CHAPTER 5

Agreement and Equilibrium with Minimal Introspection

Abstract. Standard models in epistemic game theory make strong assumptions about agents' knowledge of their own beliefs. Agents are typically assumed to be *introspectively omniscient*: if an agent believes an event with probability p, she is certain that she believes it with probability p. This paper investigates the extent to which this assumption can be relaxed while preserving some standard epistemic results. Geanakoplos (1989) claims to provide an Agreement Theorem using the "truth" axiom, together with the property of *balancedness*, a significant relaxation of introspective omniscience. I provide an example which shows that Geanakoplos's statement is incorrect. I then introduce a new property, *local balancedness*, which allows us both to correct Geanakoplos's result, and to extend it to cases where the truth axiom may fail. I exploit this general Agreement Theorem to provide novel epistemic conditions for correlated and Nash equilibrium, both of which relax the assumption of introspective omniscience. In all three cases, the results are also extended to infinite state spaces.

JEL Classification Number: C70, C72; D82

Keywords: Agreement Theorem, Nash Equilibrium, Correlated Equilibrium, common prior assumption, epistemic game theory, interactive epistemology, bounded rationality.

5.1. Introduction

5.1.1. Motivation. People often don't know what they want.[1] They don't know what they want for dinner, they don't know who they want to win the game, they don't know where they'd prefer to go on vacation, and so on. The simplest way of modeling this uncertainty is to take it at face value: agents are uncertain about their preferences.

This kind of uncertainty is not just pervasive; in many cases it seems intuitively rational. Learning one's own preferences can require considerable time, effort, and cost. People who try many varieties of wine, for example, expend effort and resources to learn not just about wine, but also about their own preferences for different wines. Similarly,

[1]Thanks to Frank Arntzenius, Amanda Friedenberg, Peter Fritz, Stephen Morris, Sujoy Mukerji and Timothy Williamson for discussion of aspects of this material.

it takes investors a great deal of experience to learn their own levels of risk tolerance. In a range of areas, "you can't know what you like until you try it". But if the cost of determining one's own preferences over a given choice set is greater than the expected value derived from knowing those preferences, it may be rational to remain uncertain of them.

In many cases, it is natural to model agents' uncertainty about their own preferences as uncertainty about their utility function, and this has been the usual approach in economics. But in other cases, agents' uncertainty about their preferences seems better represented as uncertainty about their beliefs. Agents are often unable to draw fine distinctions in their hypothetical choice behavior. For example, most people don't know at what odds they would bet on Hilary Clinton's election to the presidency. While this fact can be modeled by agents' uncertainty of their own utilities, a more natural strategy is to represent these people as uncertain about which probability distribution represents their preferences. As before, being uncertain of this distribution can be intuitively rational. Since it can take a great deal of time and effort to imagine bets with sufficient clarity to predict one's own choice behavior, there may be cases where it is efficient to remain uncertain of one's own probabilistic beliefs.

This uncertainty about one's own preferences—and the resulting uncertainty about one's own beliefs—can be rational in a formal sense as well. Agents who satisfy Savage's axioms on preferences may still be uncertain of their own preferences, and thus of their own beliefs.[2] As a matter of fact, even agents who are rational in Savage's sense may still be uncertain whether they satisfy Savage's axioms at all. In an important paper, Stephen Morris (1996) states additional axioms which can be added to Savage's to ensure that agents *are* certain of their beliefs. But Morris himself concludes that the new axioms on preferences are unpersuasive from a normative standpoint.[3] Even rational agents may fail to know their own beliefs.

Standard epistemic models abstract from the phenomenon of agents' uncertainty about their own beliefs. In Harsanyi type structures, for example, agents are generally assumed to "know" their own type. Since each type is associated with a single probability distribution, by knowing his or her own type, each agent knows his or her own distribution. Let an agent be *introspectively omniscient* just in case, if the agent believes an event

[2]For this formal observation, see Morris 1996 and Epstein and Wang 1996: 1351. Savage himself also made some interesting remarks on the conceptual side of the issue (1967: 308).
[3]Morris 1996: 22.

with probability p, she is certain of believing that event with probability p.[4] In standard Harsanyi type structures, then, agents are introspectively omniscient. Moreover, since every type of every player satisfies this condition, the players in the model have common belief that each of them is introspectively omniscient.

Introspective omniscience implies two more familiar "introspection axioms" on certainty. An agent is *positively introspective* if, if she is certain of an event, she is certain of being certain of that event. She is *negatively introspective* if, if she is not certain of an event, she is certain of not being certain of it. Negative and positive introspection together are still weaker than introspective omniscience: an agent may satisfy both of these conditions but still fail to be certain what probability p (where p is less than 1) she assigns to a given event.

Uncertainty about preferences can lead to failures of both positive and negative introspection (and thus of introspective omniscience). First, consider an example which merely reports an empirical fact. I am uncertain whether my preferences would be represented as assigning the event of there being a banking crisis in 2008 probability 1 or merely very high probability. If in fact I assign this event probability less than 1, then I fail to be negatively introspective; if in fact I assign it probability 1, I fail to be positively introspective.

But less brute, subtler examples can also be given. As mentioned earlier, even agents who satisfy Savage's axioms on preferences may be uncertain whether they are rational in Savage's sense at all. Someone might reasonably assign positive probability to the event that she herself is ambiguity averse, and so be uncertain whether she is rational in Savage's sense. Such a person would be disposed to bet at some odds that there is no single distribution which represents her preferences. If in fact there is a distribution which represents her preferences, she fails to be positively introspective: she may be certain of some event, but since she is uncertain whether she is represented by a single probability distribution at all, she is not certain of being certain of this event.

These observations raise the question: do standard epistemic analyses survive without the usual, idealized assumptions about introspection? Our first example will show that the simple answer is: "not in full generality". This paper then takes up the task of finding weak sufficient conditions which do guarantee that these standard results hold. As a test case, I will restrict attention to one class of prominent models, in which it is assumed

[4] I will use "is certain of E" and "E is a certainty" to mean an agent "assigns E probability 1" and "E is an event the agent assigns probability 1", respectively.

that agents' interim beliefs are consistent with a "common prior". In this common prior setting, I provide a new agreement theorem and new epistemic conditions for correlated equilibrium and Nash equilibrium.

An earlier literature explored the extent to which the Agreement Theorem survives relaxing some of these axioms, and, in particular, negative introspection.[5] But this work was carried out in the presence of the "truth" axiom—that if an agent is certain of an event, that event obtains. The truth axiom is now recognized to be overly strong: rational agents, too, may be wrong about the state of the world. In this paper, I will be exploring not just failures of introspection axioms, but failures of introspection axioms for agents who also may be wrong about the state of the world.[6]

5.1.2. Example. Aumann (1987) showed that if introspectively omniscient agents update on a common prior, commonly know the game, and are commonly known to be rational, then the prior forms a correlated equilibrium distribution for the game in question. We do not yet have a formal definition of the game, rationality or the common prior, but I will present a simple example somewhat informally, to give a feel for the topic of the paper.

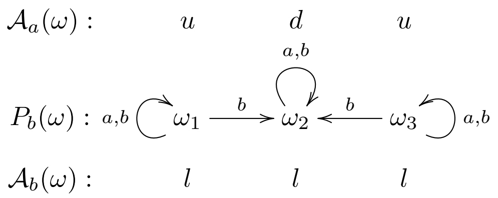
EXAMPLE 5.1.1. In this example, there are three states of the world ($\Omega = \{\omega_1, \omega_2, \omega_3\}$), two players, a and b, and two actions for each player, $A_a = \{u, d\}$; $A_b(\{l, r\}$. At each world, player a considers only that world possible; a always knows the exact state of the world. (Formally, $P_a(\omega_1) = \{\omega_1\}$; $P_a(\omega_2) = \{\omega_2\}$ $P_a(\omega_3) = \{\omega_3\}$.) Player b is not so lucky. If the state of the world is ω_2, then b knows that this is the true state. But if the state of the world is ω_1, player b is uncertain whether the state is ω_1 or ω_2. And if the state is ω_3, player b is uncertain whether the true state is ω_2 or ω_3. (Formally, $P_b(\omega_1) = \{\omega_1, \omega_2\}$; $P_b(\omega_2) = \{\omega_2\}$; $P_b(\omega_3) = \{\omega_2, \omega_3\}$.) The players assign probabilities to events as if they were updating on a common prior μ which assigns $\frac{1}{3}$ to each possible state. So, for example, in ω_1, a assigns $\{\omega_1\}$ probability $1 = \mu(\{\omega_1\}|P_a(\omega_1))$, while b assigns the same event $\frac{1}{2} = \mu(\{\omega_1\}|P_b(\omega_1))$.

[5]Bacharach 1985, Cave 1983, Geanakoplos 1989, Monderer and Samet 1989, Rubinstein and Wolinsky 1990, Samet 1990, Shin 1993, Brandenburger et al. 1992. Two further motivations for studying failures of negative introspection are well known. First, the condition of "plausibility" on unawareness structures (see Dekel et al. 1998) requires a failure of negative introspection. Second, a plausible constraint on the idealization of a rational agent is that the set of sentences of which the agent is certain at least be recursively enumerable. Moreover, it seems plausible that a rational agent could be certain of the theorems of arithmetic. But if such an agent in addition satisfies negative introspection, then it can be shown that the agent is certain of a contradiction: his or her theory of the world is logically inconsistent (McGee 1991 n. 7, Shin and Williamson 1994, Proposition 1).

[6]Agreement Theorems have also been proven with introspective omniscience, but without truth: see, e.g., Bonanno and Nehring 1999 and Hellman 2012. I discuss these results in Section 3.4.

The players' payoffs are the same regardless of the actual state; they are given by the matrix in Figure 5.1.1 where a picks rows and b picks columns.

Players' actions at each state are also represented in Figure 5.1.1 (along with their possibility correspondences). As is easily checked, at each state, each player's actions maximizes subjective expected utility relative to their probabilistic beliefs at that state. (As in this figure, I will use directed graphs throughout the paper to represent simple models. In these figures, directed edges are labelled to indicate that the edge is part of a given player's graph. The *out-neighborhood* of a vertex $\omega \in \Omega$ in a player i's graph is the event i considers possible at ω: in symbols, $P_i(\omega) = N^+_{G_i}(\omega)$.)



	l	r
u	$3,0$	$2,2$
d	$3,3$	$0,0$

The game

FIGURE 5.1.1. Example of Failure of Correlated Equilibrium

The players, as usual, commonly believe the universe $\Omega = \{\omega_1, \omega_2, \omega_3\}$. Thus, since the players' beliefs are consistent with the prior at every world, and they maximize expected utility at every world, they commonly believe that they are rational and that their beliefs are consistent with the prior. By the same reasoning, they also commonly believe that their payoffs are the ones depicted in the matrix.

But the prior induces a distribution over action profiles such that (u, l) is played with probability $\frac{2}{3}$, and (d, l) is played with probability $\frac{1}{3}$. This distribution is not a correlated equilibrium distribution: given a's play, the correlated distribution would require that b play r at least as often as a plays u.

As later definitions will make clear, this result fails even though the players obey the "truth axiom" and also positive introspection.

5.1.3. Interpretation and Outline. As this example illustrates, standard common prior-based results cannot be extended if we allow even slight weakenings of the usual introspection assumptions. The bulk of this paper is dedicated to providing weak conditions on introspection which are sufficient to "resurrect" these standard results. The possibility of providing necessary conditions is discussed in the conclusion.

But these results are best understood negatively. They exhibit the way in which standard common prior-based results rely on introspection assumptions of one form or another. The fact that these common prior-based analyses depend on introspection assumptions can be seen as a conceptual difficulty with these analyses, and with the common prior assumption itself.

The plan of the paper is then as follows. Section 5.2 lays out the basic model. Geanakoplos (1989) "Theorem 6" provides what would be, if it were correct, the most general known Agreement Theorem for interim beliefs. I present a counterexample to Geanakoplos's statement, which motivates the search for a correct, general Agreement Theorem. Section 5.3 introduces the property of local balancedness, an extension of the property used by Geanakoplos (1989). I then provide a new result which corrects Geanakoplos's, extends it to cases where the truth axiom may fail, and also extends it to infinite state spaces. This is the main result of the paper. The rest of the paper exploits this agreement theorem to provide epistemic conditions for correlated equilibrium and Nash equilibrium. Section 5.4 presents two results for correlated equilibrium: one extends the result of Brandenburger et al. (1992); the other has a new form. Section 5.5 presents epistemic conditions for Nash equilibrium which improve those of Bach and Tsakas (2014) by allowing failures of introspection. In each case, the details of the epistemic conditions are somewhat involved. The main payoff is rather that we *can* use the new agreement theorem to give these epistemic conditions.

Section 5.6 concludes; discussion of related work is given at the close of each section.

5.2. The Basic Model: Earlier Use of Balancedness

5.2.1. The Model. A pure epistemic model will be a tuple $\langle I, (\Omega, \mathcal{T}), (p_i)_{i \in I} \rangle$ including a finite set of players $I = 1, 2, ...n$; and a topological space $(\Omega, \mathcal{T})$, which is separable and completely metrizable (it is "Polish"). The Borel algebra generated by the topology is then used to give the measurable state space $(\Omega, \mathcal{B})$, where elements of $\mathcal{B}$ are *events*. In the sequel, I will often omit reference to $\mathcal{B}$ and $\mathcal{T}$; when the algebra is clear from context I write simply Ω. In Sections 2 and 3, "model" will mean a pure epistemic model; in Sections 4 and 5, a "model" will be a game model.

