Two Definitions of Correlated Equilibrium*

Christian W. Bach** and Andrés Perea***

EPICENTER Working Paper No. 18 (2018)

Abstract. Correlated equilibrium has been introduced by Aumann (1974). Often, in the literature, correlated equilibrium is defined in a simplified as well as more direct way, and sometimes called canonical correlated equilibrium or correlated equilibrium distribution. In fact, we show that the simplified notion of correlated equilibrium is not equivalent – neither doxastically nor behaviourally – to the original from an ex post perspective. We then compare both solution concepts in terms of reasoning. While correlated equilibrium can be characterized by common belief in rationality and a common prior, the simplified variant additionally requires the one-theory-per-choice condition. Since this condition features a correctness of beliefs property, the latter solution concept exhibits a larger degree of Nash equilibrium flavour than the former.

Keywords: Aumann models, canonical correlated equilibrium, common prior, complete information, correlated equilibrium, correlated equilibrium distribution, epistemic characterizations, epistemic game theory, one-theory-per-choice condition, solution concepts, static games.

---

* We are grateful to Pierpaolo Battigalli, Giacomo Bonanno, Amanda Friedenberg, to participants of the Manchester Economic Theory Workshop (MET2018), of the Thirteenth Conference on Logic and the Foundations of Game and Decision Theory (LOFT2018), to seminar participants at Maastricht University, and at the University of Liverpool for useful as well as constructive comments.

** Department of Economics, University of Liverpool Management School, Chatham Street, Liverpool, L69 7ZH, UNITED KINGDOM; EPICENTER, School of Business and Economics, Maastricht University, 6200 MD Maastricht, THE NETHERLANDS. Email: c.w.bach@liverpool.ac.uk

*** Department of Quantitative Economics, School of Business and Economics, Maastricht University, 6200 MD Maastricht, THE NETHERLANDS; EPICENTER, School of Business and Economics, Maastricht University, 6200 MD Maastricht, THE NETHERLANDS. Email: a.perea@maastrichtuniversity.nl

1 Introduction

The solution concept of correlated equilibrium, which has been proposed by Aumann (1974), is constructed within an epistemic framework based on possible worlds, information partitions, and a common prior probability measure. Often, in scientific articles and game theory textbooks, a more direct definition of correlated equilibrium is used that simply formulates correlated equilibrium as a probability measure on choice combinations. The latter solution concept is sometimes called canonical correlated equilibrium (e.g. Forges, 1990) or correlated equilibrium distribution (e.g. Aumann, 1987) in the literature. The question arises whether these two definitions are actually interchangeable or whether they represent two different solution concepts.

The two notions can be compared from two perspectives, which differ with respect to whether information has been received (ex post) or not (ex ante) by the players. It is well-known that from the ex ante perspective correlated equilibrium and canonical correlated equilibrium coincide. More precisely, the induced probability measure on choice combinations of a correlated equilibrium using the common prior only (and not the players' information) is equal to some canonical correlated equilibrium, and vice versa. However, the relevant perspective for reasoning and decision-making in games is ex post. Indeed, the posterior belief of a player about his opponents' choices – conditionalized on his information in the case of correlated equilibrium and conditionalized on one of his choices in the case of canonical correlated equilibrium – constitute the outcome of the player's reasoning. In other words, the players' posterior beliefs represent a solution concept *doxastically*. Optimal choice in line with a player's reasoning then characterizes the respective solution concept *behaviourally*. An appropriate comparison of solution concepts in terms of their game-theoretic semantics thus needs to address these two – doxastic and behavioural – dimensions.

Here, we show that correlated equilibrium and canonical correlated equilibrium are neither doxastically nor behaviourally equivalent. First of all, inspired by the game in Aumann and Dreze's (2008) Figure 2A, we illustrate that correlated equilibrium and canonical correlated equilibrium may induce different sets of first-order beliefs i.e. beliefs about the respective opponents' choice combinations, ex post. Secondly, we construct an example where correlated equilibrium and canonical correlated equilibrium also differ behaviourally, i.e. in terms of optimal choice. Hence, correlated equilibrium and canonical correlated equilibrium constitute two distinct solution concepts for static games. In order to understand their difference an epistemic perspective is pursued with standard type-based epistemic models of games. First of all, transformations from Aumann's epistemic framework to type-based models and back are defined. We show that these transformations turn correlated equilibria into epistemic models that satisfy a common prior assumption as well as contain types expressing common belief in rationality, and vice versa. An epistemic characterization of correlated equilibrium in terms of common belief in rationality and a common prior from an ex post perspective then ensues. We then introduce the epistemic condition of one-theory-per-choice. Intuitively, a reasoner satisfying this condition never

uses in his entire belief hierarchy distinct first-order beliefs to explain the same choice for any player. We give an epistemic characterization of canonical correlated equilibrium in terms of common belief in rationality, a common prior, and the one-theory-per-choice condition from an ex post perspective. In terms of reasoning, canonical correlated equilibrium thus constitutes a more demanding solution concept than correlated equilibrium. Conceptually, the one-theory-per-choice condition contains a correctness of beliefs assumption. Accordingly, the reasoner does not only always explain a given choice by the same first-order belief throughout his entire belief hierarchy, but he also believes his opponents to believe he does so, and he believes his opponents to believe their opponents to believe he does so, etc. Furthermore, the reasoner does not only believe any opponent to explain a given choice by the same first-order belief throughout his entire belief hierarchy, but he also believes his opponents to believe he does so, and he believes his opponents to believe their opponents to believe he does so, etc. Since its epistemic characterization exhibits correctness of beliefs properties, canonical correlated equilibrium has a Nash equilibrium flavour, while Aumann's original solution concept of correlated equilibrium does not.

We proceed as follows. In Section 2, the two definitions of correlated equilibrium within the framework of static games are recalled. It is then shown in Section 3 that the two solution concepts are not equivalent – neither doxastically nor behaviourally. In Section 4, a standard type-based epistemic framework is presented which is later used to analyze correlated equilibrium and canonical correlated equilibrium. Both solution concepts are characterized epistemically in Section 5 and their difference is explained. Finally, some conceptual issues are discussed in Section 6. In particular, the relation to Nash equilibrium is addressed.

2 Preliminaries

A static game is modelled as a tuple $\Gamma = \big(I, (C_i)_{i\in I}, (U_i)_{i\in I}\big)$, where I is a finite set of players, C_i denotes player i's finite choice set, and $U_i : \times_{j\in I} C_j \to \mathbb{R}$ is player i's utility function, which assigns a real number $u_i(c)$ to every choice combination $c \in \times_{j\in I} C_j$. For the class of static games the solution concept of correlated equilibrium has been introduced by Aumann (1974) and given an epistemic foundation in terms of universal rationality and a common prior from an ex ante perspective by Aumann (1987).[1] Loosely speaking, in a correlated equilibrium the players' choices are required to satisfy a best response property given a probability measure on the opponents' choice combinations derived from a common prior via Bayesian updating within some information structure.

[1] Note that Aumann (1987) actually gives an epistemic characterization of canonical correlated equilibrium from an ex ante perspective. However, since correlated equilibrium and canonical correlated equilibrium are equivalent from an ex ante perspective, Aumann's (1987) epistemic characterization also applies to correlated equilibrium.

In fact, the notion of correlated equilibrium is embedded in the epistemic framework of Aumann models, which describe the players' knowledge and beliefs in terms of information partitions. Formally, an *Aumann model* of a game Γ is a tuple $\mathcal{A}^\Gamma = \big(\Omega, \pi, (\mathcal{I}_i)_{i\in I}, (\sigma_i)_{i\in I}\big)$, where Ω is a finite set of all possible worlds, $\pi \in \Delta(\Omega)$ is a common prior probability measure on the set of all possible worlds, $\mathcal{I}_i$ is an information partition on Ω for every player $i \in I$ such that $\pi\big(\mathcal{I}_i(\omega)\big) > 0$ for all $\omega \in \Omega$, with $\mathcal{I}_i(\omega)$ denoting the cell of $\mathcal{I}_i$ containing ω, and $\sigma_i : \Omega \to C_i$ is an $\mathcal{I}_i$-measurable choice function for every player $i \in I$. Conceptually, the $\mathcal{I}_i$-measurability of σ_i ensures that i entertains no uncertainty whatsoever about his own choice, i.e. $\sigma_i(\omega') = \sigma_i(\omega)$ for all $\omega' \in \mathcal{I}(\omega)$. Note that beliefs of players are explicitly expressible in Aumann models of games. Indeed, beliefs are obtained via Bayesian conditionalization on the common prior given the respective player's information. More precisely, an event $E \subseteq \Omega$ consists of possible worlds, and player i's belief in E at a world ω is defined as $b_i(E, \omega) := \pi\big(E \mid \mathcal{I}_i(\omega)\big) = \frac{\pi\left(E \cap \mathcal{I}_i(\omega)\right)}{\pi\left(\mathcal{I}_i(\omega)\right)}$. For instance, given a choice combination s_{-i} of player i's opponents, the set $\{\omega \in \Omega : \sigma_j(\omega) = s_j \text{ for all } j \in I \setminus \{i\}\}$ denotes the event that i's opponents play according to s_{-i}.

Within the framework of Aumann models and in line with Aumann (1974), the notion of correlated equilibrium – sometimes also called objective correlated equilibrium – is formally defined as follows.

Definition 1. *Let Γ be a game, and $\mathcal{A}^\Gamma$ an Aumann model of it with choice functions $\sigma_i : \Omega \to C_i$ for every player $i \in I$. The tuple $(\sigma_i)_{i\in I}$ of choice functions constitutes a* correlated equilibrium*, if for every player $i \in I$, and for every world $\omega \in \Omega$, it is the case that*

$$\sum_{\omega' \in \mathcal{I}_i(\omega)} \pi\big(\omega' \mid \mathcal{I}_i(\omega)\big) \cdot U_i\big(\sigma_i(\omega), \sigma_{-i}(\omega')\big) \geq \sum_{\omega' \in \mathcal{I}_i(\omega)} \pi\big(\omega' \mid \mathcal{I}_i(\omega)\big) \cdot U_i\big(c_i, \sigma_{-i}(\omega')\big)$$

for every choice $c_i \in C_i$.

Intuitively, a choice function tuple constitutes a correlated equilibrium, if for every player, the choice function specifies at every world a best response given the common prior conditionalized on the player's information and given the opponents' choice functions.

