l_2. The contradiction appeared at the fifth level of mutual confidence of rationality, since $C_{A_2}\mathcal{R}$, $C_{A_2}C_{B_2}\mathcal{R}$, $C_{A_2}C_{B_2}C_{A_0}\mathcal{R}$, $C_{A_2}C_{B_2}C_{A_0}C_{B_1}\mathcal{R}$ and $C_{A_2}C_{B_2}C_{A_0}C_{B_1}C_{A_0}\mathcal{R}$ imply that $C_{A_2}l_2$ in contradiction with $\alpha_2 < 1$. What we find is that the fourth level of mutual confidence in rationality implies that l_2 and L_2 obtain. However, it is not necessarily the case that NB obtains. Since we have that $\alpha_2 = \frac{6.5}{9}$ the event $Burn$ cannot be eliminated, and the outcome $(Burn, L_1, l_1)$ is consistent with the fourth level of iteration of confidence of rationality. Note that it is A_0's indifference between $Burn$ and NB that sustains this outcome. ■

Proof of Proposition 22. The proof closely follows the verbal description provided in Section 2. Let α_1 denote the probability that A_0 assigns to l_1 and α_2 the probability she assigns to l_2. Condition (1) states that the is common confidence of the following statements

$$P^{\alpha}_{A_0}l_2 \iff P^{\alpha}_{A_1}l_2 \iff P^{\alpha}_{A_2}l_2 \ \forall\alpha \tag{84}$$

$$P^{\alpha}_{A_0}l_1 \iff P^{\alpha}_{A_1}l_1 \iff P^{\alpha}_{A_2}l_1 \ \forall\alpha \tag{85}$$

from conditions (2) and (3) we find that

$$\neg P^{\frac{6.5}{9}}_{A_0}l_1 \Longrightarrow NB \tag{86}$$

since the confidence of A_0 in A_2's rationality implies that she expects at least a payoff of 4 from choosing NB. Since A_0 is rational and is confident that A_1 is rational she will conceivably choose $Burn$ only if she expects at least a payoff of 4, this implies that she assigns at least a probability of $\frac{6.5}{9}$ to l_1. From (5) we have confidence of B_1 in (2) and (3) which implies that he is confident in 86, since we have $C_{B_1}\neg NB$ from the epistemic form, we conclude

$$C_{B_1}P^{\frac{6.5}{9}}_{A_0}l_1 \tag{87}$$

Since B_1 is confident in 85 and the rationality of A_1 we have that 87 implies that he is confident in L_1. With the rationality of B_1 we have that this implies l_1. From (6) we have that A_0 is confident in each of these steps, hence

$$u^{6.5}_{A_0}Burn \tag{88}$$

from (7) B_2 is confident of 88 and since he is confident that A_0 and A_2 are rational and that A_0 is confident that A_2 is rational and that they have the same conjecture we find that

$$C_{B_2} P_{A_0}^{\frac{6.5}{9}} l_2 \tag{89}$$

B_2's rationality implies now l_2. From (8) we find that A_0 is confident of l_2, i.e., that $\alpha_2 = 1$. The rationality of A_2 implies L_2 and the rationality of A_0 together with the confidence of A_0 in the rationality of A_2 implies NB as required. ■