For a Polish space such as Ω, let $\Delta(\Omega)$ be the set of probability measures on Ω, where $\Delta(\Omega)$ itself is endowed with the weak topology. The *belief maps* $p_i : \Omega \to \Delta(\Omega)$ represent the players' probabilistic beliefs at elements of Ω; these too are required to be measurable (where the algebra on $\Delta(\Omega)$ is taken to be the Borel algebra of the weak topology). A player i is said to be certain of an event $E \in \mathcal{B}$ at a state ω if and only if $p_i(\omega)(E) = 1$.

Now for a measure μ on Ω, let *supp*(μ) denote the support of μ, $\{\omega : \omega \in N_\omega \in \mathcal{T} \Rightarrow \mu(N_\omega > 0)\}$, that is, the set of states such that every open neighborhood of every state has positive measure. We use this idea to define *possibility correspondences* from the belief maps, as follows: $P_i(\omega) = supp(p_i(\omega))$. A player i is then said to *believe* an event $E \in \mathcal{B}$ at a state $\omega \in \Omega$ if and only if $P_i(\omega) \subseteq E$. It will be useful also to have *belief functions* $B_i : \mathcal{B} \to \mathcal{B}$, defined in terms of the possibility correspondences: $B_i(E) := \{\omega : P_i(\omega) \subseteq E\}$. A player i's belief function takes each event E to the event of i's believing E. Note that these definitions allow that a player may be certain of an event which he or she does not believe.

Later, we will describe events in terms of elements of the model. For example, for some $p^* \in \Delta(\Omega)$, we may wish to speak of the event that i's probabilistic beliefs are equal to p^*. To do this, I will use square brackets as follows: $[p_i = p^*]$ will denote the event of i's probabilistic beliefs being equal to the distribution p^*, that is, the event $\{\omega : p_i(\omega) = p^*\}$. Similarly, if E is an event and $k \in [0, 1]$, I will write $[p_i(E) = k]$ for the event that i assigns E probability k: $\{\omega : p_i(\omega)(E) = k\}$.

Introspective omniscience is defined as follows:

Introspective Omniscience: $\forall \omega \in \Omega,\ P_i(\omega) \subseteq \{\omega' : p_i(\omega') = p_i(\omega) \wedge P_i(\omega') = P_i(\omega)\}$

As noted in the introduction, introspective omniscience implies the following two conditions:

Positive Introspection: $\forall \omega, \omega'\ \omega' \in P_i(\omega) \to (P_i(\omega') \subseteq P_i(\omega))$ (also called (4))
Negative Introspection: $\forall \omega, \omega'\ \omega' \in P_i(\omega) \to P_i(\omega) \subseteq P_i(\omega')$ (also called (5))

These axioms are more familiar when stated using the notion of belief. Positive introspection says that if an agent believes an event, then she believes that she believes it $(B_i(E) \subseteq B_iB_i(E))$. Negative introspection says that if she does not believe an event, she believes that she does not believe it (using $\neg$ for relative complement within Ω, $\neg B_i(E) \subseteq B_i(\neg B_i(E))$). (Note that introspective omniscience also implies the analogous conditions for certainty, but in our setting it is more natural to consider the ones for belief.)

Aumann's (1976) original models were "partitional". Every agent in these models satisfies (T), (4) and (5), where (T) is:

Truth: $\forall \omega\ \omega \in P_i(\omega)$ (also called (T))

(in fact, the conjunction of (T) and (5) entails (4)).

We wish to describe not just one player's beliefs, but the relationship between players' beliefs. Accordingly, we introduce some key interactive notions.

DEFINITION 5.2.1. Fix a model $\mathcal{M}$. An event E is *self-evident to i* if and only if for all $\omega \in E$, $P_i(\omega) \subseteq E$. An event is *public* to a group $G \subseteq I$ if it is self-evident to all $i \in G$.

A group $G \subseteq I$ *commonly believes* an event F at ω just in case there is some event E with E public to G, $\omega \in E$, and, for all $i \in G$, $\forall \omega \in E, P_i(\omega) \subseteq F$.[7] Just as with belief, we define a function $CB_G(F)$ which takes events to the event of their being commonly believed by G.

This paper studies cases in which agents' interim beliefs are consistent with a common prior. We define a prior as follows:

DEFINITION 5.2.2. Fix a model $\mathcal{M}$ and a group $G \subseteq I$. An event $E \in \mathcal{B}$ is *common prior consistent* for a group $G \subseteq I$ if there exists a measure $\mu \in \Delta(\Omega)$ such that

(i) $\mu(E) = 1$;

for every $i \in G$, for every $\omega \in E$,

(iia)$\mu(P_i(\omega)) > 0$; and

(iib) $p_i(\omega) = \mu(\cdot|P_i(\omega))$.

μ is then a *common prior for G over E*. If $G = I$, μ is *a common prior over E*.

Note that this definition is consistent with more recent justifications of the common prior assumption, as a form of consistency among interim beliefs (Bonanno and Nehring 1999; Feinberg 2000). The common prior is not assumed to be unique or to be a "genuine" representation of agents' beliefs at some prior time.

5.2.2. Earlier Definitions of Balancedness; Geanakoplos's Theorem. Geanakoplos (1989) defines his property of "balancedness" for a restricted class of models.

DEFINITION 5.2.3. A model $\mathcal{M}$ is *regular* if:

(i) $|\Omega|$ is finite;

(ii) Ω is common prior consistent for the group I; and,

(iii) for some common prior μ over Ω, for all $\omega \in \Omega$, $\mu(\{\omega\}) > 0$.

[7]Monderer and Samet (1989) prove that in finite spaces, the characterization in terms of *public* events (suggested originally by Aumann 1976, and informally by Lewis 1969: 52-7; cf. also Friedell 1969) is equivalent to the characterization as an infinite intersection: letting $B^1_G(E) = \cap_{i\in G} B_i(E)$ and $B^n_G(E) = \cap_{i\in G} B_i B^{n-1}_G(E) \cap B^{n-1}_B(E)$, $CB_G(E) = \cap_{n\in\mathbb{N}} B^n_G(E)$. Their result holds in our "regular models" (see below, Definition 5.2.3). Kajii and Morris (1997, Theorem 9) prove an analogous result for measure spaces in general, using equivalence classes of measurable sets defined by symmetric difference under a prior μ. Their result applies to a well-behaved subset of the models considered here. But since I am taking the P_i as primitive, and not defined from certainties, I do not need this machinery.

(Note that condition (iii) implies that the algebra $\mathcal{B}$ must contain all singletons.) In regular models, we can replace the p_i with the specification of a prior μ, since together with the P_i, this is enough to determine players' interim probabilistic beliefs.

Balancedness comes in two varieties. In the first instance it is defined with respect to a specific event. The second, general definition is then given with respect to all *self-evident* events. Here is Geanakoplos's definition:

DEFINITION 5.2.4. Fix a regular model $\mathcal{M}$. A possibility correspondence P_i is balanced *with respect to* an event $E \subseteq \Omega$ if and only if there exists some function $\lambda : \{P_i(\omega) | \omega \in \Omega\} \to \mathbb{R}$ such that for every $\omega \in \Omega$, $\sum_{C \in \{P_i(\omega') | \omega' \in \Omega\}} \lambda(C)\chi_C(\omega) = \chi_E(\omega)$.

P_i is then *balanced* if and only if, for every self-evident event E, it is balanced with respect to E.

Note that in the unqualified definition of balancedness, *different* functions λ may be used to "balance" each self-evident event: all that is required is that for each self-evident event, there is *some* such λ.

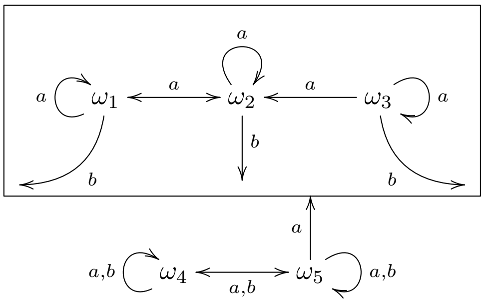
Geanakoplos (1989) claims that agents in regular models agents who are balanced and satisfy (T) cannot agree to disagree ("Theorem 6"):

CLAIM 5.2.5. Let $\mathcal{M}$ be a regular model and $G \subseteq I$ a group where for each $i \in G$, P_i is balanced and satisfies (T). If there is a state $\omega \in \Omega$, an event E and for each $i \in G$, a $k_i \in [0, 1]$ such that it is commonly known that i assigns E probability k_i, then for each i, j, $k_i = k_j$.

The following example shows that this claim is incorrect.

EXAMPLE 5.2.6. Let $\mathcal{M}$ be a model with: $I = \{a, b\}$, $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5\}$.

The model is depicted in Figure 5.2.1. (Note that the directed graph is altered for the sake of simplicity; arrows pointing to the box indicate that an agent considers the worlds in the box possible.) The table describes what players consider possible at each world. Their probabilities at each world ω are consistent with the uniform prior μ, defined so that $\forall \omega \in \Omega$, $\mu(\omega) = \frac{1}{5}$. For this μ, for all $\omega \in \Omega$ and $i \in I$, $p_i(\omega) = \mu(\cdot | P_i(\omega))$. Thus the model is regular.



	$P_a(\omega_n)$	$P_b(\omega_n)$
ω_1	$\{\omega_1, \omega_2\}$	$\{\omega_1, \omega_2, \omega_3\}$
ω_2	$\{\omega_1, \omega_2\}$	$\{\omega_1, \omega_2, \omega_3\}$
ω_3	$\{\omega_2, \omega_3\}$	$\{\omega_1, \omega_2, \omega_3\}$
ω_4	$\{\omega_4, \omega_5\}$	$\{\omega_4, \omega_5\}$
ω_5	Ω	$\{\omega_4, \omega_5\}$

FIGURE 5.2.1. Countermodel to Geanakoplos's Agreement Theorem

It is easily seen that both agents in the example are *balanced.* For b, this is trivial. The only self-evident events are $\{\omega_1, \omega_2, \omega_3\}$, $\{\omega_4, \omega_5\}$ and the universe Ω. The first is balanced by $\lambda(P_b(\omega_1)) = 1$, and for all other admissible E, $\lambda(E) = 0$. The second is balanced by $\lambda(P_b(\omega_4)) = 1$ and for all other admissible E, $\lambda(E) = 0$. For Ω, we let λ be 1 on both $P_b(\omega_1)$ and on $P_b(\omega_4)$.

For a, balancedness is similarly trivial for the self-evident events $\{\omega_1 \omega_2\}$, $\{\omega_2, \omega_3\}$ $\{\omega_4, \omega_5\}$ and Ω. The only mildly tricky case is the event $\{\omega_1, \omega_2, \omega_3\}$. Here we set $\lambda(P_a(\omega_5)) = 1$, $\lambda(P_a(\omega_4) = -1$, and for all other admissible E, $\lambda(E) = 0$.

The event $E = \{\omega_1, \omega_2, \omega_3\}$ is public, and thus a matter of common belief. Consider now the event $F = \{\omega_1, \omega_3\}$. Given the prior μ, it is easy to see that for all $\omega \in E$, $p_a(\omega)(F) = \mu(F|P_a(\omega)) = \frac{1}{2}$. But it is equally easy to see that $\forall \omega \in E$, $P_b(\omega)(F) = \mu(F|E) = \frac{2}{3}$. So it is both common belief that $p_a(F) = \frac{1}{2}$ and $p_b(F) = \frac{2}{3}$. The agents "agree to disagree".

The problem in this example arises because a's correspondence can be balanced only by using sets "outside" of the relevant self-evident event. Geanakoplos's "balancing" sum allows this correspondence to count as balanced because it ranges over the whole set $\{P_i(\omega)|\omega \in \Omega\}$. We need some restriction to rule out the counterexample.

Brandenburger et al. (1992) restrict this sum. They also call their property balancedness, but I will call it *subset balancedness*:

DEFINITION 5.2.7. Fix a regular model $\mathcal{M}$. A possibility correspondence is subset balanced *with respect to* an event $E \subseteq \Omega$ if and only if there exists some function $\lambda : \{P_i(\omega)|\omega \in \Omega\} \to \mathbb{R}$ such that $\forall \omega \in \Omega$, $\sum_{C \in \{P_i(\omega')|P_i(\omega') \subseteq E\}} \lambda(C)\chi_C(\omega) = \chi_E(\omega)$.

As usual it is subset balanced if it is subset balanced with respect to all self-evident events.

Note here that the sum ranges only over elements of $\{P_i(\omega)|P_i(\omega) \subseteq E\}$.

If subset balancedness were used instead of balancedness plus truth (T), then we could prove a version of the Agreement Theorem. But this is only because we require that the correspondence be subset balanced on the whole universe. Later, however, we will want to relax the truth axiom, and to use the Agreement Theorem "locally" within a specific self-evident event, without requiring any "global" properties on the whole universe. But under these conditions, subset balancedness is ill-behaved. To see this, consider a modification of Example 5.2.6 so that $P_a(\omega_5) = \{\omega_1, \omega_2, \omega_3\}$. P_a is now subset balanced with respect to the self-evident event $E = \{\omega_1, \omega_2, \omega_3\}$. Since this event is still self-evident, and since for all $\omega \in E$ the possibility correspondences are unchanged, the same argument as above shows that a and b can agree to disagree in their probabilities for F.[8] The property of local balancedness, introduced in the next section, will avoid this kind of problem.