Aumann structures induce for every player a probability measure at every world about the respective opponents' choices – typically called first-order belief – via an appropriate projection of the conditionalized common prior. Given a game Γ a first-order belief $\beta_i \in \Delta(C_{-i})$ of some player $i \in I$ is *possible in a correlated equilibrium*, if there there exists an Aumann model $\mathcal{A}^\Gamma$ of Γ such that the tuple $(\sigma_j)_{j\in I}$ constitutes a correlated equilibrium and with some world $\hat{\omega} \in \Omega$ such that

$$\beta_i(c_{-i}) = \pi\big(\{\omega' \in \mathcal{I}_i(\hat{\omega}) : \sigma_{-i}(\omega') = c_{-i}\} \mid \mathcal{I}_i(\hat{\omega})\big)$$

for all $c_{-i} \in C_{-i}$.

From a behavioural viewpoint it is ultimately of interest what choices a player can make given a particular line of reasoning and decision-making fixed by specific epistemic assumptions or by a specific solution concept. Formally, given a game Γ a choice $c_i^* \in C_i$ of some player $i \in I$ is *optimal in a correlated equilibrium*, if there exists an Aumann model $\mathcal{A}^\Gamma$ of Γ such that the tuple $(\sigma_j)_{j \in I}$ constitutes a correlated equilibrium and with some world $\hat{\omega} \in \Omega$ such that

$$\sum_{\omega' \in \mathcal{I}_i(\hat{\omega})} \pi\big(\omega' \mid \mathcal{I}_i(\hat{\omega})\big) \cdot U_i\big(c_i^*, \sigma_{-i}(\omega')\big) \geq \sum_{\omega' \in \mathcal{I}_i(\hat{\omega})} \pi\big(\omega' \mid \mathcal{I}_i(\hat{\omega})\big) \cdot U_i\big(c_i, \sigma_{-i}(\omega')\big)$$

for all $c_i \in C_i$.

Often, in the literature and in textbooks, the following more direct – and simpler – definition of correlated equilibrium is used.

Definition 2. *Let Γ be a game, and $\rho \in \Delta(\times_{i \in I} C_i)$ a probability measure on the players' choice combinations. The probability measure ρ constitutes a* canonical correlated equilibrium*, if for every player $i \in I$, and for every choice $c_i \in C_i$ of player i such that $\rho(c_i) > 0$, it is the case that*

$$\sum_{c_{-i} \in C_{-i}} \rho(c_{-i} \mid c_i) \cdot U_i(c_i, c_{-i}) \geq \sum_{c_{-i} \in C_{-i}} \rho(c_{-i} \mid c_i) \cdot U_i(c_i', c_{-i})$$

for every choice $c_i' \in C_i$.

Intuitively, a probability measure on the players' choice combinations constitutes a canonical correlated equilibrium, if every choice that receives positive probability is optimal given the probability measure conditionalized on the very choice itself.

Also, the solution concept of canonical correlated equilibrium naturally induces for every player a first-order belief for each of his choices via Bayesian conditionalization. Given a game Γ, a first-order belief $\beta_i \in \Delta(C_{-i})$ of some player $i \in I$ is *possible in a canonical correlated equilibrium*, if there there exists a canonical correlated equilibrium $\rho \in \Delta(\times_{j \in I} C_j)$ and a choice $\hat{c}_i \in C_i$ of player i with $\rho(\hat{c}_i) > 0$ such that

$$\beta_i(c_{-i}) = \rho(c_{-i} \mid \hat{c}_i)$$

for all $c_{-i} \in C_{-i}$.

Finally, optimal choice with a canonical correlated equilibrium also needs to be fixed in order to relate the two definitions of correlated equilibrium behaviourally. Formally, given a game Γ, a choice $c_i^* \in C_i$ of some player $i \in I$ is *optimal in a canonical correlated equilibrium*, if there exists a canonical correlated equilibrium $\rho \in \Delta(\times_{j \in I} C_j)$ and a choice $\hat{c}_i \in C_i$ of player i with $\rho(\hat{c}_i) > 0$ such that

$$\sum_{c_{-i} \in C_{-i}} \rho(c_{-i} \mid \hat{c}_i) \cdot U_i(c_i^*, c_{-i}) \geq \sum_{c_{-i} \in C_{-i}} \rho(c_{-i} \mid \hat{c}_i) \cdot U_i(c_i', c_{-i})$$

for all $c_i' \in C_i$.

3 Difference of the Two Definitions

With two notions of correlated equilibrium existing in the literature the natural question emerges whether they are equivalent or not from an ex post perspective. The two solution concepts can be compared doxastically as well as behaviourally.

Suppose that a first-order belief $\beta_i \in \Delta(C_{-i})$ is possible in a canonical correlated equilibrium of some game Γ, i.e. $\beta_i(c_{-i}) = \rho(c_{-i} \mid \hat{c}_i)$ for all $c_{-i} \in C_{-i}$ for some canonical correlated equilibrium $\rho \in \Delta(\times_{j \in I} C_j)$ of Γ and for some choice $\hat{c}_i \in C_i$ with $\rho(\hat{c}_i) > 0$. Construct an Aumann structure $\mathcal{A}^\Gamma$ with $\Omega := \{\omega^{(c_j)_{j \in I}} : (c_j)_{j \in I} \in \times_{j \in I} C_j \text{ such that } \rho\big((c_j)_{j \in I}\big) > 0\}$, $\mathcal{I}_j := \big\{\{\omega^{(c_j, c_{-j})} \in \Omega : c_{-j} \in C_{-j}\} : c_j \in C_j \text{ with } \rho(c_j) > 0\big\}$ for all $j \in I$, $\pi\big(\omega^{(c_j)_{j \in I}}\big) := \rho\big((c_j)_{j \in I}\big)$ for all $\omega^{(c_j)_{j \in I}} \in \Omega$, and $\sigma_j(\omega^{(c_k)_{k \in I}}) = c_j$ for all $\omega^{(c_k)_{k \in I}} \in \Omega$ and for all $j \in I$. As ρ constitutes a canonical correlated equilibrium, observe that

$$\sum_{\omega \in \mathcal{I}_i(\omega^{(\hat{c}_i, c_{-i})})} \pi\big(\omega \mid \mathcal{I}_i(\omega^{(\hat{c}_i, c_{-i})})\big) \cdot U_i\big(\sigma_i(\omega^{(\hat{c}_i, c_{-i})}), \sigma_{-i}(\omega)\big)$$

$$= \sum_{c_{-i} \in C_{-i}} \rho(c_{-i} \mid \hat{c}_i) \cdot U_i(c_i, c_{-i}) \geq \sum_{c_{-i} \in C_{-i}} \rho(c_{-i} \mid \hat{c}_i) \cdot U_i(c_i', c_{-i})$$

$$= \sum_{\omega \in \mathcal{I}_i(\omega^{(\hat{c}_i, c_{-i})})} \pi\big(\omega \mid \mathcal{I}_i(\omega^{(\hat{c}_i, c_{-i})})\big) \cdot U_i\big(c_i', \sigma_{-i}(\omega)\big)$$

holds for every choice $c_i' \in C_i$ and for every player $i \in I$, i.e. $(\sigma_j)_{j \in I}$ constitutes a correlated equilibrium. It is also the case that $\rho(c_{-i} \mid \hat{c}_i) = \pi\Big(\{\omega \in \mathcal{I}_i\big(\omega^{(\hat{c}_i, c_{-i})}\big) : \sigma_{-i}\big(\omega\big) = c_{-i}\} \mid \mathcal{I}_i\big(\omega^{(\hat{c}_i, c_{-i})}\big)\Big)$. Consequently, the following remark obtains.

Remark 1. Let Γ be a static game, $i \in I$ some player, and $\beta_i^* \in \Delta(C_{-i})$ some first-order belief of player i. If β_i^* is possible in a canonical correlated equilibrium, then β_i^* is possible in a correlated equilibrium.

The definition of optimal choice in a solution concept together with Remark 1 directly implies that optimality in a canonical correlated equilibrium implies optimality in a correlated equilibrium.

Remark 2. Let Γ be a static game, $i \in I$ some player, and $c_i^* \in C_i$ some choice of player i. If c_i^* is optimal in a canonical correlated equilibrium, then c_i^* is optimal in a correlated equilibrium.

However, it is now shown by means of an example that the converse of Remark 1 does not hold.

Example 1. Consider the two player game between *Rowena* and *Colin* depicted in Figure 1, which is due to Aumann and Dreze (2008, Figure 2A).[2]

Let $\big(\Omega, \pi, (\mathcal{I}_i)_{i \in I}, (\sigma_i)_{i \in I}\big)$ be an Aumann model of the game, where

[2] In fact, Aumann and Dreze (2008) use the game depicted in Figure 1 to show that *Rowena*'s expected payoff in a canonical correlated equilibrium can be different if

		Colin		
		L	C	R
Rowena	T	$0, 0$	$4, 5$	$5, 4$
	M	$5, 4$	$0, 0$	$4, 5$
	B	$4, 5$	$5, 4$	$0, 0$

Fig. 1. A two player static game between *Rowena* and *Colin*.

- $I = \{Rowena, Colin\}$,
- $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6, \omega_7\}$,
- $\pi \in \Delta(\Omega)$ with $\pi(\omega_1) = \pi(\omega_3) = \frac{1}{12}$ and $\pi(\omega) = \frac{1}{6}$ for all $\omega \in \Omega \setminus \{\omega_1, \omega_3\}$,
- $\mathcal{I}_{Rowena} = \big\{\{\omega_1\}, \{\omega_2, \omega_3\}, \{\omega_4, \omega_5\}, \{\omega_6, \omega_7\}\big\}$,
- $\mathcal{I}_{Colin} = \big\{\{\omega_1, \omega_3, \omega_5\}, \{\omega_2, \omega_7\}, \{\omega_4, \omega_6\}\big\}$,
- $\sigma_{Rowena}(\omega_1) = \sigma_{Rowena}(\omega_2) = \sigma_{Rowena}(\omega_3) = T$, $\sigma_{Rowena}(\omega_4) = \sigma_{Rowena}(\omega_5) = M$, and $\sigma_{Rowena}(\omega_6) = \sigma_{Rowena}(\omega_7) = B$,
- $\sigma_{Colin}(\omega_1) = \sigma_{Colin}(\omega_3) = \sigma_{Colin}(\omega_5) = R$, $\sigma_{Colin}(\omega_2) = \sigma_{Colin}(\omega_7) = C$, and $\sigma_{Colin}(\omega_4) = \sigma_{Colin}(\omega_6) = L$.

Observe that $(\sigma_i)_{i \in I}$ constitutes a correlated equilibrium of the game. Also, the first-order belief $\beta^*_{Rowena} \in \Delta(C_{Colin})$ of *Rowena* such that $\beta^*_{Rowena}(R) = 1$ is possible in a correlated equilibrium, as $\mathcal{I}_{Rowena}(\omega_1) = \{\omega_1\}$ and $\sigma_{Colin}(\omega_1) = R$.

Suppose that there exists a canonical correlated equilibrium $\rho \in \Delta(C_{Rowena} \times C_{Colin})$ with $\rho(\cdot \mid c_{Rowena}) = \beta^*_{Rowena}$ for some $c_{Rowena} \in C_{Rowena}$ such that $\rho(c_{Rowena}) > 0$. Since c_{Rowena} is optimal for $\rho(\cdot \mid c_{Rowena}) = \beta^*_{Rowena}$, it is the case that $c_{Rowena} = T$. Hence, $\rho(\cdot \mid T) = \beta^*_{Rowena}$ and thus $\rho(R \mid T) = 1$. Consequently, $\rho(T, R) > 0$ as well as $\rho(T, L) = \rho(T, C) = 0$. Then, $\rho(M, C) = \rho(B, C) = 0$, as otherwise C is strictly dominated by L on $\{M, B\}$, contradicting the optimality of C given $\rho(\cdot \mid C) \in \Delta(\{M, B\})$. Then, $\rho(B, L) = \rho(B, R) = 0$, as otherwise B is strictly dominated by M on $\{L, R\}$, contradicting the optimality of B given $\rho(\cdot \mid B) \in \Delta(\{L, R\})$. Then, $\rho(M, L) = 0$, as otherwise L is strictly dominated by R on $\{M\}$, contradicting the optimality of L given $\rho(\cdot \mid L) \in \Delta(\{M\})$. Then, $\rho(M, R) = 0$, as otherwise M is strictly dominated by T on $\{R\}$, contradicting the optimality of M given $\rho(\cdot \mid M) \in \Delta(\{R\})$. Therefore, it is the case that $\rho(T, R) = 1$. However, R is not optimal given $\rho(\cdot \mid R)$, a contradiction. Hence, the first-order belief $\beta^*_{Rowena} \in \Delta(C_{Colin})$ of *Rowena* such that $\beta^*_{Rowena}(R) = 1$ is not possible in a canonical correlated equilibrium. ♣

The preceding example establishes the following remark.

Remark 3. There exists a game Γ, a player $i \in I$, and a first-order belief $\beta^*_i \in \Delta(C_{-i})$ of player i such that β^*_i is possible in a correlated equilibrium but β^*_i is not possible in a canonical correlated equilibrium.

---

the game is doubled in the sense that each of her choices are listed twice. The game is thus changed but only the solution concept of canonical correlated equilibrium is considered. Here, we keep the game fixed, but switch between the solution concepts of correlated equilibrium and canonical correlated equilibrium.

Actually, in Example 1 the induced optimal choices are equal for both solution concepts despite their difference in terms of possible first-order beliefs. Indeed, observe that $\rho \in \Delta(C_{Rowena} \times C_{Colin})$ with $\rho(c) = \frac{1}{9}$ for all $c \in C_{Rowena} \times C_{Colin}$ constitutes a canonical correlated equilibrium of the game depicted in Figure 1 and for every player it is the case that every choice is optimal in ρ. Also, the correlated equilibrium $(\sigma_i)_{i \in I}$ of this game from Example 1 exhibits the property that for every player it is the case that every choice is optimal.

Yet, both definitions of correlated equilibrium can also be distinct in terms of induced optimal choice as the next example shows.

Example 2. Consider the two player game between *Alice* and *Bob* depicted in Figure 2.

		Bob			
		e	f	g	h
Alice	a	$1,1$	$2,3$	$3,2$	$0,1$
	b	$3,2$	$1,1$	$2,3$	$2,2$
	c	$2,3$	$3,2$	$1,1$	$1,3$
	d	$3,0$	$0,0$	$0,0$	$0,1$

Fig. 2. A two player static game between *Alice* and *Bob*.

Suppose the Aumann model $\big(\Omega, \pi, (\mathcal{I}_i)_{i \in I}, (\hat{\sigma})_{i \in I}\big)$ of the game, where

- $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6, \omega_7\}$,
- $\pi(\omega_1) = \pi(\omega_2) = \pi(\omega_5) = \pi(\omega_6) = \pi(\omega_7) = \frac{1}{6}$ and $\pi(\omega_3) = \pi(\omega_4) = \frac{1}{12}$,
- $\mathcal{I}_{Alice} = \{\{\omega_1, \omega_2\}, \{\omega_3\}, \{\omega_4, \omega_5\}, \{\omega_6, \omega_7\}\}$,
- $\mathcal{I}_{Bob} = \{\{\omega_3, \omega_4, \omega_6\}, \{\omega_1, \omega_7\}, \{\omega_2, \omega_5\}\}$,
- $\sigma_{Alice}(\omega_1) = \sigma_{Alice}(\omega_2) = a$, $\sigma_{Alice}(\omega_3) = \sigma_{Alice}(\omega_4) = \sigma_{Alice}(\omega_5) = b$, and $\sigma_{Alice}(\omega_6) = \sigma_{Alice}(\omega_7) = c$,
- $\sigma_{Bob}(\omega_1) = \sigma_{Bob}(\omega_7) = f$, $\sigma_{Bob}(\omega_2) = \sigma_{Bob}(\omega_5) = g$, and $\sigma_{Bob}(\omega_3) = \sigma_{Bob}(\omega_4) = \sigma_{Bob}(\omega_6) = e$.

Observe that $(\sigma_{Alice}, \sigma_{Bob})$ constitute a correlated equilibrium. Also, the choice d of *Alice* is optimal in the correlated equilibrium $(\sigma_{Alice}, \sigma_{Bob})$, since d is optimal for *Alice* at world ω_3.

However, it is now shown that d cannot be optimal in a canonical correlated equilibrium. Towards a contradiction, suppose that there exists a canonical correlated equilibrium $\rho \in \Delta(C_{Alice} \times C_{Bob})$, for which d is optimal. Then, $\rho(e \mid c_1) = 1$ for some choice $c_1 \in C_{Alice}$ with $\rho(c_1) > 0$, as otherwise c would be strictly better than d for *Alice*. Since c_1 needs to be optimal for $\rho(\cdot \mid c_1)$, it must be the case that $c_1 = b$ or $c_1 = d$.

Suppose that $c_1 = d$. Then, $\rho(e \mid d) = 1$ implies that $\rho(e) > 0$, which in turn implies that e is optimal for $\rho(\cdot \mid e)$. As $\rho(d \mid e) > 0$, the choice h is thus better than e, a contradiction.

Alternatively, suppose that $c_1 = b$, and thus $\rho(e \mid b) = 1$. It has to be the case that $\rho(d) = 0$, as otherwise d is optimal for $\rho(\cdot \mid d)$, hence $\rho(e \mid d) = 1$, a contradiction. Because $\rho(d) = 0$ and $\rho(e \mid b) = 1$, it follows that $\rho(b, g) = 0$ as well as $\rho(d, g) = 0$. Therefore, $\rho(b \mid g) = \rho(d \mid g) = 0$ if $\rho(g) > 0$. Yet, if $\rho(g) > 0$, then f is better than g against $\rho(\cdot \mid g)$, because in that case $\rho(b \mid g) = \rho(d \mid g) = 0$. This is a contradiction, and thus $\rho(g) = 0$. Consequently, if $\rho(a) > 0$, then $\rho(g \mid a) = 0$, and thus c is better than a against $\rho(\cdot \mid a)$, a contradiction, hence $\rho(a) = 0$.

Since $\rho(a) = \rho(d) = 0$ as well as $\rho(e \mid b) = 1$, it is the case that $\rho(a, f) = \rho(d, f) = \rho(b, f) = 0$, and therefore $\rho(c \mid f) = 1$ if $\rho(f) > 0$. But then, if $\rho(f) > 0$, the choice e is better than f against $\rho(\cdot \mid f)$, a contradiction, and thus $\rho(f) = 0$.

As $\rho(f) = \rho(g) = 0$, it is the case that $\rho(f \mid c) = \rho(g \mid c) = 0$ if $\rho(c) > 0$. Hence, if $\rho(c) > 0$, the choice b is better than c against $\rho(\cdot \mid c)$, a contradiction, and thus $\rho(c) = 0$.

Since $\rho(a) = \rho(c) = \rho(d) = 0$ as well as $\rho(e \mid b) = 1$, it is the case that $\rho(b, e) = 1$. But then $\rho(b \mid e) = 1$, and thus g is better than e against $\rho(\cdot \mid e)$, a contradiction.

Consequently, there exists no canonical correlated equilibrium for which d is optimal. ♣

Thus, the following remark ensues.

Remark 4. There exists a game Γ, some player $i \in I$, and some choice $c_i^* \in C_i$ of player i such that c_i^* is optimal in a correlated equilibrium but c_i^* is not optimal in a canonical correlated equilibrium.

Due to Remarks 3 and 4 correlated equilibrium and canonical correlated equilibrium differ both doxastically as well as behaviourally. Hence, the two notions actually constitute genuinely distinct solution concepts for static games.

4 Epistemic Models

Reasoning in games is usually modelled by belief hierarchies about the underlying space of uncertainty. Due to Harsanyi (1967-68) types can be used as implicit representations of belief hierarchies. The notion of an epistemic model provides the framework to formally describe reasoning in games.

Definition 3. *Let Γ be a static game. An* epistemic model *of Γ is a tuple $\mathcal{M}^\Gamma = \big((T_i)_{i \in I}, (b_i)_{i \in I}\big)$, where for every player $i \in I$*

- *T_i is a finite set of types,*
- *$b_i : T_i \to \Delta(C_{-i} \times T_{-i})$ assigns to every type $t_i \in T_i$ a probability measure $b_i[t_i]$ on the set of opponents' choice type combinations.*

Given a game and an epistemic model of it, belief hierarchies, marginal beliefs, as well as marginal belief hierarchies can be derived from every type. For instance, every type $t_i \in T_i$ induces a belief on the opponents' choice combinations by marginalizing the probability measure $b_i[t_i]$ on the space C_{-i}. Note that no additional notation is introduced for marginal beliefs, in order to keep notation as sparse as possible. It should always be clear from the context which belief $b_i[t_i]$ refers to.

Besides, we follow a one-player perspective approach, which considers game theory as an interactive extension of decision theory. Accordingly, all epistemic concepts – including iterated ones – are defined as mental states inside the mind of a single person. A one-player approach seems natural in the sense that reasoning is formally represented by epistemic concepts and any reasoning process prior to choice does indeed take place entirely *within* the reasoner's mind. Formally, this approach is parsimonious in the sense that states, describing the beliefs of all players, do not have to be invoked in epistemic models of games.

Some further notions and notation are now introduced. For that purpose consider a game Γ, an epistemic model $\mathcal{M}^\Gamma$ of it, and fix two players $i, j \in I$ such that $i \neq j$.