5.3. Local Balancedness: The Agreement Theorem

5.3.1. Local Balancedness. To give the intuition for local balancedness, I will first define it in the finite case, providing two results which relate it to better-known properties from epistemic models.

Subset balancedness uses a particular restriction on the sum in the definition of balancedness. Local balancedness is a variation on this theme. It will be convenient to have the following notation. The set $\mathcal{C}_E$ will denote the events the agent *could* believe, if E obtains: $\mathcal{C}_E = \{C | \exists\omega(\omega \in E \wedge P_i(\omega) = C)\}$. The elements of $\mathcal{C}_E$ are events "local" to E in the sense that, if E occurs, they might represent the information of the agent in question.

Definition 5.3.1. Fix a regular model $\mathcal{M}$. A possibility correspondence P_i is *locally balanced with respect to* an event $E \subseteq \Omega$ if and only if there exists a function $\lambda : \{P_i(\omega) | \omega \in \Omega\} \to \mathbb{R}$ such that for all $\omega \in \Omega$, $\sum_{C \in \mathcal{C}_E} \lambda(C)\chi_C(\omega) = \chi_E(\omega)$.

A possibility correspondence is locally balanced *within* Ω if it is locally balanced with respect to every self-evident event.

This is not our official definition, since it is restricted to regular models. But the definition for regular models allows us to prove some preliminary results which help to give a sense for the property.

[8] To see that this correspondence is not subset balanced *globally*, consider the self-evident event $\{\omega_1, \omega_2, \omega_3, \omega_5\}$.

First, let $\mathcal{M}$ be a regular model, $i \in I$ a player so that P_i satisfies (T), and an E which is self-evident to i. Then it is easy to see that P_i is subset balanced with respect E if and only if it is locally balanced with respect to E. The two properties (for self-evident events) differ only once we have relaxed the truth axiom.

Next, we can show, following Geanakoplos:

Proposition 5.3.2. *(Geanakoplos) Let $\mathcal{M}$ be a regular model and i a player. If P_i satisfies* (4) *and* (T)*, P_i is locally balanced.*

Geanakoplos proves this proposition for his property of balancedness, but the argument is the same for local balancedness.

Proof. By induction on $|\Omega|$. The base case is trivial. Now we suppose the result holds for all $|\Omega| < n$, and consider $|\Omega| = n$. Find some ω^* such that $|P_i(\omega^*)|$ is minimal among $\{P_i(\omega)|\omega \in \Omega\}$. By (4), every $\omega' \in P_i(\omega^*)$, has $P_i(\omega') \subseteq P_i(\omega^*)$. Since $|P_i(\omega^*)|$ is minimal, it must be that $P_i(\omega') = P_i(\omega^*)$. Now construct a new model, $\mathcal{F}'$, where $\Omega' = \Omega \setminus P_i(\omega^*)$, and $P_i^{'}(\omega) = P_i(\omega) \setminus P_i(\omega^*)$. The new P_i' itself satisfies (4) and (T), and $|\Omega'| < n$, so P_i' is locally balanced, by the induction hypothesis. Now consider an E which is self evident for the original P_i. If $E \cap P_i(\omega^*) = \emptyset$, then E is self-evident to the new P_i' as well, so we use the λ which witnesses the local balancedness of P_i'. If $E = P_i(\omega^*)$, we simply use $\lambda(P_i(\omega^*)) = 1$, and $\lambda(P_i(\omega)) = 0$ for all other sets $P_i(\omega)$. Finally if $P_i(\omega^*) \subsetneq E$, we first consider $F = E \setminus P_i(\omega^*)$, which must be self-evident for P_i'. We use the λ which witnesses the local balancedness of P' on F. This leaves $P_i(\omega^*)$ unbalanced, but we can simply set $\lambda(P_i(\omega^*))$ to whatever value is needed. $\square$

It is an immediate corollary of this proposition that every partitional correspondence is locally balanced (since partitional correspondences obey both (T) and (4)). Thus local balancedness affords more generality than the standard partitions of Aumann 1976. Moreover, it is known that in finite spaces, the Agreement Theorem holds in the presence of positive introspection and truth (Bacharach 1985, Samet 1990). If we can provide an Agreement Theorem using local balancedness, it will be more general than those results.

We will also extend these results to infinite spaces, where local balancedness will prove to be more flexible than standard conditions. Samet 1992 shows that, in infinite spaces, positive introspection (4) plus truth (T) is not enough to ensure that the Agreement Theorem holds. Local balancedness not only generalizes the finite case; it also provides the needed condition in infinite spaces.

Before we define local balancedness in infinite spaces, we extend the notion of a characteristic function, using intersection instead of membership.

DEFINITION 5.3.3. Let $\chi_C : \mathscr{P}(\Omega) \to \{0,1\}$ be the function with $\chi_C(F) = 1$ if $C \cap F \neq \emptyset$, and $\chi_C(F) = 0$ otherwise.

Instead of using individual states in the definition of local balancedness, we now use cells of finite partitions. Recall that a partition $\mathcal{P}$ of a set E is *at least as fine* as a partition $\mathcal{Q}$ of E if for every $P \in \mathcal{P}$, there is some $Q \in \mathcal{Q}$ such that $P \subseteq Q$.

In addition to using finite partitions, we have to require finiteness in a further respect. Note that when we move from regular models to infinite spaces, the sum in the definition of local balancedness is no longer guaranteed to be well defined. For a given i and E, $|\mathcal{C}_E|$ is no longer be guaranteed to be finite, so even in the countable case we would need to sum relative to a sequence. In the official definition of local balancedness, we avoid this difficulty by simply stipulating that the balancing function λ assign non-zero value to at most finitely many elements of $\mathcal{C}_E$.

DEFINITION 5.3.4. Fix a model $\mathcal{M}$. A possibility correspondence P_i is *locally balanced with respect to an event* $E \in \mathcal{B}$ if and only if there is some partition $\mathcal{Q}$ of Ω which

(i) induces a finite partition of E; and

(ii) for all finite partitions $\mathcal{P}$ of Ω which are at least as fine as $\mathcal{Q}$, there exists a function $\lambda : \{P_i(\omega)|\omega \in \Omega\} \to \mathbb{R}$ so that

(a) $|\{C \in \mathcal{C}_E : \lambda(C) \neq 0\}| \in \mathbb{N}$; and

(b) for every $P \in \mathcal{P}$, $\sum_{C \in \mathcal{C}_E} \lambda(C)\chi_C(P) = \chi_E(P)$.

A correspondence is locally balanced within E if it is locally balanced with respect to every self-evident $E' \subseteq E$.

The definition of balancedness *within* E is not the global definition we used earlier for balancedness and subset balancedness. Those definitions use the special case where a correspondence is (locally) balanced within Ω.

Since $\mathcal{P}$ is itself finite, (iia) and (iib) together have the consequence that there is a finite cover of E contained in $\mathcal{C}_E$. In other words, there is an event F such that $|F|$ is finite and $\{P_i(\omega)|\omega \in F\}$ covers E.

We will be particularly interested in the following class of events:

DEFINITION 5.3.5. Fix a model $\mathcal{M}$ and a group $G \subseteq I$. A common prior consistent event E is *consistently balanced for* G if for every $i \in G$, P_i is locally balanced within E.

E is *consistently balanced* if it is consistently balanced for I.

5.3.2. Agreement Theorem. The next lemma is key to the Agreement Theorem. But in a way it is more flexible than the Agreement Theorem, and will also provide the basis for the later epistemic conditions. The lemma gives conditions under which an agent's probabilistic beliefs throughout a self-evident event agree with the distribution obtained by conditionalizing a prior on that event.

LEMMA 5.3.6. *Let $\mathcal{M}$ be a model, i a player, E an event which is consistently balanced for i, and $E^* \subseteq E$ an event which is self-evident for i. If there is an event F and a $k \in [0,1]$ such that $E^* \subseteq [p_i(F) = k]$ then for any common prior μ over E, $k = \mu(F|E^*)$.*

PROOF. By hypothesis, there is a k such that $E^* \subseteq [p_i(F) = k]$. Since μ is a common prior, we know that $\mu(C) > 0$ for all $C \in \{C|\exists\omega(\omega \in E^* \wedge P_i(\omega) = C)\} = \mathcal{C}_{E^*}$, and moreover $\dfrac{\mu(F \cap C)}{\mu(C)} = k$ for all such C.

For a given $C \in \mathcal{C}_E$, we rewrite as $\mu(F \cap C) = k\mu(C)$. Choose a finite partition $\mathcal{Q}$ of E^* which is sufficiently fine that, for all $C \in \mathcal{C}_E$ which are given non-zero value by a λ which balances E^*, $\mathcal{Q}$ induces a partition of $F \cap C$, and $\mathcal{Q}$ itself witnesses local balancedness with respect to E^*. (It's clear that such a partition exists by the definition of local balancedness.) So for a given C we have:

$$\sum_{Q \in \mathcal{Q}} \chi_C(Q)\chi_F(Q)\mu(Q) = \sum_{Q \in \mathcal{Q}} k\chi_C(Q)\mu(Q)$$

Since this is true for each $C \in \mathcal{C}_{E^*}$ to which λ assigns nonzero value, we have

$$\sum_{C \in \mathcal{C}_{E^*}} \lambda(C) \sum_{Q \in \mathcal{Q}} \chi_C(Q)\chi_F(Q)\mu(Q) = \sum_{C \in \mathcal{C}_{E^*}} \lambda(C) \sum_{Q \in \mathcal{Q}} k\chi_C(Q)\mu(Q) \tag{5.3.1}$$

for any function λ. By local balancedness with respect to E^*, and by our choice of $\mathcal{Q}$, for any cell $Q \in \mathcal{Q}$, we have, for an appropriately chosen λ:

$$\chi_{E^*}(Q) = \sum_{C \in \mathcal{C}_{E^*}} \lambda(C)\chi_C(Q).$$

Using this λ, equation 5.3.1 becomes:

$$\sum_{Q \in \mathcal{Q}} \chi_{E^*}(Q)\chi_F(Q)\mu(Q) = \sum_{Q \in \mathcal{Q}} k\chi_{E^*}(Q)\mu(Q)$$

Or, equivalently: $\mu(F|E^*) = k$, as required. □

With this lemma in hand, the Agreement Theorem follows easily:

Proposition 5.3.7. *(Agreement Theorem) Let $\mathcal{M}$ be a model, $G \subseteq I$ a group, and E an event which is consistently balanced for G. Suppose there is an $\omega \in E$, and an event F so that for each $i \in G$, there is a $k_i \in [0,1]$ such that $\omega \in CB_G([p_i(F) = k_i])$. Then for all $i, j \in G$, $k_i = k_j$.*

Proof. For each i, we know there is a public event $E_i \subseteq [p_i(F) = k_i]$. For each such E_i, $\omega \in E_i$, so choosing one for each player, $\bigcap_{i \in G} E_i$ is nonempty and thus also public; we denote this intersection E^*. Now pick an arbitrary $i \in G$. For any $C \in \mathcal{C}_{E^*}$, we write, as above: $\dfrac{\mu(F \cap C)}{\mu(C)} = k_i$. By the hypothesis that each agent is locally balanced with respect to E^*, it follows by Lemma 5.3.6, that, for each i, $k_i = \mu(F|E^*)$. Since μ, E^* and F are the same for all agents in G, $k_i = k_j$ for all $i, j \in G$. $\square$

An extensive earlier literature examined failures of introspective omniscience in connection with the Agreement Theorem and in the presence of the truth axiom (Bacharach 1985, Geanakoplos 1989, Rubinstein and Wolinsky 1990, Samet 1990, Shin 1993). Geanakoplos's theorem is the most general one presented in this initial body of work; Proposition 5.3.7 corrects this result, and extends it both to infinite spaces, and possibility correspondences which do not satisfy (T).

5.3.3. Conditioning Events and "False" Beliefs. More recently, Bonanno and Nehring (1999) proved an Agreement Theorem for finite $KD45$ models (where, in addition to (4) and (5), we have, as here: (D) $\forall \omega \in \Omega$ $P_i(\omega) \neq \emptyset$). But there is an important conceptual difference between their setting and the one used here. In defining consistency with a common prior, Bonanno and Nehring take the agent's "posterior" at ω (thought of as "interim beliefs") to be defined as $\mu(\cdot|\{\omega'|P_i(\omega') = P_i(\omega)\})$. In words, the agent's interim beliefs are as if the agent had conditioned on the event of her having the beliefs she has. This is conceptually difficult to understand; intuitively, agents should conditionalize (or be consistent with regard to) their information, not the event of their having the information they in fact have.

An example may help to illustrate the point.

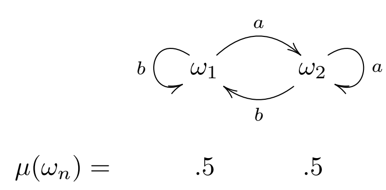


Figure 5.3.1. Conditioning Events

Example 5.3.8. Let $\mathcal{M}$ be a model with $\Omega = \{\omega_1, \omega_2\}$, and $I = \{a, b\}$. The possibility correspondences are as follows: $\forall \omega \in \Omega$, $P_a(\omega) = \{\omega_2\}$ while $P_b(\omega) = \{\omega_1\}$ (this is depicted in the figure). Finally, $p_a(\omega_1) = 0$, $p_a(\omega_2) = 1$, $p_b(\omega_1) = 1$ and $p_b(\omega_2) = 0$. Intuitively, these probabilistic beliefs are consistent with the uniform prior μ, defined so that $\forall \omega \in \Omega$, $\mu(\omega) = .5$. Player a conditionalizes this prior on what she believes; b conditionalizes the prior on what he believes. But according to Bonanno and Nehring's definition, as is easily checked, these beliefs are not consistent with a common prior.

This is the conceptual difficulty just noted. The P_i are supposed to represent what the agent believes. So the agent should conditionalize on $P_i(\omega)$ itself—what she believes—not on the event that she has the beliefs she in fact has. Even if we do not interpret the relationship between the "prior" and "posterior" as given by the agents' "update", it is unclear how to interpret the resulting property of "consistency".

This definition of conditioning events is not specific to Bonanno and Nehring. In results we will discuss later, Aumann and Brandenburger (1995), Barelli (2009) and Bach and Tsakas (2014), all similarly define the conditioning events as the set of all worlds where agents have a given type. As a formal matter, this means that if any agent has a false belief at a state, the only admissible prior will be one which assigns 0 to this state. But if the prior assigns a state 0 , then *all* agents' interim beliefs must assign that state 0. So although the models nominally relax the truth axiom, the common prior assumption plus this definition of conditioning events effectively restores it. From the standpoint of the probabilities, the agents in the models are *de facto* partitional.

Hellman (2012) also proves an Agreement Theorem for $KD45$ models. His theorem applies to $KD45$ models not included in the hypothesis of Proposition 5.3.7, but Proposition 5.3.7 describes many models in which the P_i are not $KD45$.

5.4. Correlated Equilibrium

Example 5.1.1 showed that standard epistemic conditions for correlated equilibrium may fail if agents are not introspectively omniscient. In this section, I will provide two different sets of epistemic conditions for correlated equilibrium. Both results replace the usual assumption of introspective omniscience with properties related to local balancedness.

5.4.1. Game Models, Conjectures, Rationality, and Games. To study games, we must first extend our pure epistemic models. Let a game model be a tuple

$$\mathcal{M} = \langle I, (\Omega, \mathcal{T}), (p_i)_{i \in I}, A, (\mathcal{A}_i)_{i \in I}, (u_i)_{i \in I} \rangle.$$

The new elements of the model are:

- a finite set of action profiles : $A = A_1 \times A_2, ...A_n$, where each A_i is also finite (and is assumed to be endowed with the discrete algebra);
- for each i, a measurable function $\mathcal{A}_i : \Omega \to A_i$ mapping states to actions; and
- for each i, a state-dependent utility function $u_i : \Omega \times A \to \mathbb{R}$, which for all ω, is integrable with respect to $(\Omega, \mathcal{B}, p_i(\omega))$.

In this section and Section 5, when I write "model", I will mean a game model. Since game models are extensions of epistemic models, the definitions in terms of pure epistemic models can be carried over straightforwardly to this new class of models.

As before we will want to consider certain events, such as the event that a player takes a given action. For a given $a_i \in A_i$, we use $[a_i] = \{\omega | \mathcal{A}_i(\omega) = a_i\}$ to denote the event that i plays action a_i. (This notation is slightly different than the one for conjectures and utilities.) As usual, we will use A_{-i} for $A_1 \times A_2 ... A_{i-1} \times A_{i+1} ... A_n$, and a_{-i} for an element of this set (recall that I use n for $|I|$). The event $[a_{-i}]$ is then defined in the obvious way.

In earlier results, we considered groups of players, subsets of I. When dealing with actions and utilities it becomes clumsy to achieve this kind of generality. So the main results in the following sections will use I as the group in question. This loss of generality is very slight, and it could be remedied easily, by requiring that some players' utilities be insensitive to the actions of others, and defining games relative to a group. But these extensions are both cumbersome and mathematically trivial, so I won't pursue them here.

We will consider one restriction on the relationship between actions and beliefs.

Definition 5.4.1. An event E is *action evident* for i if for all $\omega \in E$, $\exists a_i \in A_i$ such that $P_i(\omega) \subseteq [a_i] \cap E$.

This definition is motivated by the plausible idea that if players are playing a game (E obtains), they know their own actions. Formally, action evidence means more than this: each player's actions are self-evident to him or her within E.

Beliefs over actions will be particularly important. A player i's *conjecture* at a world ω, $\phi_i(\omega) \in \Delta(A_{-i})$, is a distribution over the action profiles of *other* players, defined as follows: $\phi_i(\omega)(a_{-i}) := p_i(\omega)([a_{-i}])$. i's *conjecture about* j at ω, $\phi_i^j(\omega)$ is defined by $\phi_i^j(\omega)(a_j) = p_i(\omega)([a_j])$.

A player i is *rational* at ω^* if

$$\int_\Omega u_i(\omega, \mathcal{A}_i(\omega^*), \mathcal{A}_{-i}(\omega))dp_i(\omega^*) \geqslant \int_\Omega u_i(\omega, b_i, \mathcal{A}_{-i}(\omega))dp_i(\omega^*)$$

for all $b_i \in A_i$.

I will use Rat_i for the event that i is rational (the set of states ω such that i is rational at ω), and Rat_I for $\cap_{i\in I} Rat_i$.

For fixed $\omega \in \Omega$, $u_i(\omega, \cdot)$ denotes a state *in*dependent utility function. A *game* Γ is then a tuple of pure actions and state independent payoff functions $\Gamma = ((A)_{i\in I}, (u_i(\omega_i^*, \cdot))_{i\in I})$. Note again that the game is not defined for an arbitrary group, but only for I itself.

Fixing a game $\Gamma = ((A)_{i\in I}, (u_i(\omega_i^*, \cdot))_{i\in I})$, the event of a player i *believing in* Γ, denoted $B_i(\Gamma)$, is the event of i believing that he has the utilities he would have in Γ, that is, $B_i(\Gamma) = B_i[u_i = u_i(\omega^*, \cdot)]$. Note that for i to believe in the game, he doesn't have to have any beliefs about *others'* utilities; i believes in the game in the very limited sense that i believes his or her own utilities are as they would be in Γ.

Earlier, we used consistency with a common prior, where the prior must agree with interim beliefs on all events. When considering games we won't make this strong requirement. Instead, we'll use priors *over actions*. To do this, we extend the notion of action-consistency introduced by Barelli (2009).

Definition 5.4.2. Fix a model $\mathcal{M}$. An event $E \in \mathcal{B}$ is action consistent among a group $G \subseteq I$ if and only if there exists a probability distribution $\mu \in \Delta(\Omega)$ such that

(i) $\mu(E) = 1$;

for every $i \in G$, and every $\omega \in E$,

(iia) $\mu(P_i(\omega)) > 0$; and

(iib) for every $a_{-i} \in A_{-i}$ $\phi_i(\omega)([a_{-i}]) = \mu([a_{-i}]|P_i(\omega))$.

Call any such μ an *action prior* over E. If $G = I$, we call μ simply an action prior over E.

Recall that our definition of conditioning events is different from the one used by Barelli and others, as noted above in Section 3.4. Thus the above definition is slightly different than Barelli's.

Every common prior is an action prior, but the converse does not hold in general. Action priors behave like a common prior *over actions* in E, by ensuring that players' *conjectures* (not necessarily the full complement of their beliefs) relate to each other as if they were derived from a common prior.

5.4.2. Positive Local Balancedness and Brandenburger et al. (1992). Brandenburger et al. (1992) use a version of subset balancedness to provide epistemic conditions for correlated equilibrium. Here, I will show that we can provide a related result using local balancedness.

First, we modify the definition of local balancedness by restricting the sign of the "balancing" function λ:

Definition 5.4.3. Fix a model $\mathcal{M}$. A possibility correspondence P_i is *positively locally balanced with respect to* an event $E \in \mathcal{B}$ if and only if it is locally balanced with respect to E by a function $\lambda : \{P_i(\omega) | \omega \in \Omega\} \to \mathbb{R}_{>0}$.

It is positively locally balanced within E if it is positively locally balanced with respect to every self-evident $E^* \subseteq E$.

Instead of requiring that the agent be locally balanced with respect to *every* self-evident event, we need only the following weaker property:

Definition 5.4.4. Fix a model $\mathcal{M}$. An agent $i \in I$ is *balanced in action* within E if, for every $a_i \in A_i$, P_i is locally balanced with respect to $[a_i] \cap E$.

i is *positively balanced in action* within E if, for every $a_i \in A_i$, i is positively locally balanced with respect to $[a_i] \cap E$.

If E itself is action-evident then for each a_i, $[a_i] \cap E$ is self-evident. So if a correspondence P_i is locally balanced within an action-evident event E, it is balanced in action within E. But since not every self-evident event need be of this form, a correspondence P_i may be balanced in action within E, but not locally balanced within E.

Now, by analogy to the case of consistently balanced events, we use action consistency and balancedness in action to define:

Definition 5.4.5. Fix a model $\mathcal{M}$ and a group $G \subseteq I$. An action consistent event E is *(positively) consistently action balanced* for G if for every $i \in G$, P_i is (positively) balanced in action within E.

An action consistent event E is (positively) consistently action balanced if for every $i \in I$, P_i is (positively) balanced in action within E.

The general definition (not restricted by "positively") will be used later, in Section 5. The next proposition—our first epistemic condition for correlated equilibrium—uses the notion of a positively consistently action balanced event:

Proposition 5.4.6. *(Brandenburger et al. 1992) Let $\mathcal{M}$ be a model, Γ a game, and E an action evident, positively consistently action balanced event. If $E \subseteq \bigcap_{i \in I}(B_i(\Gamma) \cap Rat_i)$, then the distribution $\hat{\mu} \in \Delta(A)$, given by $\hat{\mu}(a) := \mu([a])$, is an objective correlated equilibrium distribution for Γ.*

Brandenburger et al. (1992) provide the basis of this proposition. The above statement generalizes their result in three ways. (i) It allows for false beliefs (whereas they require truth), (ii) it uses action-consistency in place of a common prior, and, finally (iii) it extends the result to infinite spaces. Note also that we use local balancedness in place of subset balancedness.

The proof, however, is a straightforward extension of theirs, so I simply recite it in an appendix for completeness.

5.4.3. Local Action Balancedness: A Second Result. Now we turn to a new kind of epistemic condition for correlated equilibrium, which will use our earlier results about agreement. The basic strategy will be to show that the agents each agree with an action prior conditioned on their own action, in spite of not being introspectively omniscient.

First, we want to ensure that our earlier Lemma 5.3.6 about agreement carries over to action priors:

Lemma 5.4.7. *Let $\mathcal{M}$ be a model, i a player, and E an event which is consistently balanced for i. Suppose there is an event $E^* \subseteq E$ which is action evident for i so that for some $A^*_{-i} \subseteq A_{-i}$, and some $k \in [0,1]$, $E^* \subseteq [p_i(\{\omega | \mathcal{A}_{-i}(\omega) \in A^*_{-i}\}) = k]$. Then for any action prior μ for i over E, $k = \mu(F|E^*)$.*

Proof. Observe that Lemma 5.3.6 used only the fact that the prior μ is a prior for P_i *for the relevant event F*, that is, for all $\omega \in E$, $\mu(F|P_i(\omega)) = p_i(\omega)(F)$. The requirement that F have a fixed projection on A_{-i} suffices to ensure that the action prior μ is a prior for F, and the proof goes through as above. □

With this Lemma in hand, we state epistemic conditions for correlated equilibrium:

Proposition 5.4.8. *Let $\mathcal{M}$ be a model, Γ a game, and E an action consistent event. Suppose there is an action-evident $E^* \subseteq E$ such that for every $i \in I$*

(a) $E^* \subseteq B_i(\Gamma) \cap Rat_i$;

(b) i is balanced in action within E^*;

(c) for each $a_i \in A_i$, $\exists \phi_i^* \in \Delta(A_{-i})$ such that $E^* \cap [a_i] \subseteq [\phi_i = \phi_i^*]$;

Then for any action prior μ over E, the distribution $\hat{\mu} \in \Delta(A)$ given by $\hat{\mu}(a) := \mu([a]|E^*)$ is an objective correlated equilibrium distribution for Γ.

This proposition is independent of the one given by Brandenburger et al. (1992). Some correspondences are positively locally balanced within E but fail to be balanced in action within E. Conversely, some correspondences are balanced in action within a given E, and have constant conjectures in their actions, as above, but fail to be positively balanced within E.