A type $t_i \in T_i$ is said to *deem possible* some choice type combination (c_{-i}, t_{-i}) of his opponents, if $b_i[t_i]$ assigns positive probability to (c_{-i}, t_{-i}). Analogously, a type $t_i \in T_i$ deems possible some opponent type $t_j \in T_j$, if $b_i[t_i]$ assigns positive probability to t_j.

For each choice type combination (c_i, t_i), the *expected utility* is given by

$$u_i(c_i, t_i) = \sum_{c_{-i} \in C_{-i}} \big(b_i[t_i](c_{-i}) \cdot U_i(c_i, c_{-i})\big).$$

Intuitively, the common prior assumption in economics states that every belief in models with multiple agents is derived from a single probability distribution, the so-called common prior. In the epistemic framework of Definition 3 all beliefs are furnished by the types. The common prior assumption thus imposes a condition on the types, requiring all beliefs to be derived from a single probability distribution on the basic space of uncertainty and the players' types.

Definition 4. *Let Γ be a static game, and $\mathcal{M}^\Gamma$ an epistemic model of it. The epistemic model $\mathcal{M}^\Gamma$ satisfies the* common prior assumption*, if there exists a probability measure $\varphi \in \Delta\big(\times_{j \in I}(C_j \times T_j)\big)$ such that for every player $i \in I$, and for every type $t_i \in T_i$ it is the case that $\varphi(t_i) > 0$ and*

$$b_i[t_i](c_{-i}, t_{-i}) = \frac{\varphi(c_i, c_{-i}, t_i, t_{-i})}{\varphi(c_i, t_i)}$$

for all $c_i \in C_i$ with $\varphi(c_i, t_i) > 0$, and for all $(c_{-i}, t_{-i}) \in C_{-i} \times T_{-i}$. The probability measure φ is called common prior*.*

Accordingly, every type's induced belief function obtains from a single probability measure – the common prior – via Bayesian updating. Note that the

common prior is defined on the full space of uncertainty, i.e. on the set of all the players' choice type combinations, while belief functions are defined on the space of respective opponents' choice type combinations only. The common prior assumption could be interpreted by means of an interim stage set-up, in which every player $i \in I$ observes the pair (c_i, t_i) on which he then conditionalizes. Moreover, note that our common prior assumption according to Definition 4 is equivalent to the conjunction of Dekel and Siniscalchi's (2015) Definition 12.13 with their Definition 12.15.

Intuitively, an optimal choice yields at least as much payoff as all other options, given what the player believes his opponents to choose. Formally, optimality is a property of choices given a type. A choice $c_i^* \in C_i$ is said to be *optimal* for the type t_i, if

$$u_i(c_i^*, t_i) \geq u_i(c_i, t_i)$$

for all $c_i \in C_i$.

A player believes in rationality, if he only deems possible choice type pairs – for each of his opponents – such that the choice is optimal for the respective type. Formally, a type $t_i \in T_i$ is said to *believe in rationality*, if t_i only deems possible choice type combinations $(c_{-i}, t_{-i}) \in C_{-i} \times T_{-i}$ such that c_j is optimal for t_j for every opponent $j \in I \setminus \{i\}$. Note that belief in rationality imposes restrictions on the first two layers of a player's belief hierarchy, since the player's belief about his opponents' choices as well as the player's belief about his opponents' beliefs about their respective opponents' choices are affected.

The conditions on interactive reasoning can be taken to further – arbitrarily high – layers in belief hierarchies.

Definition 5. *Let Γ be a static game, $\mathcal{M}^\Gamma$ an epistemic model of it, and $i \in I$ some player.*

- *A type $t_i \in T_i$ expresses* 1-fold belief in rationality*, if t_i believes in rationality.*
- *A type $t_i \in T_i$ expresses* k-fold belief in rationality *for some $k > 1$, if t_i only deems possible types $t_j \in T_j$ for all $j \in I \setminus \{i\}$ such that t_j expresses $k-1$-fold belief in rationality.*
- *A type $t_i \in T_i$ expresses* common belief in rationality*, if t_i expresses k-fold belief in rationality for all $k \geq 1$.*

A player satisfying common belief in rationality entertains a belief hierarchy in which the rationality of all players is not questioned at any level. Observe that if an epistemic model for every player only contains types that believe in rationality, then every type also expresses common belief in rationality. This fact is useful when constructing epistemic models with types expressing common belief in rationality.

Consider two players $i \in I$ and $j \in I$ not necessarily distinct. A type t_j of player j is called *belief-reachable* from a type t_i of player i, if there exists a finite sequence $(t^1, \ldots, t^N)$ of types with $N \in \mathbb{N}$, where $t^{n+1} \in \text{supp}(b_k[t^n])$ such that $t^n \in T_k$ for all $n \in \{1, \ldots, N-1\}$, and $t^1 = t_i$ as well as $t^N = t_j$. Intuitively, if a type t_j is belief-reachable from a type t_i, the former is not excluded in the

interactive reasoning by the latter. The set $T_j(t_i)$ contains all belief-reachable types of player j from t_i. Similarly, a choice type pair $(c_j, t_j) \in C_j \times T_j$ is called *belief-reachable* from t_i, if there exists a finite sequence $(t^1, \ldots, t^N)$ of types with $N \in \mathbb{N}$, where $t^{n+1} \in \text{supp}(b_k[t^n])$ for some $k \in I$ such that $t^n \in T_k$ for all $n \in \{1, \ldots, N-1\}$, $t^1 = t_i$ as well as $t^N = t_j$, and $b_k(t^{N-1})(c_j, t_j) > 0$. The set of belief-reachable choice type pairs of player j from t_i is denoted by $(C_j \times T_j)(t_i)$. Intuitively, if a choice type pair (c_j, t_j) is belief-reachable from a type t_i, the former is not excluded in the interactive reasoning by the latter.

The following lemma ensures that belief reachability preserves common belief in rationality.

Lemma 1. *Let Γ be a static game, $\mathcal{M}^\Gamma$ an epistemic model of it, $i, j \in I$ some players, $t_i \in T_i$ a type of player i, and $t_j \in T_j$ a type of player j. If t_i expresses common belief in rationality and t_j is belief reachable from t_i, then t_j expresses common belief in rationality.*

Proof. Assume that t_j is belief reachable from t_i in $N > 1$ steps, i.e. there exists a finite sequence $(t^1, \ldots, t^N)$ of types with $t^{n+1} \in \text{supp}(b_k[t^n])$ as well as $t^1 = t_i$ and $t^N = t_j$. Towards a contradiction suppose that t_j does not express common belief in rationality. Then, there exists $k > 0$ such that t_j does not express k-fold belief in rationality. However, as t_i deems possible t_j at the N-level of its induced belief hierarchy, t_i thus violates $(N + k)$-fold belief in rationality and a fortiori common belief in rationality, a contradiction. ■

The choice rule of rationality and the reasoning concept of common belief in rationality give rational choice under common belief in rationality. More precisely, a choice $c_i^* \in C_i$ is said to be *rational under common belief in rationality*, if there exists an epistemic model $\mathcal{M}^\Gamma$ of Γ with a type $t_i \in T_i$ of i such that c_i^* is optimal for t_i and t_i expresses common belief in rationality. Similarly, a choice $c_i^* \in C_i$ is said to be *rational under common belief in rationality with a common prior*, if there exists an epistemic model $\mathcal{M}^\Gamma$ of Γ satisfying the common prior assumption with a type $t_i \in T_i$ of i such that c_i^* is optimal for t_i and t_i expresses common belief in rationality. Besides, a first-order belief $\beta_i^* \in \Delta(C_{-i})$ is said to be *possible under common belief in rationality with a common prior*, if there exists an epistemic model $\mathcal{M}^\Gamma$ of Γ satisfying the common prior assumption with a type $t_i \in T_i$ of i such that $b_i[t_i](c_{-i}) = \beta_i^*(c_{-i})$ for all $c_{-i} \in C_{-i}$ and t_i expresses common belief in rationality

5 Epistemic Comparison of the Two Definitions

Before the two solution concepts of correlated equilibrium and canonical correlated equilibrium are juxtaposed epistemically, the structural relationship between Aumann models and epistemic models is investigated.

On the one hand, epistemic models can be derived from Aumann models as follows.

Definition 6. *Let Γ be a static game, and $\mathcal{A}^\Gamma$ an Aumann model of Γ. For every player $i \in I$, construct a set $T_i := \{t_i^{P_i} : P_i \in \mathcal{I}_i\}$, a function $\eta_i : \Omega \to T_i$ such that $\eta_i(\omega) = t_i^{\mathcal{I}_i(\omega)}$ for all $\omega \in \Omega$, a function $b_i : T_i \to \Delta(C_{-i} \times T_{-i})$ such that $b_i[t_i^{P_i}](c_{-i}, t_{-i}) = \sum_{\omega \in P_i : \sigma_{-i}(\omega) = c_{-i}, \eta_{-i}(\omega) = t_{-i}} \pi(\omega \mid P_i)$ for all $(c_{-i}, t_{-i}) \in C_{-i} \times T_{-i}$ and for all $t_i^{P_i} \in T_i$. The epistemic model $\eta(\mathcal{A}^\Gamma)$ of Γ thus obtained is called the $\mathcal{A}^\Gamma$-*induced epistemic model *of Γ.*

Accordingly, based on an Aumann model the functions η_i for every player $i \in I$ provide the ingredients for an epistemic model. In particular, these epistemic models satisfy the common prior assumption as will – among other things – be shown below in Theorem 1.

Conversely, epistemic models with a common prior also induce Aumann models.

Definition 7. *Let Γ be a static game, and $\mathcal{M}^\Gamma$ an epistemic model of Γ satisfying the common prior assumption with common prior φ. Construct a set $\Omega := \{\omega^{(c_i, t_i)_{i \in I}} : c_i \in C_i, t_i \in T_i \text{ for all } i \in I \text{ such that } \varphi\big((c_i, t_i)_{i \in I}\big) > 0\}$, a function $\pi \in \Delta(\Omega)$ such that $\pi\big(\omega^{(c_i, t_i)_{i \in I}}\big) = \varphi\big((c_i, t_i)_{i \in I}\big)$ for all $\omega^{(c_i, t_i)_{i \in I}} \in \Omega$, as well as for every player $i \in I$ a function $\sigma_i : \Omega \to C_i$ such that $\sigma_i\big(\omega^{(c_j, t_j)_{j \in I}}\big) = c_i$ for all $\omega^{(c_j, t_j)_{j \in I}} \in \Omega$, and a partition $\mathcal{I}_i$ of Ω such that $\mathcal{I}_i\big(\omega^{(c_j, t_j)_{j \in I}}\big) = \{\omega^{(c_i, t_i, c'_{-i}, t'_{-i})} \in \Omega : c'_{-i} \in C_{-i}, t'_{-i} \in T_{-i}\}$ for all $\omega^{(c_j, t_j)_{j \in I}} \in \Omega$. The Aumann model $\theta(\mathcal{M}^\Gamma)$ of Γ thus obtained is called the $\mathcal{M}^\Gamma$-*induced* Aumann model *of Γ.*

Note that given some game Γ, the structure $\eta(\mathcal{A}^\Gamma)$ can be expressed as the image of a function from the collection of all Aumann models of Γ as domain to the collection of all epistemic models of Γ as range, and the structure $\theta(\mathcal{M}^\Gamma)$ can be expressed as the image of a function from the collection of all epistemic models for Γ satisfying the common prior assumption as domain to the collection of all Aumann models of Γ as range.