Proof. Since E^* is action-evident, for each i and each a_i, $E \cap [a_i]$ is self-evident. Now for each i, P_i is balanced in action and ϕ_i is constant in each $E^* \cap [a_i]$. So by Lemma 5.4.7, for every $\omega \in E \cap [a_i]$, and $a_{-i} \in A_{-i}$, $\phi_i(\omega)([a_{-i}]) = \hat{\mu}([a_{-i}]|[a_i]))$. Since each player is rational with respect to each of these conjectures in the game Γ, it follows immediately that $\hat{\mu}$ is a correlated equilibrium distribution of Γ. $\square$

5.5. Epistemic Conditions for Nash

Aumann and Brandenburger (1995) provided epistemic conditions for games with more than 2 players *via* the agreement theorem. Example 5.1.1 shows that the Agreement Theorem cannot be extended to generalized possibility correspondences. With slight alteration, the example could be used to show that we also cannot extend Aumann and Brandenburger's result to this more general setting.

The question then arises: can we use the Agreement Theorem of Section 5.3 to provide epistemic conditions for Nash along the lines of Aumann and Brandenburger's original? Is the new Agreement Theorem also robust to more recent extensions of Aumann and Brandenburger's result? In this Section, I answer the question in the affirmative. The details are somewhat involved, and the new result does not give a particularly clear form to Aumann and Brandenburger's core insights. But still it is of some interest that such a result can be given.

5.5.1. Nash for Interim Beliefs without a Prior. We start by re-stating some standard facts about Nash equilibrium—not in the common prior setting. These results are very clear conceptually, and they also form the basis of the more involved, technical proposition which follows.

To state these results, we need one more definition. A player i's conjectures about j and k are *independent* if, for all $a_j \in A_j$ and $a_k \in A_k$, $\phi_i^j(\omega)([a_j])\phi_i^k(\omega)([a_k]) = \phi_i^{j,k}(\omega)([a_j] \cap [a_k])$. A conjecture $\phi_i(\omega)$ is independent if it is independent for all $j, k \in I \setminus \{i\}$. A player's conjecture being independent is itself an event, which I will denote by $Ind(\phi_i)$.

In general, for two measures μ_1 and μ_2, I will write the product measure as $\mu_1 \otimes \mu_2$. If more than two distributions are involved, I will use the symbol for product; for example, $\phi_i(\omega) = \prod_{j \in I \setminus \{i\}} marg_{A_j} \phi_i(\omega) = marg_{A_1} \phi_i(\omega) \otimes marg_{A_2} \phi_i(\omega) \otimes ...marg_{A_{i-1}} \phi_i(\omega) \otimes marg_{A_{i+1}} \otimes ...marg_{A_b} \phi_i(\omega)$ expresses the statement that i's conjecture at ω is independent.

To set the stage for the common-prior based result, we first develop simple, minimal epistemic conditions for Nash. The following Lemma is now standard:

LEMMA 5.5.1. *Let $\mathcal{M}$ be a model and $\Gamma = ((A)_{i \in I}, (u_i(\omega_i^*, \cdot))_{i \in I})$ a game. Suppose that for some ω, for some $i, j \in I$ with $i \neq j$,*

(a) for some $\phi_i^* \in \Delta(A_{-i})$, $\omega \in B_j[\phi_i = \phi_i^*]$,

(b) $\omega \in B_j B_i \Gamma$ and

(c) $\omega \in B_j(Rat_i)$.

Then for any $a_i \in A_i$ such that $\phi_j(\omega)([a_i]) > 0$, $\sum_{a_{-i} \in A_{-i}} \phi_i^*([a_{-i}]) u_i(\omega^*, a_i, a_{-i}) \geq \sum_{a_{-i} \in A_{-i}} \phi_i^*([a_{-i}]) u_i(\omega^*, b_i, a_{-i})$, for all $b_i \in A_i$.

Under conditions (a)-(c) any action which has positive support in j's conjecture about i is rational for i with respect to the conjecture j believes i has.

PROOF. By hypothesis, at ω, j believes three events: $[\phi_i = \phi_i^*]$, $B_i\Gamma = \{\omega' \in \Omega | P_i(\omega') \subseteq [u_i(\omega^*, \cdot)]\}$ and Rat_i. It follows that $p_j(\omega)$ assigns 1 to these three events. Since $\phi_j(\omega)([a_i]) > 0$, $p_j(\omega)$ assigns positive probability also to $[\phi_i = \phi_i^*] \cap B_i\Gamma \cap Rat_i \cap [a_i]$. So since this intersection is nonempty, it must be that a_i maximizes $u_i(\omega^*, \cdot)$, with respect to the conjecture ϕ_i^*. □

We then have:

PROPOSITION 5.5.2. *Let $\mathcal{M}$ be a model and $\Gamma = ((A)_{i \in I}, (u_i(\omega_i^*, \cdot))_{i \in I})$ a game. Suppose that for some world $\omega \in \Omega$, for every player $i \in I$, there is a $\sigma_i \in \Delta(A_i)$ so that* *for every $j, k \in I \setminus \{i\}$, $\phi_j^i(\omega) = \phi_k^i(\omega) = \sigma_i$. Suppose further that, for every $i \in I$, there is some player $j \in I \setminus \{i\}$, such that*

(a) for every $k \in I \setminus \{i\}$, $\omega \in B_j[\phi_i^k = \phi_i^k(\omega)]$;

(b) $\omega \in B_j B_i(\Gamma)$,

(c) $\omega \in B_j(Rat_i)$,

(d) $\omega \in B_j(Ind(\phi_i))$.

Then $(\sigma_1, \sigma_2 ... \sigma_I)$ is a Nash Equilibrium of Γ.

Notice that (a) in this Proposition is different from (a) in the previous Lemma. Here, it is not just that there is *some* conjecture j believes i has; the believed conjecture must (a) agree with i's actual marginal conjectures (i's marginal conjectures at ω) *and* (d) represent others' actions as independent.

PROOF. Consider an arbitrary player i. By hypothesis, there is some $j \neq i$ such that j believes that, for every $k \in I \setminus \{i\}$, $i's$ conjecture about k is ϕ_i^{k*}, where in fact this is correct and $\phi_i^{k*} = \phi_i^k(\omega)$. Since every player agrees in their marginal conjectures at ω, then for all $k \in I \setminus \{i\}$, this $\phi_i^k(\omega) = \sigma_k$. Moreover, at ω, j believes that i's full conjecture is $\prod_{k \in I \setminus \{i\}} \phi_i^k(\omega) = \prod_{k \in I \setminus \{i\}} \sigma_k$, j believes that i is rational, and believes that i believes i has utility $u_i(\omega_i^*, \cdot)$. By Lemma 5.5.1, it follows that every a_i assigned positive probability by $\phi_j^i(\omega) = \sigma_i$ is a best response to $\prod_{k \in I \setminus \{i\}} \sigma_k$, when i's payoffs are given by $u_i(\omega_i^*, \cdot)$. Since i was chosen arbitrarily, this holds for every player, and thus $(\sigma_1, \sigma_2 ... \sigma_n)$ is a Nash Equilibrium of Γ. $\square$

Versions of this fact are given in Brandenburger and Dekel (1987: Proposition 4.1), Aumann and Brandenburger (1995) Remark 7.1, Liu (2010 (unpublished)), Dekel and Siniscalchi (2014: 21-2) and Battigalli et al. (2014). As usual, this result is more general than the standard ones, since I have made no assumptions whatsoever about introspection. In fact, this proposition is very general, since it does not even require local balancedness.

5.5.2. Independence. As Proposition 5.5.2 illustrates, two claims must be established to give the epistemic conditions for Nash: agreement of marginal conjectures, and independence of believed conjectures. For agreement, we will simply use the earlier Lemma 5.4.7. Independence will require two new lemmas.

In later results we will be using events E which are both action evident and consistently action balanced (see Definition 5.4.5). Our first lemma will show that if E satisfies these properties, it turns out that each P_i is also locally balanced with respect to E itself.

LEMMA 5.5.3. *Let $\mathcal{M}$ be a model, i a player, and E an event which is action consistent and action-evident for i. If P_i is balanced in action within E, then P_i is locally balanced with respect to E.*

PROOF. $\mathcal{P} = \{\{[a_i] \cap E\} | a_i \in A_j\}$ is a finite partition of E. Moreover, since E is action-evident, each of these events is self-evident to i, so for any distinct $a_i, \hat{a}_i \in A_i$ of $\mathcal{C}_{E\cap[a_i]} \cap \mathcal{C}_{E\cap[\hat{a}_i]} = \emptyset$. Fix an enumeration of the $a_i \in A_i$, and denote the nth action as a_i^n. For each such a_i^n, by hypothesis, there exists $\mathcal{Q}_n$ a finite partition of $E \cap [a_i^n]$, such that, for $\mathcal{Q}_n$, and any finer finite partition of $E \cap [a_i^n]$, there is some λ_n which witnesses the local balancedness of P_i with respect to this $E \cap [a_i^n]$. For each a_i^n pick such a $\mathcal{Q}_n$, and now consider the union of all these partitions $\bigcup_{1 \leqslant n \leqslant k} \mathcal{Q}_n = \mathcal{Q}$. This new $\mathcal{Q}$ partitions E, and is finite. We now construct λ so that, if $\omega \in P_n$, $\lambda(P_i(\omega)) = \lambda_n(P_i(\omega))$. (Given any finer but still finite partition, we can construct λ in the same way.) These λ witness the local balancedness of P_i with respect to E. $\square$

Now we turn to independence directly.

LEMMA 5.5.4. *Let $\mathcal{M}$ be a model, i a player, and E an event which is action evident, and consistently action balanced for i. If for some $\phi_i^* \in \Delta(A_{-i})$, $E \subseteq [\phi_i = \phi_i^*]$, then for any action prior μ for i over E, $marg_A\mu = marg_{A_i} \otimes marg_{A_{-i}}$, and for every $\omega \in E$, $\phi_i(\omega) = marg_{A_{-i}}\mu$.*

PROOF. By constancy of i's conjectures within the action-evident E, and i's local balancedness with respect to $E \cap [a_i]$, for each $a_i \in A_i$, it follows from Lemma 5.4.7 that, for $\omega \in E$ with $\mathcal{A}_i(\omega) = a_i$,

$$\phi_i(\omega)([a_{-i}]) = \mu([a_i, a_{-i}] | [a_i]) \tag{5.5.1}$$

for all $a_{-i} \in A_{-i}$. Rearranging, for every $\omega \in E$ with $\mathcal{A}_i(\omega) = a_i$, we have $\mu([a_i, a_{-i}]) = \phi_i(\omega)([a_{-i}])\mu([a_i])$, for all $a_{-i} \in A_{-i}$. Since this holds for each $a_i \in A_i$, it follows that $marg_A\mu = marg_{A_i}\mu \otimes marg_{A_{-i}}\mu$. But then for each a_i, it's clear that $marg_A(\mu(\cdot|[a_i]) = marg_{A_{-i}}\mu$. Since P_i is action balanced within the action evident E, and $\exists \phi^*$ such that $E \subseteq [\phi_i = \phi_i^*]$, it follows from Lemma 5.4.7 that, for any $\omega \in E$, $\phi_i(\omega) = marg_A\mu(\cdot|[\mathcal{A}_i(\omega)]) = marg_{A_{-i}}\mu$. $\square$

5.5.3. Pairwise Epistemic Conditions. Following Bach and Tsakas (2014), we can give epistemic conditions for Nash equilibrium using a notion of *pairwise* general action consistency, defined relative to a graph on the set of players. Our strategy will be to build up agreement and independence pair-by-pair along the edges of the graph, to show that the conjectures of the whole group agree, and that each conjecture treats all others as acting independently.

First, we recall some graph-theoretic definitions. An *undirected graph* $G = (N, \mathcal{E})$ consists of a set of vertices and a set of edges described by the binary symmetric relation $\mathcal{E} \subseteq N \times N$ (where $(i,j) \in \mathcal{E}$ describes the same edge as $(j,i) \in \mathcal{E}$). A sequence $(i_k)_{k=1}^m$ of players is a *path* if and only if $(i_k, i_{k+1}) \in \mathcal{E}$ for all $k \in \{1, ..., m-1\}$. A graph is *connected* if for every $i, j \in N$, there is a path $(i_k)_{k=1}^m$ with $i = i_1$ and $j = i_m$. A *cut vertex* of a connected graph is a node $i \in N$ such that, if it is deleted (and any edges containing it are removed), the resulting graph is not connected. A graph is *biconnected* if it has no cut vertices.

Definition 5.5.5. Fix a model $\mathcal{M}$. Let a state $\omega \in \Omega$ be pairwise action consistent if and only if there exists a biconnected graph $H = (I, \mathcal{E})$ where for each $(j,k) \in \mathcal{E}$, there exists an action evident, consistently action balanced event E_{jk} for the group $G_{jk} = \{j, k\}$ such that:

(i) $\omega \in E_{j,k}$; and

(ii) for some $\phi_j^* \in \Delta(A_{-j})$, $\phi_k^* \in \Delta(A_{-k})$, $E_{jk} \subseteq [\phi_j = \phi_j^*] \cap [\phi_k = \phi_k^*]$.

This definition sets the stage for our final proposition:

Proposition 5.5.6. *Let $\mathcal{M}$ be a model, Γ a game, and $\omega \in \Omega$ a pairwise action consistent state.*

Suppose that for every $i \in I$, there exists some $j \in I$ with $j \neq i$, such that:

(a) $\omega \in B_j[\phi_i = \phi_i(\omega)]$

(b) $\omega \in B_j B_i(\Gamma)$

(c) $\omega \in B_j(Rat_i)$.

Then all $j \in I \setminus \{i\}$ share the same conjecture σ_i about i, and the vector of these conjectures about individuals, $(\sigma_1, ..., \sigma_I)$ is a Nash Equilibrium for the game Γ.

Proof. We break the proof into three stages: (1) agreement; (2) independence; (3) Nash.

Agreement. Pick an arbitrary $i \in I$, and an arbitrary $j \in I \setminus \{i\}$. We first show, by induction on path length in H, that $\phi_j^i(\omega) = \phi_k^i(\omega)$ for all $k \in I \setminus \{i, j\}$. Suppose, for the base case, that $(j,k) \in \mathcal{E}$: k is reachable from j in a single step. By Lemma 5.5.3, Definition 5.5.5 $(i)-(ii)$ imply j and k's local balancedness with respect to E_{jk}. We now use Lemma 5.4.7, and Definition 5.5.5 $(i) - (ii)$, to obtain $\phi_j(\omega)([a_i]) = \phi_k(\omega)([a_i])$, for all $a_i \in A_i$.

Now we suppose this holds for any k_n reachable in n steps on the graph H, and show that it holds for any k_{n+1} reachable in $n+1$ steps on H. Any such k_{n+1} is reachable in one step from some k_n with $(k_n, k_{n+1}) \in \mathcal{E}$. So Definition 5.5.5 $(i) - (ii)$ hold for $G_{k_n,k_{n+1}} = \{k_n, k_{n+1}\}$, where $k_n \in I$ is some player reachable in n steps from this initial j. Now another application of Lemma 5.4.7 gives us that $\phi_{k_n}(\omega)([a_i]) = \phi_{k_{n+1}}(\omega)([a_i])$, for all $a_i \in A_i$. But, by the induction hypothesis, $\phi_{k_n}(\omega)([a_i]) = \phi_j(\omega)([a_i])$, for all $a_i \in A_i$. Since the graph H is biconnected, and I is finite, we can reach every element of $I \setminus \{i\}$ in a finite number of steps, without passing through i,. Moreover, since the initial i was chosen arbitrarily, it follows that for any $i \in I$, all $j, k \in I \setminus \{i\}$ agree in their conjecture about i.

Independence. We now use agreement in marginal conjectures to show that all players' conjectures are independent. First we show that, for every $i, j, k, l \in I$, with $(i, j), (k, l) \in \mathcal{E}$, $marg_A \mu_{ij} = marg_A \mu_{kl}$, by showing that an arbitrary subpath $(a, b), (b, c) \in \mathcal{E}$ preserves equality of $marg_A \mu_{ab} = marg_A \mu_{bc}$. By b's local action balancedness relative to E_{ab} and Lemma 5.5.4, we have $marg_A \mu_{ab} = marg_{A_b} \mu \otimes \phi_b(\omega)$. By Lemma 5.4.7, a's constancy in E_{ab}, and a's local balancedness with respect to E_{ab} (given, once again by using Lemma 5.5.3, and Definition 5.5.5$(i) - (ii)$), we have $marg_{A_b} \mu = \phi_a^b = \sigma_b$, where σ_b is the marginal conjecture shared among all players other than b about b. Thus $marg_A \mu_{ab} = \sigma_b \otimes \phi_b(\omega)$. The same argument shows that $\mu_{bc} = \sigma_b \otimes \phi_b(\omega)$, and thus, that $\mu_{ab} = \mu_{bc}$. Since a, b, c were chosen arbitrarily, any such path preserves equality. By the (bi)connectedness of H, there is such a path from any i to any l. Every step in any such path preserves equality of the relevant action-priors, and thus, for all $i, j, k, l \in I$, with $(i, j), (k, l) \in \mathcal{E}$, $marg_A \mu_{ij} = marg_A \mu_{kl}$, as required. From now on we denote this distribution on actions simply by μ.

Now we show that $\mu = \prod_{i \in I} marg_{A_i} \mu$. Pick an arbitrary $i \in I$. By connectedness of H, there is some j with $(i, j) \in \mathcal{E}$. By the argument just given $marg_A \mu_{ij} = \mu$. Moreover, by i's local action balancedness with respect to E_{ij}, we have $marg_A \mu_{ij} = \mu = marg_{A_i} \mu \otimes marg_{A_{-i}} \mu$. Since this holds for all $i \in I$, we have, by Aumann and Brandenburger (1995) Lemma 4.6, that $\mu = \prod_{i \in I} marg_{A_i} \mu$. (The full argument is by induction on the cardinality of I, and may be found in their paper.)

To complete the proof of independence, it remains to show only that, for every i, $\phi_i = marg_{A_{-i}} \mu$. By the connectedness of H, every i has at least one neighbor, j, and for every $(i, j) \in \mathcal{E}$, there is some E_{ij}, with μ_{ij} an action prior on E_{ij}, and i locally balanced with respect to E_{ij}. By the constancy of i's conjecture in E_{ij}, we can use Lemma 5.4.7, to derive that $\phi_i(\omega) = marg_{A_{-i}}\mu_{ij}$. But, by the earlier argument $marg_{A_{-i}}\mu_{ij} = marg_{A_{-i}}\mu$, as required.

Nash. With agreement and independence in hand, Nash follows by Lemma 5.5.1. □

Proposition 5.5.6 generalizes Bach and Tsakas (2014) in two ways. First, it allows belief in rationality to float free from the graph structure. In the above statement, belief in rationality may fail to form even a connected graph.[9] Second, local action balancedness is weaker than their assumption that agents are introspectively omniscient. Note once again that the definition of conditioning events used here is different from theirs as well.

Polak (1999) observed that, in Aumann and Brandenburger's Theorem B, common knowledge of conjectures plus common knowledge of payoffs implies common knowledge of rationality. As a result, in perfect information games, where payoffs are assumed to be commonly known, satisfaction of Aumann and Brandenburger's weak conditions still implies common knowledge of rationality. Barelli (2009) and Bach and Tsakas (2014) have shown that this result need not follow, by giving weaker epistemic conditions for Nash Equilibrium. Proposition 5.5.6 and Proposition 5.5.2 share this feature, though the more general and transparent Lemma 5.5.1 has a more striking version of it.

5.6. Conclusion: Interpretation of Local Balancedness

Without introspective omniscience, some common prior-based results fail. Local balancedness is considerably weaker than introspective omniscience, and it allows us to preserve some standard results.

But this leaves us with two questions. First, are these the weakest constraints which can be given which validate these results? Second, and somewhat relatedly, does local balancedness have any interesting interpretation? Given positive introspection, for example, the Agreement Theorem holds. Positive introspection has a simple interpretation, and some might be happy maintaining positive introspection in this setting. But in others, e.g. for correlated equilibrium, positive introspection is not enough. (Recall that the agents in Example 5.1.1 were both positively introspective—they obeyed (4)—but the prior there failed to form a correlated equilibrium.) For local balancedness, we don't seem to have an interpretation which is as clear as the one for positive introspection.

[9]For example, with four players, a, b, c, d: if the other conditions are satisfied, it suffices that a believe that b is rational and b believe that a is rational, while c believe that d is rational, and d believes that c is rational. a and b may each believe c and d not to be rational, and c and d may return the favor.

Ideally we would like to find some way of understanding local balancedness in terms of the constraints it imposes on players' beliefs and knowledge.

The first question can be answered partly in the affirmative. As I have noted, the main result of the paper is the Agreement Theorem. Geanakoplos provided a partial converse to his version of this result. Recall that we can specify regular models just by specifying a prior over the state space in the model. Then if $\mathcal{M} = \langle \Omega, I, (P)_{i \in I}, \mu \rangle$ is a regular model, and $i \in I$ is a player, then a regular model $\mathcal{M}' == \langle \Omega, I', (P')_{i \in I}, \mu' \rangle$ is *i-equivalent* if $i \in I'$, and $P_i = P'_i$.

Then we can state Geanakoplos's converse, using local balancedness, as follows:

Proposition 5.6.1. *(Geanakoplos 1989) Let $\mathcal{M}$ be a regular model, and $i \in I$ a player. If P_i is not locally balanced, there is an i-equivalent $\mathcal{M}'$ and a $j \in I'$ such that P_j is locally balanced within Ω, and i and j agree to disagree.*

This converse applies only to regular models. It also requires quantifying i-equivalent models, that is, over possible priors and possible balanced possibility correspondences. It says nothing about agents on a specific occasion, given the beliefs they in fact have. But unfortunately, stronger converses are not available. If two agents have possibility correspondences which are ill-behaved in exactly the same ways—that is, $\forall \omega \; P_i(\omega) = P_j(\omega)$ and $p_i(\omega) = p_j(\omega)$—then they cannot agree to disagree, because they cannot disagree at all. This converse may be the best we can hope for.[10]

How should we understand local balancedness? Geanakoplos suggested that "Balancedness is a condition under which one can say that every $\omega \in E$ is equally scrutinized by the information correspondence E. Every element $C \in \mathcal{P}$ has an intensity $\lambda(C)$, and the sum of the intensities with which each $\omega \in E$ is considered possible by P is the same, namely 1." (1989: 18) The following model, however, is both balanced and locally balanced, although it does not appear to exhibit "equal scrutiny" in any intuitive sense.

[10]Hellman 2012 gives more exact necessary conditions, but his theorem is of an essentially different kind than the ones studied in this paper. The statement of Hellman's conditions on agents' beliefs are essentially interactive: a given possibility correspondence does not satisfy them (or fail to satisfy them) if a group has not been specified. This makes his result essentially different from the standard theorems, where interactive results are derived from at least *prima facie* single-agent constraints.

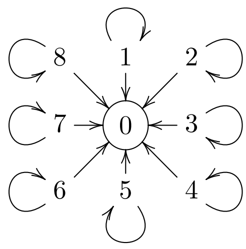


FIGURE 5.6.1. Local Balancedness and Truth do Not Imply Equal Scrutiny (0 can see itself)

World 0 is seen from 9 worlds, whereas all other 8 worlds in the model are seen only from one, namely, themselves. The example could be extended to yield any finite "degree of scrutiny" for one world, while every other world is seen by only itself.

So we are still in need of an intuitive characterization. In the case of positive introspection, we give an intuitive characterization of a model-theoretic property by considering the propositional modal logic which characterizes the class of models which satisfy this property (or, in the logical terminology the class of "frames"). One might hope that, as with positive introspection, we can similarly use a propositional modal logic to give a transparent characterization of which models are locally balanced. Formally, this would mean finding a logic of belief which is valid exactly on the class of models $\mathcal{M}$ in which agents are (positively) locally balanced within the universe Ω of $\mathcal{M}$. If a class of models can be characterized in this way, we say that the class of models is *definable*. Unfortunately, the following proposition shows that such an equivalence cannot be given; it thus suggests that properties related to local balancedness may not admit an intuitive characterization in terms of belief.

PROPOSITION 5.6.2. *The class of (positively) locally balanced models is undefinable by a normal modal logic.*

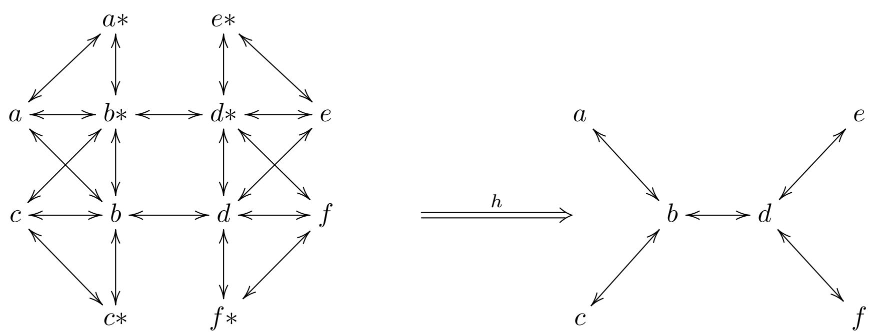


FIGURE 5.6.2. A Locally Balanced model which has a Bounded Morphic Image that is not Locally Balanced

Recall that for conditions such as $(T) + (4)$ on possibility correspondences, the class of models is definable by a normal modal logic, in this case, the logic $S4$.