It is now shown that the transformations between Aumann models and epistemic models connect correlated equilibrium with common belief in rationality and a common prior.

Theorem 1. *Let Γ be a static game.*

(*i*) *Let $\mathcal{A}^\Gamma$ be an Aumann model of Γ, and $\eta(\mathcal{A}^\Gamma)$ be the $\mathcal{A}^\Gamma$-induced epistemic model of Γ. If $(\sigma_i)_{i \in I}$ in $\mathcal{A}^\Gamma$ constitutes a correlated equilibrium, then all types in $\eta(\mathcal{A}^\Gamma)$ express common belief in rationality and $\eta(\mathcal{A}^\Gamma)$ satisfies the common prior assumption.*

(*ii*) *Let $\mathcal{M}^\Gamma$ be an epistemic model of Γ satisfying the common prior assumption, and $\theta(\mathcal{M}^\Gamma)$ be the $\mathcal{M}^\Gamma$-induced Aumann model of Γ. If all types in $\mathcal{M}^\Gamma$ express common belief in rationality, then $(\sigma_i)_{i \in I}$ in $\theta(\mathcal{M}^\Gamma)$ constitutes a correlated equilibrium.*

Proof. For part (i) of the theorem, let $\omega \in \Omega$ be some world and $t_i^{\mathcal{I}_i(\omega)}$ some type of some player $i \in I$. Consider some player $j \in I \setminus \{i\}$ and some choice type pair $(c_j, t_j) \in C_j \times T_j$ of player j such that $b_i[t_i^{\mathcal{I}_i(\omega)}](c_j, t_j) > 0$. As

$$b_i[t_i^{\mathcal{I}_i(\omega)}](c_{-i}, t_{-i}) = \sum_{\omega' \in \mathcal{I}_i(\omega): \sigma_{-i}(\omega') = c_{-i}, t_{-i}^{\mathcal{I}_{-i}(\omega')} = t_{-i}} \pi\big(\omega' \mid \mathcal{I}_i(\omega)\big),$$

there exists a world $\omega' \in \mathcal{I}_i(\omega)$ such that $\pi(\omega') > 0$, $\sigma_{-i}(\omega') = c_{-i}$, and $t_{-i}^{\mathcal{I}_{-i}(\omega')} = t_{-i}$. Since $(\sigma_k)_{k \in I}$ constitutes a correlated equilibrium, $\sigma_j(\omega') = c_j$ is optimal for j's first-order belief at ω' which is the same as $t_j^{\mathcal{I}_j(\omega')}$'s first-order belief by construction of $\eta(\mathcal{A}^\Gamma)$. Because $t_j^{\mathcal{I}_j(\omega')} = t_j$, the choice c_j is optimal for t_j's first-order belief and $t_i^{\mathcal{I}_i(\omega)}$ thus believes in j's rationality. As $t_i^{\mathcal{I}_i(\omega)}$ as well as $t_j^{\mathcal{I}_j(\omega')}$ have been chosen arbitrarily, all types in $\eta(\mathcal{A}^\Gamma)$ believe in rationality, and consequently express common belief in rationality too.

Define a a probability measure $\varphi \in \Delta\big(\times_{j \in I}(C_j \times T_j)\big)$ such that for all $(c_j, t_j^{P_j})_{j \in I} \in \times_{j \in I}(C_j \times T_j)$

$$\varphi\big((c_j, t_j^{P_j})_{j \in I}\big) := \begin{cases} \pi(\cap_{j \in I} P_j), & \text{if } c_j = \sigma_j(P_j) \text{ for all } j \in I, \\ 0, & \text{otherwise.} \end{cases}$$

It is now shown that $\eta(\mathcal{A}^\Gamma)$ satisfies the common prior assumption, by establishing that for all $j \in I$ and $t_j^{P_j} \in T_j$, it is the case that

$$b_j[t_j^{P_j}](c_{-j}, t_{-j}^{P_{-j}}) = \frac{\varphi\big(c_j, t_j^{P_j}, c_{-j}, t_{-j}^{P_{-j}}\big)}{\varphi\big(c_j, t_j^{P_j}\big)}$$

for all $c_j \in C_j$ with $\varphi(c_j, t_j^{P_j}) > 0$, and for all $(c_{-j}, t_{-j}^{P_{-j}}) \in C_{-j} \times T_{-j}$. Note that $\varphi(c_j, t_j^{P_j}) > 0$ only holds if $c_j = \sigma_j(P_j)$. It thus has to be established that

$$b_j[t_j^{P_j}](c_{-j}, t_{-j}^{P_j}) = \frac{\varphi\Big(\big(\sigma_j(P_j), t_j^{P_j}\big), (c_{-j}, t_{-j}^{P_j})\Big)}{\varphi\big(\sigma_j(P_j), t_j^{P_j}\big)}$$

for all $(c_{-j}, t_{-j}^{P_{-j}}) \in C_{-j} \times T_{-j}$ and for all $t_j^{P_j} \in T_j$. Consider some $P_j \in \mathcal{I}_j$ and distinguish two cases (I) and (II).

Case (I). Suppose that $P_j \cap (\cap_{k \in I \setminus \{j\}} P_k) \neq \emptyset$ and $c_k = \sigma_k(P_k)$ for all $k \in I \setminus \{j\}$. Observe that

$$b_j[t_j^{P_j}](c_{-j}, t_{-j}^{P_{-j}}) = b_j[t_j^{P_j}](\sigma_{-j}(P_{-j}), t_{-j}^{P_{-j}})$$

$$= \sum_{\omega' \in P_j: \sigma_{-j}(\omega') = c_{-j}, t_{-j}^{\mathcal{I}_{-j}(\omega')} = t_{-j}^{P_{-j}}} \pi(\omega' \mid P_j)$$

$$= \sum_{\omega' \in P_j : \omega' \in P_k \text{ for all } k \in I \setminus \{j\}} \pi(\omega' \mid P_j)$$

$$= \frac{\pi(\cap_{k \in I} P_k)}{\pi(P_j)}$$

$$= \frac{\varphi\big(\sigma_j(P_j), t_j^{P_j}, \sigma_{-j}(P_{-j}), t_{-j}^{P_{-j}}\big)}{\sum_{\hat{P}_{-j} \in \mathcal{I}_{-j}} \pi\big(P_j \cap (\cap_{k \in I \setminus \{j\}} \hat{P}_k)\big)}$$

$$= \frac{\varphi\big(\sigma_j(P_j), t_j^{P_j}, \sigma_{-j}(P_{-j}), t_{-j}^{P_{-j}}\big)}{\sum_{\hat{P}_{-j} \in \mathcal{I}_{-j}} \varphi\big(\sigma_j(P_j), t_j^{P_j}, \sigma_{-j}(\hat{P}_{-j}), t_{-j}^{\hat{P}_{-j}}\big)}$$

$$= \frac{\varphi\big(\sigma_j(P_j), t_j^{P_j}, \sigma_{-j}(P_{-j}), t_{-j}^{P_{-j}}\big)}{\sum_{(c_{-j}, t_{-j}) \in C_{-j} \times T_{-j}} \varphi\big(\sigma_j(P_j), t_j^{P_j}, c_{-j}, t_{-j}\big)}$$

$$= \frac{\varphi\big(\sigma_j(P_j), t_j^{P_j}, \sigma_{-j}(P_{-j}), t_{-j}^{P_{-j}}\big)}{\varphi\big(\sigma_j(P_j), t_j^{P_j}\big)}.$$

Case (II). Suppose that $P_j \cap (\cap_{k \in I \setminus \{j\}} P_k) = \emptyset$ or $c_k \neq \sigma_k(P_k)$ for some $k \in I \setminus \{j\}$. Then,

$$b_j[t_j^{P_j}](c_{-j}, t_{-j}^{P_{-j}}) = 0 = \frac{\varphi\big(\sigma_j(P_j), t_j^{P_j}, c_{-j}, t_{-j}^{P_{-j}}\big)}{\varphi\big(\sigma_j(P_j), t_j^{P_j}\big)}$$

holds by definition. Hence, $\eta(\mathcal{A}^\Gamma)$ satisfies the common prior assumption.

For part (ii) of the theorem, let $(c_j, t_j)_{j \in I} \in \times_{j \in I}(C_j \times T_j)$ be some choice type combination of all players such that $\varphi\big((c_j, t_j)_{j \in I}\big) > 0$. Consider the world $\omega^{(c_j, t_j)_{j \in I}} \in \Omega$ in $\theta(\mathcal{M}^\Gamma)$ and a choice $c_i' \in C_i$ of some player $i \in I$. Then,

$$\sum_{\omega' \in \mathcal{I}_i\left(\omega^{(c_j, t_j)_{j \in I}}\right)} \pi\Big(\omega' \mid \mathcal{I}_i\big(\omega^{(c_j, t_j)_{j \in I}}\big)\Big) \cdot U_i\big(c_i', \sigma_{-i}(\omega')\big)$$

$$= \sum_{\omega' \in \mathcal{I}_i\left(\omega^{(c_j, t_j)_{j \in I}}\right)} \frac{\pi(\omega')}{\pi\Big(\mathcal{I}_i\big(\omega^{(c_j, t_j)_{j \in I}}\big)\Big)} \cdot U_i\big(c_i', \sigma_{-i}(\omega')\big)$$

$$= \sum_{(c_{-i}', t_{-i}') \in C_{-i} \times T_{-i} : \varphi(c_i, t_i, c_{-i}', t_{-i}') > 0} \frac{\varphi(c_i, c_{-i}', t_i, t_{-i}')}{\varphi(c_i, t_i)} \cdot U_i(c_i', c_{-i}')$$

$$= \sum_{(c_{-i}', t_{-i}') \in C_{-i} \times T_{-i} : b_i[t_i](c_{-i}', t_{-i}') > 0} b_i[t_i](c_{-i}', t_{-i}') \cdot U_i(c_i', c_{-i}')$$

$$= u_i(c_i', t_i),$$

where the third equality follows from the fact that $\mathcal{M}^\Gamma$ satisfies the common prior assumption with common prior φ. Now, consider some world $\omega^{(c_j, t_j)_{j \in I}} \in \Omega$