PROOF. By the Goldblatt-Thomason theorem, a class of models is definable by a normal modal logic only if it is closed under bounded morphic images.[11] In figure 5.6.2 the left-hand model is positively locally balanced (and therefore locally balanced). To see this, note that the only self-evident event is the universe, and that $\{P_i(a^*), P_i(e^*), P_i(c^*), P_i(f^*)\}$ partitions the universe. The right-hand model is not locally balanced. To see this, consider the uniform prior μ defined so that: $\forall \omega \in \Omega$, $\mu(\omega) = \frac{1}{6}$. While $\mu(\{a, c, e, f\}) = \frac{2}{3}$, $\mu(\{a, c, e, f\}|P_i(\omega)) = \frac{1}{2}$ for all $\omega \in \Omega$. Since $\mu(\{a, c, e, f\}|P_i(\omega))$ is constant within Ω, but does not agree with $\mu(\{a, c, e, f\}|\Omega)$, it follows by contraposition of Lemma 5.3.6 that the correspondence is not locally balanced. But the model on the right-hand side of Figure 5.6.2 is a bounded morphic image of the one on the left. Consider the function h which takes every unasterisked world to itself, but takes the asterisked worlds to their unasterisked counterparts. Call the accessibility relation (the relation of directed adjacency) of the left-hand model 'R', and the accessibility relation of the right hand model 'R''. It is straightforward to check that (i) wRv implies $h(w)R'h(v)$; (ii) $h(w)R'v'$ implies that there is some v such that wRv and $h(v) = v'$; (iii) h is surjective. $\square$

Appendix

PROPOSITION. *(Brandenburger et al. 1992) Let $\mathcal{M}$ be a model, Γ a game, and E an action evident, positively consistently action balanced event. If $E \subseteq \bigcap_{i \in I}(B_i(\Gamma) \cap Rat_i)$, then the distribution $\hat{\mu} \in \Delta(A)$ given by $\hat{\mu}(a) := \mu([a])$ is an objective correlated equilibrium distribution for Γ.*

PROOF. (Brandenburger et al. 1992) Following Brandenburger et al, fixing the game $\Gamma = ((A)_{i \in I}, (u_i(\omega_i^*, \cdot))_{i \in I})$ and given a distribution $\nu \in \Delta(A)$, with $\nu(\{a_i\} \times A_{-i})) > 0$, let

$Q_\nu(a_i) = \{q \in \Delta(A) : supp(q) \subseteq supp(\nu(\cdot|a_i)),$

$$\sum_{a_{-i} \in A_{-i}} q(a_{-i}) u_i(\omega_i^*, a_i, a_{-i}) \geqslant \sum_{a_{-i} \in A_{-i}} q(a_{-i}) u_i(\omega_i^*, b_i, a_{-i}) \text{ for all } b_i \in A_i.\}$$

(We use *supp*, as usual, to denote the smallest closed set with measure 1, and use the discrete topology on the finite set of action profiles.) Observe that $Q_\nu(a_i)$ is a compact,

[11] For a standard reference, and proof of the Goldblatt-Thomason theorem, see Blackburn et al. 2001: Sections 3.8 and 5.4.

convex subset of $\Delta(A_{-i})$. We now show that for any action-prior μ on E, $\hat{\mu}$, defined as above, is a correlated equilibrium distribution. This claim is equivalent to the claim that, for each i, for every $a_i \in A_i$ such that $\mu([\{a_i\} \times A_{-i}]) > 0$, $marg_{A_{-i}}\hat{\mu}(\cdot|[a_i]) \in Q_{\hat{\mu}}(a_i)$.

For every $i \in I$, $a_i \in A_i$ and $\omega \in E \cap [a_i]$, it is straightforward to show that $\phi_i(\omega) \in Q_{\hat{\mu}}(a_i)$. Since μ is an action prior, for every $i \in I$, and $\omega \in E$, $supp(\phi_i(\omega)) \subseteq supp(marg_{A_{-i}}(\hat{\mu}))$. Moreover, by the fact that $E \subseteq \bigcap_{i \in I}(B_i(\Gamma) \cap Rat_i)$, we have that for every $i \in I$, $a_i \in A_i$ and $\omega \in E \cap [a_i]$, a_i is optimal with respect to $\phi_i(\omega)$, and $u_i(\omega_i^*, \cdot)$. Thus for every $i \in I$, $a_i \in A_i$ and $\omega \in E \cap [a_i]$, $\phi_i(\omega) \in Q_{\hat{\mu}}(a_i)$.

By the convexity of $Q_{marg_A\mu}(a_i)$, it now suffices to show that, for every $i \in I$, for every $a_i \in A_i$, $marg_{A_{-i}}\hat{\mu}(\cdot|[a_i]) \in conv\{\phi_i(\omega)|\omega \in E \cap [a_i]\}$, where $conv$ denotes the convex hull of the set of all such conjectures. By action consistency, we have $\phi_i(\omega)([a_{-i}]) = \mu([a_{-i}]|P_i(\omega))$, for every $a_{-i} \in A_{-i}$. Recall that given an event E and a player i, we use $\mathcal{C}_E$ to denote the set $\mathcal{C}_E = \{C|\exists\omega(\omega \in E \wedge P_i(\omega) = C)\}$. Consider an arbitrary $i \in I$, $a_i \in A_i$, and $a_{-i} \in A_{-i}$, so that $\mu([a_{-i}]|E \cap [a_i]) > 0$. Now choose a finite partition $\mathcal{Q}$, but fine enough to partition $E \cap [a_i, a_{-i}]$, and fine enough to witness the positive local balancedness of i with respect to $E \cap [a_i]$. Since i is positively locally balanced in action within E,

$$\mu([a_{-i}]|[a_i]) = \frac{1}{\mu([a_i])} \sum_{Q \in \mathcal{Q}} \chi_{[a_i]\cap E}(Q)\chi_{[a_{-i}]}(Q)\mu(Q)$$

$$= \frac{1}{\mu([a_i])} \sum_{Q \in \mathcal{Q}} \sum_{C \in \mathcal{C}_{E\cap[a_i]}} \lambda(C)\chi_C(Q)\chi_{[a_{-i}]}(Q)\mu(Q)$$

for an appropriately chosen λ. It follows that:

$$\mu([a_{-i}]|[a_i]) = \frac{1}{\mu([a_i])} \sum_{C \in \mathcal{C}_{E\cap[a_i]}} \lambda(C)\mu(C)\frac{1}{\mu(C)} \sum_{Q \in \mathcal{Q}} \chi_C(Q)\chi_{[a_{-i}]}(Q)\mu(Q)$$

$$= \frac{1}{\mu([a_i])} \sum_{C \in \mathcal{C}_{E\cap[a_i]}} \lambda(C)\mu(C)\mu([a_{-i}]|C). \tag{5..1}$$

But local positive balancedness gives us that $\frac{1}{\mu([a_i])}\sum_{C \in \mathcal{C}_{E\cap[a_i]}} \lambda(C)\mu(C) = 1$, and, moreover, that every value in the sum is at least 0. So 5..1 gives us

$$\mu([a_{-i}]|[a_i]) \in conv\{\phi_i(\omega)([a_{-i}])|\omega \in E \cap [a_i]\} \tag{5..2}$$

This holds for all $a_{-i} \in A_{-i}$, so $marg_{A_{-i}}\mu(\cdot|[a_i])$ lies in the convex hull of the $\phi_i(\omega)$ such that $\omega \in E \cap [a_i]$, as required. Since $Q_{marg_A\mu}(a_i)$ is convex, it follows by 5..2 that $marg_{A_{-i}}\mu$ lies in $Q_{marg_A\mu}(a_i)$ as well, for every a_i such that $marg_A\mu(\{a_i\} \times A_{-i})) > 0$. So $marg_A\mu = \hat{\mu}$ is a correlated equilibrium for Γ. $\square$

Bibliography

Aarnio, Maria Lasonen. 2010. Unreasonable knowledge. *Philosophical Perspectives*, **24**(1), 1–21. 123

Aaronson, Scott. 2005. The Complexity of Agreement. *Symposium on the Theory of Computing*, Extended Abstract. Full Version at www.scottaaronson.com/papers/agree–econ.pdf. 96

Akkoyunlu, Eralp A., Ekanadham, Kattamuri, & Huber, R. V. 1975. Some Constraints and Tradeoffs in the Design of Network Communications. *SIGOPS Oper. Syst. Rev.*, **9**(5), 67–74. 13

Almotahari, Mahrad, & Glick, Ephraim. 2010. Context, content, and epistemic transparency. *Mind*, **119**(476), 1067–1086. 65

Arad, Ayala, & Rubinstein, Ariel. 2012. The 11-20 Money Request Game: A Level-k Reasoning Study. *American Economic Review*, **102**(7), 3561–73. 17

Aumann, Robert J. 1976. Agreeing to Disagree. *The Annals of Statistics*, **4**(6), 1236–1239. 18, 95, 102, 140, 141, 145

Aumann, Robert J. 1985. On the non-transferable utility value: A comment on the Roth-Shafer examples. *Econometrica: Journal of the Econometric Society*, 667–677. 27

Aumann, Robert J. 1987. Correlated Equilibrium as an Expression of Bayesian Rationality. *Econometrica*, **55**(1), 1–18. 137

Aumann, Robert J. 1998. Common Priors: A Reply to Gul. *Econometrica*, **66**(4), 929–938. 115

Aumann, Robert J., & Brandenburger, Adam. 1995. Epistemic Conditions for Nash Equilibrium. *Econometrica*, **63**(5), 1161–1180. 25, 149, 154, 156, 159

Aumann, Robert J., & Hart, Sergiu. 2006 (unpublished). *Agreeing on decisions.* The Hebrew University of Jerusalem, Center for the Study of Rationality. 109

Aumann, Robert J., Hart, Sergiu, & Perry, Motty. 2005 (unpublished). *Conditioning and the sure-thing principle.* Center for the Study of Rationality. 108, 109

Bach, Christian W., & Tsakas, Elias. 2014. Pairwise epistemic conditions for Nash equilibrium. *Games and Economic Behavior*, **85**(0), 48 – 59. 25, 139, 149, 157, 160

Bacharach, Michael. 1985. Some extensions of a claim of Aumann in an axiomatic model of knowledge. *Journal of Economic Theory*, **37**(1), 167 – 190. 103, 105, 108, 137, 145, 148

Barelli, Paolo. 2009. Consistency of Beliefs and Epistemic Conditions for Nash and Correlated Equilibria. *Games and Economic Behavior*, **67**, 363–375. 25, 113, 149, 151, 160

Barwise, Jon. 1988. Three views of common knowledge. *Pages 365–379 of: Proceedings of the 2nd Conference on Theoretical Aspects of Reasoning about Knowledge*. Morgan Kaufmann Publishers Inc. 42

Battigalli, Pierpaolo, Brandenburger, Adam, Friedenberg, Amanda, & Siniscalchi, Marciano. 2014. *Strategic Uncertainty: The Epistemic Approach to Game Theory.* 25, 156

Bernheim, B. Douglas. 1984. Rationalizable Strategic Behavior. *Econometrica*, **52**(4), 1007–1028. 27, 35

Binmore, Ken, & Samuelson, Larry. 2001. Coordinated action in the electronic mail game. *Games and Economic Behavior*, **35**(1), 6–30. 25

Blackburn, Patrick, de Rijke, Maarten, & Venema, Yde. 2001. *Modal Logic.* Cambridge University Press. 163

Bonanno, Giacomo, & Nehring, Klaus. 1997 (September). *Agreeing to Disagree: A Survey.* http://www.econ.ucdavis.edu/faculty/bonanno/PDF/agree.pdf. 113

Bonanno, Giacomo, & Nehring, Klaus. 1999. How to make sense of the common prior assumption under incomplete information. *International Journal of Game Theory*, **28**(3), 409–434. 113, 137, 141, 148

Bonanno, Giacomo, & Nehring, Klaus. 2000. Common Belief with the Logic of Individual Belief. *Mathematical Logic Quarterly*, **46**(1), 49–52. 93

Brandenburger, Adam, & Dekel, Eddie. 1987. Rationalizability and correlated equilibria. *Econometrica*, **55**(6), 1391–1402. 27, 35, 156

Brandenburger, Adam, & Keisler, Jerome H. 2006. An Impossibility Theorem on Beliefs in Games. *Studia Logica*, **84**, 211–240. 83

Brandenburger, Adam, Dekel, Eddie, & Geanakoplos, John. 1992. Correlated Equilibrium with Generalized Information Structures. *Games and Economic Behavior*, **4**, 182–201. 137, 139, 143, 152, 153, 154, 163

Briggs, Rachael. 2009. Distorted Reflection. *The Philosophical Review*, **118**(1), 59–85. 121

Camerer, Colin. 2003. *Behavioral game theory: Experiments in strategic interaction.* Princeton University Press. 22

Cave, Jonathan A.K. 1983. Learning to agree. *Economics Letters*, **12**(2), 147 – 152. 103, 105, 137

Chwe, Michael Suk-Young. 2001. *Rational Ritual: Culture, Coordination, and Common Knowledge.* Princeton University Press. 36

Clark, Herbert H. 1996. *Using language.* Cambridge University Press. 44

Clark, HH, & Marshall, CR. 1981. Definite reference and mutual knowledge. 10–64. 44

Cohen, Stewart, & Comesaña, Juan. 2013. Williamson on Gettier Cases and Epistemic Logic. *Inquiry*, **56**(1), 15–29. 78

Collins, John. 1997. *How We Can Agree to Disagree.* http://collins.philo.columbia.edu/disagree.pdf. 121

Conee, Earl, & Feldman, Richard. 2004. *Evidentialism: Essays in Epistemology.* Oxford University Press. 103

Cowen, Tyler, & Hanson, Robin. 2014 (forthcoming). Are disagreements honest? *Journal of Economic Methodology.* 96

Dekel, Eddie, & Siniscalchi, Marciano. 2014 (14 January). *Epistemic Game Theory.* 25, 156

Dekel, Eddie, Lipman, Bart L, & Rustichini, Aldo. 1998. Standard state-space models preclude unawareness. *Econometrica*, **66**(1), 159–73. 7, 128, 137

Easwaran, Kenny. 2014. Regularity and Infinitesimals. *Philosophical Review*, **123**(1), 1–41. 97

Elga, Adam. 2007. Reflection and Disagreement. *Noûs*, **41**(3), 478–502. 121

Epstein, Larry G, & Wang, Tan. 1996. " Beliefs about Beliefs" without Probabilities. *Econometrica: Journal of the Econometric Society*, 1343–1373. 135

Fagin, Ronald, Halpern, Joseph Y., Moses, Yoram, & Vardi, Moshe Y. 1995. *Reasoning about Knowledge.* MIT Press. 13, 18, 100

Feinberg, Yossi. 2000. Characterizing Common Priors in the Form of Posteriors. *Journal of Economic Theory*, **91**(2), 127 – 179. 113, 141

Feldman, Richard. 2007. Reasonable Religious Disagreements. *Chap. 17, pages 194–214 of:* Antony, Louise (ed), *Philosophers without Gods: Meditations on Atheism and the Secular.* Oxford University Press. 97

Friedell, Morris F. 1969. On the structure of shared awareness. *Behavioral Science*, **14**(1), 28–39. 18, 102, 141

Geanakoplos, John. 1989 (May). *Game Theory Without Partitions, and Applications to Speculation and Consensus.* Cowles Foundation Discussion Papers 914. Cowles Foundation for Research in Economics, Yale University. 7, 134, 137, 139, 141, 142, 148, 161

Geanakoplos, John. 1994. Common Knowledge. *Chap. 40, pages 1437–1496 of:* Aumann, Robert J., & Hart, Sergiu (eds), *Handbook of Game Theory with Economic Applications*, vol. 2. North Holland. 113

Gilbert, Margaret. 1989. *On social facts.* Princeton University Press. 41

Goodman, Jeremy. 2013. Inexact knowledge without improbable knowing. *Inquiry*, **56**(1), 30–53. 30

Gray, Jim. 1978. Notes on Data Base Operating Systems. *Pages 393–481 of: Operating Systems, An Advanced Course.* London, UK: Springer-Verlag. 13

Greco, Daniel. 2014a. Could KK be OK? *Journal of Philosophy*, **111**(4), 169–197. 10, 119

Greco, Daniel. 2014b. Iteration and Fragmentation. *Philosophy and Phenomenological Research*, n/a–n/a. 10, 78

Grimm, V., & Mengel, F. 2012. An experiment on learning in a multiple games environment. *Journal of Economic Theory*, **147**(6), 2220–2259. 28

Groenendijk, Jeroen, & Stokhof, Martin. 1991. Dynamic predicate logic. *Linguistics and Philosophy*, **14**(1), 39–100. 44

Hájek, Alan. 2010. *Staying regular.* Unpublished Manuscript. 97

Halpern, Joseph Y, & Pucella, Riccardo. 2011. Dealing with logical omniscience: Expressiveness and pragmatics. *Artificial intelligence*, **175**(1), 220–235. 102

Hanson, Robin. 2006. Uncommon Priors Require Origin Disputes. *Theory and Decision*, **61**(4), 318–328. 113

Hawthorne, John, & Magidor, Ofra. 2009. Assertion, Context, and Epistemic Accessibility. *Mind*, **118**(470), 377–397. 65, 86, 87

Hawthorne, John, & Magidor, Ofra. 2011. Assertion and Epistemic Opacity. *Mind*, **119**(476), 1087–1105. 65

Heal, Jane. 1978. Common knowledge. *The Philosophical Quarterly*, **28**(111), 116–131. 15, 41

Heifetz, Aviad. 1996. Comment on Consensus without Common Knowledge. *Journal of Economic Theory*, **70**(1), 273–77. 113

Heim, Irene. 1983. On the projection problem for presuppositions. *Pages 114–125 of: West Coast Conference in Linguistics*, vol. 2. 44

Heinemann, Frank, Nagel, Rosemarie, & Ockenfels, Peter. 2004. The Theory of Global Games on Test: Experimental Analysis of Coordination Games with Public and Private Information. *Econometrica*, **72**(5), 1583–1599. 22

Hellman, Ziv. 2012 (June). *Deludedly Agreeing to Agree*. Abstract in TARK 2013; full paper unpublished: http://www.coll.mpg.de/sites/www.coll.mpg.de/files/workshop/DeludedPartitions21.pdf. 137, 149, 161

Hintikka, Jaako. 1962. *Knowledge and Belief*. Cornell University Press. 100, 106

Kajii, Atsushi, & Morris, Stephen. 1997. Common p-Belief: The General Case. *Games and Economic Behavior*, **18**, 73–82. 141

Kamp, Hans, Van Genabith, Josef, & Reyle, Uwe. 2011. Discourse representation theory. *Pages 125–394 of: Handbook of Philosophical Logic*. Springer. 44

Karttunen, Lauri. 1974. Presupposition and linguistic context. *Theoretical linguistics*, **1**(1), 181–194. 43

Kelly, Thomas. 2013. How to be an Epistemic Permissivist. *In:* Greco, John, Steup, Matthias, & Turri, John (eds), *Contemporary Debates in Epistemology*. Blackwell Publishing Ltd. 97

Kinderman, Peter, Dunbar, Robin, & Bentall, Richard P. 1998. Theory-of-mind deficits and causal attributions. *British Journal of Psychology*, **89**(2), 191–204. 45

Kolodny, Niko, & MacFarlane, John. 2010. Ifs and oughts. *The Journal of Philosophy*, **107**(3), 115–143. 108

Kripke, Saul A. 1979. A Puzzle about Belief. *Pages 239–283 of:* Margalit, Avishai (ed), *Meaning and Use*. Dordrecht. 42

Lewis, David. 1975. Languages and language. *Minnesota Studies in the Philosophy of Science*. 66

Lewis, David K. 1969. *Convention: A philosophical study*. Harvard University Press. 18, 102, 141

Lin, Shuhong, Keysar, Boaz, & Epley, Nicholas. 2010. Reflexively mindblind: Using theory of mind to interpret behavior requires effortful attention. *Journal of Experimental Social Psychology*, **46**(3), 551–556. 45

Liu, Qinming. 2010 (unpublished). *Higher-Order Beliefs and Epistemic Conditions for Nash Equilibrium.* 25, 156

McGee, Vann. 1991. We turing machines aren't expected-utility maximizers (even ideally). *Philosophical Studies*, **64**(1), 115–123. 137

Mengel, Friederike. 2012. Learning across games. *Games and Economic Behavior*, **74**(2), 601 – 619. 28

Monderer, Dov, & Samet, Dov. 1989. Approximating Common Knowledge with Common Beliefs. *Games and Economic Behavior*, **1**(2), 170–190. 18, 102, 137, 141

Morris, Stephen. 1995. The Common Prior Assumption in Economic Theory. *Economics and Philosophy*, **11**(2), 227–253. 96

Morris, Stephen. 1996. The logic of belief and belief change: A decision theoretic approach. *Journal of Economic Theory*, **69**(1), 1–23. 135

Moses, Yoram, & Nachum, Gal. 1990. Agreeing to disagree after all. *Pages 151–168 of: Proceedings of the 3rd conference on Theoretical Aspects of Reasoning and Knowledge.* 97, 109

Nagel, Rosemarie. 1995. Unraveling in Guessing Games: An Experimental Study. *American Economic Review*, **85**(5), 1313–26. 17

Parfit, Derek. 1988 (1983). *What we together do.* Unpublished Manuscript. 108

Peacocke, Christopher. 2005. *Joint attention: Its nature, reflexivity, and relation to common knowledge.* Oxford: Oxford University Press. 41, 42

Pearce, David G. 1984. Rationalizable Strategic Behavior and the Problem of Perfection. *Econometrica*, **52**(4), 1029–1050. 27, 35

Pinker, Steven, Nowak, Martin A., & Lee, James J. 2008. The logic of indirect speech. *Proceedings of the National Academy of Sciences*, **105**(3), 833–838. 37

Polak, Ben. 1999. Epistemic Conditions for Nash Equilibrium and Common Knowledge of Rationality. *Econometrica*, **67**, 673–676. 25, 160

Portner, Paul. 2007. Imperatives and modals. *Natural Language Semantics*, **15**(4), 351–383. 44

Pryor, James. 2000. The Skeptic and the Dogmatist. *Noûs*, **34**(4), 517–549. 122

Rubinstein, Ariel. 1989. The Electronic Mail Game: Strategic Behavior Under "Almost Common Knowledge". *The American Economic Review*, **79**(3), 385–391. 21, 22, 24

Rubinstein, Ariel, & Wolinsky, Asher. 1990. On the logic of 'agreeing to disagree' type results. *Journal of Economic Theory*, **51**(1), 184 – 193. 107, 137, 148

Samet, Dov. 1990. Ignoring ignorance and agreeing to disagree. *Journal of Economic Theory*, **52**(1), 190–207. 137, 145, 148

Samet, Dov. 1992. Agreeing to Disagree in Infinite Information Structures. *International Journal of Game Theory*, **21**(3), 213–218. 145

Samet, Dov. 2010. Agreeing to disagree: The non-probabilistic case. *Games and Economic Behavior*, **69**(1), 169 – 174. 108, 109

Savage, Leonard J. 1954. *The Foundations of Statistics*. John Wiley and Sons. 108, 123

Savage, Leonard J. 1967. Difficulties in the Theory of Personal Probability. *Philosophy of Science*, **34**(4), 305–10. 135

Schiffer, Stephen R. 1972. *Meaning*. Oxford: Oxford University Press. 11

Shin, Hyun Song. 1993. Logical structure of common knowledge. *Journal of Economic Theory*, **60**(1), 1–13. 137, 148

Shin, Hyun Song, & Williamson, Timothy. 1994. Representing the knowledge of turing machines. *Theory and Decision*, **37**(1), 125–146. 137

Srinivasan, Amia. 2013. Are We Luminous? *Philosophy and Phenomenological Research*, n/a–n/a. 39

Stahl, Dale, & Wilson, Paul W. 1994. Experimental evidence on players' models of other players. *Journal of Economic Behavior & Organization*, **25**(3), 309–327. 17

Stahl, Dale, & Wilson, Paul W. 1995. On Players' Models of Other Players: Theory and Experimental Evidence. *Games and Economic Behavior*, **10**(1), 218–254. 17

Stalnaker, Robert. 1998. On the representation of context. *Journal of Logic, Language and Information*, **7**(1), 3–19. 43, 53

Stalnaker, Robert. 1999. *Context and Content: Essays on Intentionality in Speech and Thought*. Oxford: Oxford University Press. 86, 102

Stalnaker, Robert. 2002. Common ground. *Linguistics and Philosophy*, **25**(5), 701–721. 43, 53, 57

Stalnaker, Robert. 2006. On Logics of Knowledge and Belief. *Philosophical Studies*, **128**, 169–199. 31

Stalnaker, Robert C. 1974. Pragmatic presuppositions. New York: New York University Press. 43, 57

Stalnaker, Robert C. 1978. Assertion. *Pages 315–32 of:* Cole, P. (ed), *Syntax and Semantics*, vol. 9. New York Academic Press. 43, 86

Stalnaker, Robert C. 2009. On Hawthorne and Magidor on Assertion, Context, and Epistemic Accessibility. *Mind*, **118**(470), 399–409. 12, 30, 65, 86, 92, 106

Stalnaker, Robert C. 2014. *Context*. Oxford: Oxford University Press. 43, 53, 57, 60, 61, 93

Stiller, James, & Dunbar, Robin IM. 2007. Perspective-taking and memory capacity predict social network size. *Social Networks*, **29**(1), 93–104. 45

Strzalecki, Tomasz. 2010. *Depth of reasoning and higher order beliefs*. Harvard Institute of Economic Research, Harvard University. 22, 26

Van Eijck, Jan, & Kamp, Hans. 2011. *Representing discourse in context*. Elsevier. Chap. 3, pages 181–252. 44

Van Eijck, Jan, & Visser, Albert. 2012. Dynamic Semantics. *In:* Zalta, Edward N. (ed), *The Stanford Encyclopedia of Philosophy*, winter 2012 edn. 44

van Fraassen, Bas. 1984. Belief and the Will. *Journal of Philosophy*, **81**, 235–56. 121

Veltman, Frank. 1996. Defaults in update semantics. *Journal of Philosophical Logic*, **25**(3), 221–261. 44

von Fintel, Kai, & Gillies, Anthony. 2011. *Might Made Right*. Oxford University Press. 66

White, Roger. 2005. Epistemic Permissiveness. *Philosophical Perspectives*, **19**, 445–459. 97

White, Roger. 2013. *In:* Greco, John, Steup, Matthias, & Turri, John (eds), *Contemporary Debates in Epistemology*. Blackwell Publishing Ltd. 97

Williamson, Timothy. 1992. Inexact knowledge. *Mind*, **101**(402), 217–242. 117

Williamson, Timothy. 2000. *Knowledge and its Limits*. Oxford: Oxford University Press. 31, 38, 78, 98, 102, 103, 106, 110, 117, 119

Williamson, Timothy. 2007. How probable is an infinite sequence of heads? *Analysis*, **67**(3), 173–180. 97

Williamson, Timothy. 2013. Response to Cohen, Comesaña, Goodman, Nagel, and Weatherson on Gettier Cases in Epistemic Logic. *Inquiry*, **56**(1), 77–96. 76

Yalcin, Seth. 2007. Epistemic modals. *Mind*, **116**(464), 983–1026. 44, 48, 53, 57