and some player $i \in I$. Since $\varphi(c_i, t_i) > 0$, there exists a type $t_j \in T_j$ such that $b_j[t_j](c_i, t_i) > 0$ for some player $j \in I$. As t_j expresses common belief in rationality, t_j believes in i's rationality. Hence

$$u_i(c_i, t_i) \geq u_i(c_i', t_i)$$

for all $c_i' \in C_i$. Because

$$u_i(c_i', t_i) = \sum_{\omega' \in \mathcal{I}_i\left(\omega^{(c_j, t_j)_{j \in I}}\right)} \pi\Big(\omega' \mid \mathcal{I}_i\big(\omega^{(c_j, t_j)_{j \in I}}\big)\Big) \cdot U_i\big(c_i', \sigma_{-i}(\omega')\big)$$

for all $c_i' \in C_i$, and $\sigma_i\big(\omega^{(c_j, t_j)_{j \in I}}\big) = c_i$, it follows that

$$\sum_{\omega' \in \mathcal{I}_i\left(\omega^{(c_j, t_j)_{j \in I}}\right)} \pi\Big(\omega' \mid \mathcal{I}_i\big(\omega^{(c_j, t_j)_{j \in I}}\big)\Big) \cdot U_i\Big(\sigma_i\big(\omega^{(c_j, t_j)_{j \in I}}\big), \sigma_{-i}(\omega')\Big) = u_i(c_i, t_i)$$

$$\geq u_i(c_i', t_i) = \sum_{\omega' \in \mathcal{I}_i\left(\omega^{(c_j, t_j)_{j \in I}}\right)} \pi\Big(\omega' \mid \mathcal{I}_i\big(\omega^{(c_j, t_j)_{j \in I}}\big)\Big) \cdot U_i\big(c_i', \sigma_{-i}(\omega')\big)$$

holds for all $c_i' \in C_i$, and thus $(\sigma_i)_{i \in I}$ constitutes a correlated equilibrium. ■

In fact, Theorem 1 can be interpreted as a morphism between Aumann models and epistemic models that preserves some notions of optimality of choice and common prior.

An epistemic characterization of correlated equilibrium in terms of common belief in rationality and a common prior ensues as follows.

Theorem 2. *Let Γ be a static game, $i \in I$ some player, $\beta_i^* \in \Delta(C_{-i})$ some first-order belief of player i, and $c_i^* \in C_i$ some choice of player i.*

(i) The first-order belief β_i^* is possible in a correlated equilibrium, if and only if, the first-order belief β_i^* is possible under common belief in rationality with a common prior.

(ii) The choice c_i^* is optimal in a correlated equilibrium, if and only if, the choice c_i^* is rational under common belief in rationality with a common prior.

Proof. For the *only if* direction of part (i) of the theorem, let $\mathcal{A}^\Gamma$ be an Aumann model of Γ and $(\sigma_j)_{j \in I}$ a correlated equilibrium, in which β_i^* is possible. Then, there exists a world $\hat{\omega} \in \Omega$ such that $\beta_i^*(c_{-i}) = \pi\big(\{\omega' \in \mathcal{I}_i(\hat{\omega}) : \sigma_{-i}(\omega') = c_{-i}\} \mid \mathcal{I}_i(\hat{\omega})\big)$ for all $c_{-i} \in C_{-i}$. Consider the epistemic model $\eta(\mathcal{A}^\Gamma)$ of Γ. By Theorem 1 (i), the type $t_i^{\mathcal{I}_i(\hat{\omega})}$ expresses common belief in rationality, and the epistemic model $\eta(\mathcal{A}^\Gamma)$ of Γ satisfies the common prior assumption. Note that $b_i[t_i^{\mathcal{I}_i(\hat{\omega})}](c_{-i}, t_{-i}) = \sum_{\omega \in \mathcal{I}_i(\hat{\omega}) : \sigma_{-i}(\omega) = c_{-i}, \eta_{-i}(\omega) = t_{-i}} \pi\big(\omega \mid \mathcal{I}_i(\hat{\omega})\big)$ for all $(c_{-i}, t_{-i}) \in C_{-i} \times T_{-i}$, and thus $\beta_i^*(c_{-i}) = b_i[t_i^{\mathcal{I}_i(\hat{\omega})}](c_{-i})$ for all $c_{-i} \in C_{-i}$.

Therefore, the first-order belief β_i^* is possible under common belief in rationality with a common prior.

For the *if* direction of the part (i) of the theorem, suppose that β_i^* is possible under common belief in rationality with a common prior. Thus, there exists an epistemic model $\mathcal{M}^\Gamma$ of Γ with a type $t_i^* \in T_i$ such that t_i^* expresses common belief in rationality, $b_i[t_i^*](c_{-i}) = \beta_i^*(c_{-i})$ for all $c_{-i} \in C_{-i}$, and $\mathcal{M}^\Gamma$ satisfies the common prior assumption. Construct an epistemic model $(\mathcal{M}^\Gamma)' = \big((T_j')_{j\in I}, (b_j')_{j\in I}\big)$ of Γ, where for every player $j \in I$, the set T_j' of types contains those $t_j \in T_j$ from $\mathcal{M}^\Gamma$ such that $t_j \in T_j(t_i^*)$, i.e. t_j is belief-reachable from t_i^*. Note that $(\mathcal{M}^\Gamma)'$ satisfies the common prior assumption, with common prior $\varphi' \in \Delta\big(\times_{j\in I}\,(C_j \times T_j')\big)$ being $\varphi \in \Delta\big(\times_{j\in I}\,(C_j \times T_j)\big)$ from $\mathcal{M}^\Gamma$ restricted to, and normalized on, $\times_{j\in I}(C_j \times T_j')$. By Lemma 1, all types in $(\mathcal{M}^\Gamma)'$ express common belief in rationality. It then follows with Theorem 1 (ii) that $(\sigma_j)_{j\in I}$ constitutes a correlated equilibrium in $\theta\big((\mathcal{M}^\Gamma)'\big)$. As the first-order beliefs of t_i^* are the same in $(\mathcal{M}^\Gamma)$ and $(\mathcal{M}^\Gamma)'$, the first-order belief of t_i^* equals β_i^* also in $(\mathcal{M}^\Gamma)'$. Consider a world $\omega^{(c_i,t_i^*,c_{-i},t_{-i})} \in \Omega$ with $\varphi'(c_i, t_i^*, c_{-i}, t_{-i}) > 0$ for some $c_i \in C_i$, $c_{-i} \in C_{-i}$, and $t_{-i} \in T_{-i}$. Consequently, $\beta_i^*(c_{-i}) = b_i[t_i^*](c_{-i}) = \sum_{t_{-i}\in T_{-i}} \varphi(c_{-i}, t_{-i} \mid c_i, t_i^*) = \pi\Big(\{\omega \in \mathcal{I}_i\big(\omega^{(c_i,t_i^*,c_{-i},t_{-i})}\big) : \sigma_{-i}\big(\omega\big) = c_{-i}\} \mid \mathcal{I}_i\big(\omega^{(c_i,t_i^*,c_{-i},t_{-i})}\big)\Big)$. Therefore, β_i^* is possible in a correlated equilibrium.

For part (ii) of the theorem, let $\mathcal{A}^\Gamma$ be an Aumann model of Γ and $(\sigma_j)_{j\in I}$ a correlated equilibrium, in which c_i^* is optimal. Then, there exists some first-order belief $\beta_i^* \in \Delta(C_{-i})$ possible in $\mathcal{A}^\Gamma$ for which c_i^* maximizes expected utility. By part (i) of the corollary it then follows that β_i^* is also possible under common belief in rationality with a common prior, and consequently c_i^* is optimal under common belief in rationality with a common prior too. Conversely, let $\mathcal{M}^\Gamma$ be an epistemic model of Γ with a type $t_i^* \in T_i$ such that t_i^* expresses common belief in rationality, c_i^* is optimal for t_i^*, and $\mathcal{M}^\Gamma$ satisfies the common prior assumption. Let $\beta_i^* \in \Delta(C_i)$ be the first-order belief of t_i^*. Then, β_i^* is possible under common belief in rationality with a common prior. By part (i) of the corollary it then follows that β_i^* is also possible in a correlated equilibrium, and consequently c_i^* is optimal in a correlated equilibrium too. ■

From an epistemic perspective correlated equilibrium is thus – doxastically and behaviourally – equivalent to common belief in rationality with a common prior. In fact, the epistemic characterization of correlated equilibrium according to Theorem 2 is similar to Dekel and Siniscalchi (2015, Theorem 12.14). However, the two epistemic characterizations differ importantly in the sense that the latter is provided for an ex ante perspective while the former is furnished for an ex post perspective. Furthermore, a minor difference lies in the formulation of the epistemic characterization in terms of belief hierarchies (Dekel and Siniscalchi, 2015, Theorem 12.14) as opposed to types (Theorem 2). Note that the conditions used by Dekel and Sinischalchi (2015, Theorem 12.14) as well as by Theorem 2 are weaker than in Aumann (1987), where correlated equilibrium is characterized – from an ex ante perspective – in terms of universal rationality

and a common prior. More precisely, Aumann (1987) assumes that players are rational at all possible worlds, which is stronger than common belief in rationality. Intuitively, in Aumann's (1987) model no irrationality in the system is admitted at all. Besides, Brandenburger and Dekel (1987) characterize a variant of correlated equilibrium without a common prior – a posteriori equilibrium and sometimes also called subjective correlated equilibrium – by common knowledge of rationality.

Next canonical correlated equilibrium is considered from an epistemic perspective. Before the solution concept is epistemically characterized, two further doxastic conditions are introduced.

Definition 8. *Let Γ be a static game, $\mathcal{M}^{\Gamma}$ an epistemic model of it, $i, j \in I$ two players, $t_i \in T_i$ some type of player i, $\beta_j \in \Delta(C_{-j})$ some first-order belief of player j, and $c_j \in C_j$ some choice of player j. The type t_i* always explains choice *c_j by first-order belief β_j, if for all $t_j \in T_j$ such that $(c_j, t_j) \in (C_j \times T_j)(t_i)$, it is the case that*

$$b_j[t_j](c_{-j}) = \beta_j(c_{-j})$$

for all $c_{-j} \in C_{-j}$.

Accordingly, every given choice deemed possible a reasoner accompanies with the same first-order belief in his entire belief hierarchy. In this sense, throughout his reasoning any given choice is explained in a unique way.

Requiring a player to always explain any choice with a fixed first-order belief gives rise to the notion of one-theory-per-choice, as follows.

Definition 9. *Let Γ be a static game, $\mathcal{M}^{\Gamma}$ an epistemic model of it, $i \in I$ some player, and $t_i \in T_i$ some type of player i. The type t_i holds* one-theory-per-choice*, if for all $j \in I$, and for all $c_j \in C_j$, there exists $\beta_j \in \Delta(C_{-j})$ such that t_i always explains c_j by β_j.*

Intuitively, a player reasoning in line with one-theory-per-choice never – i.e. nowhere in his belief hierarchy – uses distinct first-order beliefs ("theories") for any player to explain the same choice of this player. The reasoner does thus not use more theories than necessary in his belief hierarchy, which is in this sense sparse. Besides, note that in Example 2 *Bob*'s belief hierarchy induced at world ω_3 actually violates the one-theory-per-choice condition. Indeed, *Bob* believes with probability $\frac{1}{4}$ that *Alice* chooses b while believing him to choose e, but he also believes with probability $\frac{1}{4}$ that *Alice* chooses b while believing him to choose e with probability $\frac{1}{3}$ and g with probability $\frac{2}{3}$.

In fact, the one-theory-per-choice condition contains a rather strong psychological assumption in terms of correctness of beliefs. Since at no iteration in the full belief hierarchy of a reasoner holding one-theory-per-choice any given choice is coupled with distinct first-order beliefs, the reasoner believes that his opponents are correct about how he explains any choice, he believes that his opponents believe that their opponents are correct about how he explains any choice, etc. Also, the reasoner does not only believe that any opponent only

uses a single theory to explain a given choice, but also believes that his other opponents believe so, and that they believe their opponents to believe so, etc. In particular, the following remark thus ensues.

Remark 5. Let Γ be a static game, $\mathcal{M}^\Gamma$ an epistemic model of it, $i \in I$ some player, and $t_i \in T_i$ some type of player i that holds one-theory-per-choice. Consider some player $j \in I$, some choice of player $c_j \in C_j$, and some first-order belief $\beta_j \in \Delta(C_{-j})$ of player j such that t_i always explains c_j by β_j.

(i) For all $k \in I \setminus \{i\}$, for all $t_k \in T_k$ such that $b_i[t_i](t_k) > 0$, and for all $t_i' \in T_i$ such that $b_k[t_k](t_i') > 0$, it is the case that t_i' always explains c_j by β_j.

(ii) For all $l \in I \setminus \{i, j\}$, and for all $t_l \in T_l$ such that $b_i[t_i](t_l) > 0$, it is the case that t_l always explains c_j by β_j.

Accordingly, the one-theory-per-choice condition thus contains two correctness of beliefs assumptions: a reasoner believes his opponents to be correct about all of his choice explanations as well as projects his choice explanations on any other opponent. It is even the case that common belief in these two properties – or formally in properties (i) and (ii) of Remark 5 – is implied by one-theory-per-choice, as they are taken for certain in all interactive belief iterations.

Besides, a first-order belief $\beta_i \in C_i$ is said to be *possible under common belief in rationality with a common prior and one-theory-per-choice*, if there exists an epistemic model $\mathcal{M}^\Gamma$ of Γ satisfying the common prior assumption with a type $t_i^* \in T_i$ of i such that $b_i[t_i^*](c_{-i}) = \beta_i^*(c_{-i})$ for all $c_{-i} \in C_{-i}$ and t_i^* expresses common belief in rationality as well as holds one-theory-per-choice. Similarly, a choice $c_i^* \in C_i$ is said to be *rational under common belief in rationality with a common prior and one-theory-per-choice*, if there exists an epistemic model $\mathcal{M}^\Gamma$ of Γ satisfying the common prior assumption with a type $t_i^* \in T_i$ of i such that c_i^* is optimal for t_i^* and t_i^* expresses common belief in rationality as well as holds one-theory-per-choice.

An epistemic characterization of canonical correlated equilibrium then ensues as follows.

Theorem 3. *Let Γ be a static game, $i \in I$ some player, β_i^* some first-order belief of player i, and $c_i^* \in C_i$ some choice of player i.*

(i) *The first-order belief β_i^* is possible in a canonical correlated equilibrium, if and only if, the first-order belief β_i^* is possible under common belief in rationality with a common prior and one-theory-per-choice.*

(ii) *The choice c_i^* is optimal in a canonical correlated equilibrium, if and only if, the choice c_i^* is rational under common belief in rationality with a common prior and one-theory-per-choice.*

Proof. For the *only if* direction of part (i) of the theorem, suppose that $\rho \in \Delta(\times_{j \in I} C_j)$ constitutes a canonical correlated equilibrium of Γ. For every $j \in I$

define a type space $T_j := \{t_j^{c_j} : \rho(c_j) > 0\}$ with induced belief function

$$b_j[t_j^{c_j}](c_{-j}, t_{-j}) := \begin{cases} \rho(c_{-j} \mid c_j), & \text{if } t_{-j} = t_{-j}^{c_{-j}}, \\ 0, & \text{otherwise,} \end{cases}$$

for every type $t_j^{c_j} \in T_j$. Also, define a probability measure $\varphi \in \Delta\big((C_j \times T_j)_{j \in I}\big)$ such that

$$\varphi\big((c_j, t_j)_{j \in I}\big) := \begin{cases} \rho\big((c_j)_{j \in I}\big), & \text{if } t_j = t_j^{c_j} \text{ for all } j \in I, \\ 0, & \text{otherwise,} \end{cases}$$

for all $(c_j, t_j)_{j \in I} \in (C_j \times T_j)_{j \in I}$.

Observe that

$$\frac{\varphi(c_j, t_j^{c_j}, c_{-j}, t_{-j}^{c_{-j}})}{\varphi(c_j, t_j^{c_j})} = \frac{\rho\big((c_k)_{k \in I}\big)}{\rho(c_j)} = \rho(c_{-j} \mid c_j) = b_j[t_j^{c_j}](c_{-j}, t_{-j}^{c_{-j}})$$

holds for all $(c_j, t_j^{c_j}) \in C_j \times T_j$, and thus the constructed epistemic model $\big((T_j)_{j \in I}, (b_j)_{j \in I}\big)$ satisfies the common prior assumption with common prior φ.

Next consider some type $t_j^{c_j} \in T_j$ and let $(c_k, t_k), (c_k, t_k') \in (C_k \times T_k)(t_j^{c_j})$ be belief-reachable from $t_j^{c_j}$. By definition of T_k it holds that $t_k = t_k' = t_k^{c_k}$ and thus $b_k[t_k](c_{-k}) = b_k[t_k'](c_{-k})$ trivially holds for all $c_{-k} \in C_{-k}$. Therefore, $t_j^{c_j}$ holds one-theory-per-choice. As $t_j^{c_j}$ has been chosen arbitrarily, all types in T_j hold one-theory-per-choice.

Furthermore, let $(c_k, t_k) \in C_k \times T_k$ such that $b_j[t_j^{c_j}](c_k, t_k) > 0$ for some $t_j^{c_j} \in T_j$. Then, $t_k = t_k^{c_k}$ and $b_k[t_k^{c_k}](c_{-k}) = \rho(c_{-k} \mid c_k)$ holds for all $c_{-k} \in C_{-k}$ as well as $\rho(c_k) > 0$. Since ρ is a canonical correlated equilibrium, c_k is optimal for $\rho(\cdot \mid c_k)$ and consequently optimal for $t_k^{c_k}$ too. Hence, all types believe in rationality and a fortiori all types express common belief in rationality.

Suppose that β_i^* is possible in the canonical correlated equilibrium ρ. Then, there exists some choice $\hat{c}_i \in C_i$ with $\rho(\hat{c}_i) > 0$ such that $\rho(c_{-i} \mid \hat{c}_i) = \beta_i^*(c_{-i})$ for all $c_{-i} \in C_{-i}$. Consider the type $t_i^{\hat{c}_i} \in T_i$, which indeed exists due to $\rho(\hat{c}_i) > 0$, and observe that $b_i[t_i^{\hat{c}_i}](c_{-i}) = \rho(c_{-i} \mid \hat{c}_i) = \beta_i^*(c_{-i})$ for all $c_{-i} \in C_{-i}$. Therefore, the first-order belief β_i^* is possible under common belief in rationality with a common prior and one-theory-per-choice.

For the *if* direction of part (i) of the theorem, let $\mathcal{M}^\Gamma$ be an epistemic model of Γ that satisfies the common prior assumption with common prior $\varphi \in \Delta\big(\times_{j \in I} (C_j \times T_j)\big)$, as well as $t_i^* \in T_i$ be a type such that t_i^* expresses common belief in rationality, holds one-theory-per-choice, and t_i^* holds first-order belief β_i^*. It is shown that β_i^* is possible in a canonical correlated equilibrium.

Consider some choice type pair $(c_j, t_j) \in (C_j \times T_j)(t_i^*)$ of some player $j \in I$ that is belief-reachable from t_i^*. Then, there exists a sequence $(t^1, \ldots, t^N)$ of types such that $t^1 = t_i^*$, $t^N = t_j$, $b_k[t^n](t^{n+1}) > 0$ for all $n \in \{1, \ldots, N-1\}$, for some $k \in I$, and $b_l[t^{N-1}](c_j, t_j) > 0$. As t_i^* expresses $(N-1)$-fold belief in rationality, it directly follows that c_j is optimal for t_j.

Define a probability measure $\rho \in \Delta(\times_{k\in I} C_k)$ by

$$\rho\big((c_k)_{k\in I}\big) := \begin{cases} \frac{\varphi(\times_{k\in I}\{c_k\}\times T_k)}{\varphi\left(\times_{k\in I}(C_k\times T_k)(t_i^*)\right)}, & \text{if } c_k \in C_k(t_i^*) \text{ for all } k \in I, \\ 0, & \text{otherwise,} \end{cases}$$

for all $(c_k)_{k\in I} \in \times_{k\in I} C_k$, where $C_k(t_i^*) := \{c_k \in C_k : (c_k, t_k) \in (C_k \times T_k)(t_i^*) \text{ for some } t_k \in T_k\}$.

Let $\tilde{c}_j \in C_j$ be some choice such that $\rho(\tilde{c}_j) > 0$. Thus, $\tilde{c}_j \in C_j(t_i^*)$ and there exists some type $\tilde{t}_j \in T_j$ such that $(\tilde{c}_j, \tilde{t}_j) \in (C_j \times T_j)(t_i^*)$. Since t_i^* expresses common belief in rationality, it follows, that $\tilde{c}_j$ is optimal for $\tilde{t}_j$. As $\mathcal{M}^\Gamma$ satisfies the common prior assumption, it is the case that

$$b_j[\tilde{t}_j](c_{-j}, t_{-j}) = \frac{\varphi(\tilde{c}_j, \tilde{t}_j, c_{-j}, t_{-j})}{\varphi(\tilde{c}_j, \tilde{t}_j)}$$

holds, and hence

$$b_j[\tilde{t}_j](c_{-j}) = \frac{\varphi(\tilde{c}_j, \tilde{t}_j, \{c_{-j}\} \times T_{-j})}{\varphi(\tilde{c}_j, \tilde{t}_j)}$$

for all $c_{-j} \in C_{-j}$.

Since t_i^* holds one-theory-per-choice, all types in the set $T_j(\tilde{c}_j) := \{t_j' \in T_j : (\tilde{c}_j, t_j') \in (C_j \times T_j)(t_i^*)\}$ have the same first-order belief $\beta_j \in \Delta(C_{-j})$. Consequently, for all $t_j' \in T_j(\tilde{c}_j)$ it is the case that

$$b_j[t_j'](c_{-j}) = \frac{\varphi(\{\tilde{c}_j, t_j'\} \times \{c_{-j}\} \times T_{-j})}{\varphi(\tilde{c}_j, t_j')} = \beta_j(c_{-j})$$

for all $c_{-j} \in C_{-j}$. Then,

$$\rho(c_{-j} \mid \tilde{c}_j) = \frac{\rho(\tilde{c}_j, c_{-j})}{\rho(\tilde{c}_j)} = \frac{\varphi\big(\{\tilde{c}_j\} \times T_j(\tilde{c}_j) \times \{c_{-j}\} \times T_{-j}\big)}{\varphi\big(\{\tilde{c}_j\} \times T_j(\tilde{c}_j)\big)}$$

$$\frac{\sum_{t_j' \in T_j(\tilde{c}_j)} \varphi(\{\tilde{c}_j, t_j'\} \times \{c_{-j}\} \times T_{-j})}{\sum_{t_j' \in T_j(\tilde{c}_j)} \varphi(\tilde{c}_j, t_j')} = \frac{\sum_{t_j' \in T_j(\tilde{c}_j)} \beta_j(c_{-j}) \cdot \varphi(\tilde{c}_j, t_j')}{\sum_{t_j' \in T_j(\tilde{c}_j)} \varphi(\tilde{c}_j, t_j')} = \beta_j(c_{-j})$$

for all $c_{-j} \in C_{-j}$. Thus, $\tilde{t}_j$'s first-order belief is $\beta_j = \rho(\cdot \mid \tilde{c}_j)$, and – since $\tilde{c}_j$ is optimal for $\tilde{t}_j$ – it is the case that $\tilde{c}_j$ is optimal for $\rho(\cdot \mid \tilde{c}_j)$. Therefore, ρ is a canonical correlated equilibrium.

Recall that t_i^* holds first-order belief β_i^*. It is shown that β_i^* is possible in the canonical correlated equilibrium ρ. As $\varphi(t_i^*) > 0$, and $\mathcal{M}^\Gamma$ satisfies the common prior assumption, it follows that $(\tilde{c}_i, t_i^*) \in (C_i \times T_i)(t_i^*)$ for some $\tilde{c}_i \in C_i$. In fact, there exists a player $l \in I$ such that $b_i[t_i^*](t_l) > 0$ and $b_l[t_l](\tilde{c}_i, t_i^*) > 0$. Since t_i^* holds one-theory-per-choice, β_i^* is the unique first-order belief attached to $\tilde{c}_i$ in t_i^*'s induced belief hierarchy. As $t_i^* \in T_i(\tilde{c}_i)$, it follows from above that $\beta_i^*(c_{-i}) = b_i[t_i^*](c_{-i}) = \rho(c_{-i} \mid \tilde{c}_i)$ for all $c_{-i} \in C_{-i}$. Consequently, β_i^* is possible in a canonical correlated equilibrium.

For part (ii) of the theorem, let ρ be a canonical correlated equilibrium, in which c_i^* is optimal. Then, there exists some first-order belief $\beta_i^* \in \Delta(C_{-i})$ possible in ρ for which c_i^* maximizes expected utility. By part (i) of the theorem it then follows that β_i^* is also possible under common belief in rationality with a common prior and one-theory-per-choice, thus c_i^* is optimal under common belief in rationality with a common prior and one one-theory-per-choice too. Conversely, let $\mathcal{M}^\Gamma$ be an epistemic model of Γ with a type $t_i^* \in T_i$ such that t_i^* expresses common belief in rationality, t_i^* holds one-theory-per-choice, c_i^* is optimal for t_i^*, and $\mathcal{M}^\Gamma$ satisfies the common prior assumption. Let β_i^* be t_i^*'s first-order belief. Then, β_i^* is possible under common belief in rationality with a common prior and one-theory-per-choice. By part (i) of the theorem it then follows that β_i^* is also possible in a canonical correlated equilibrium, and consequently c_i^* is optimal in a canonical correlated equilibrium too. ■

From an epistemic perspective the solution concept of canonical correlated equilibrium thus is substantially stronger than correlated equilibrium by also requiring the reasoner's thinking to be in line with the one-theory-per-choice condition, which in turn contains a correctness of beliefs assumption.

It can be concluded that correlated equilibrium and canonical correlated equilibrium are distinct solution concepts both behaviourally as well as doxastically. The epistemic characterizations via Theorems 2 and 3 shed light on understanding this difference. Indeed, canonical correlated equilibrium requires some correctness of beliefs property – the one-theory-per-choice condition – in addition to common belief in rationality and a common prior also used by correlated equilibrium. Since some correctness of beliefs assumption also constitutes the substantial reasoning property of Nash equilibrium, canonical correlated equilibrium appears to be closer to this solution concept, while correlated equilibrium seems to be more distant from it. Also, canonical correlated equilibrium can thus be seen as a more demanding solution concept than correlated equilibrium in terms of reasoning.

6 Discussion

One-Theory-per-Choice. A player reasoning in line with the one-theory-per-choice condition uses for each of his opponents' choices only a single first-order belief in his whole belief hierarchy. In other words, a player never uses two different first-order beliefs to explain the same choice in his whole belief hierarchy. The one-theory-per-choice condition thus keeps a belief hierarchy lean. Such a sparsity condition is similar to Perea's (2012) epistemic notion of simple belief hierarchies, which require a belief hierarchy to be entirely generated by a tuple of first-order beliefs. Since simple belief hierarchies are closely connected to Nash equilibrium and the one-theory-per-choice condition to canonical correlated equilibrium, the resemblance between the two conditions in terms of leanness gives canonical correlated equilibrium some Nash equilibrium flavour, which is absent from correlated equilibrium due to lacking such a leanness condition.

Common Belief in Rationality. The one-theory-per-choice condition does not have any behavioural effect if imposed in addition to common belief in rationality only. Intuitively, if a choice is rational under common belief in rationality, it is well-known that it then survives iterated elimination of strictly dominated choices. It is possible to construct an epistemic model such that there exists a single type for every surviving choice. As for every choice there then exists a unique supporting type, belief in rationality already requires a unique way of coupling opponents' choices and types in the support of a given player's induced belief function. Consequently, the one-theory-per-choice condition holds in such an epistemic model. Therefore, a choice is rational under common belief in rationality, if and only if, it is rational under common belief in rationality with one-theory-per-choice.

Thus, the one-theory-per-choice-condition does not add anything in terms of optimal choice to common belief in rationality. Only if a common prior is also assumed the one-theory-per-choice condition exhibits behavioural implications beyond common belief in rationality resulting in canonical correlated equilibrium and not in iterated elimination of strictly dominated choices. Remark 5 also distinguishes the one-theory-per-choice condition from simple belief hierarchies. Indeed, the assumption of simple belief hierarchies in conjunction with common belief in rationality behaviourally yields Nash equilibrium (Perea, 2012).

Nash Equilibrium. The epistemic analysis of Nash equilibrium (e.g. Aumann and Brandenburger, 1995; Perea, 2007; Barelli, 2009; Bach and Tsakas, 2014; Bonanno, 2017) has unveiled a correctness of beliefs assumption as the decisive epistemic property of Nash equilibrium. In fact, a correctness of beliefs property also features implicitly in the one-theory-per-choice condition: the reasoner believes that his opponents are correct about his theories, believes that his opponents believe that their opponents are correct about his theories, etc. Thus, canonical correlated equilibrium exhibits some Nash equilibrium flavour, whereas correlated equilibrium does not.

To some extent, the lack of a correctness of beliefs assumption for correlated equilibrium illustrates its fundamental difference to Nash equilibrium. Intuitively, the former solution concept only requires players to behave optimally given the opponents' choice functions, while the latter necessitates players to behave optimally given the opponents' actual choices.

Nash equilibrium can be characterized by common belief in rationality together with simple belief hierarchies. The correctness of beliefs assumptions due to simple belief hierarchies and one-theory-per-choice can be compared. As the whole belief hierarchy is generated by a single tuple of first-order beliefs, the condition simple belief hierarchies directly implies the one-theory-per-choice condition. However, it is possible in a belief hierarchy satisfying the one-theory-per-choice condition that different choices of some opponent are coupled with types inducing distinct first-order beliefs for that opponent, which is impossible for simple belief hierarchies, as all choices of a player are explained by only a single theory in the reasoner's entire belief hierarchy. Besides, simple belief hierarchies imply independence of the first-order beliefs that they are generated with, which

is not necessarily the case with belief hierarchies satisfying the one-theory-per-choice condition. Therefore, if a type holds a simple belief hierarchy, then he also holds one-theory-per-choice, while it is possible that a type holds one-theory-per-choice but no simple belief hierarchy.

The one-theory-per-choice condition thus constitutes a weaker correctness of beliefs assumption than the simplicity condition. It can then be argued that implausibility criticisms due to implicit correctness of beliefs properties affect Nash equilibrium stronger than canonical correlated equilibrium.

Besides, correctness of beliefs inherent in simple belief hierarchies or one-theory-per-choice lies entirely inside the mind of the respective reasoner. In this one-person perspective sense the notion of correctness used here is distinct from the truth axiom ("a proposition is implied by the belief in it"), which is the way correctness of beliefs is typically understood in philosophy. In fact, the truth axiom cannot be expressed in the one-person perspective type-based epistemic models used here (Definition 3), as a formal notion of state is lacking. In a sense, correctness of beliefs in the sense of simple belief hierarchies and one-theory-per-choice is a subjective property, while the truth axiom embodies an objective correctness of beliefs trait.

Common Prior Assumption. The common prior assumption is present in both Theorem 2 and Theorem 3, and thus underlies correlated equilibrium as well as canonical correlated equilibrium. Psychologically, belief hierarchies derived from a common prior can be interpreted as exhibiting a kind of symmetry in the reasoning of the respective player and his opponents. While the existence of a common prior does imply that a player believes that his opponents assign positive probability to his true belief hieararchy, a genuine correctness of beliefs property of a common priror is not directly apparent. The exploration of belief hierarchies derived from a common prior and any potential correctness of beliefs properties represents an intriguing question for further research. In any case, Nash equilibrium and canonical correlated equilibrium implicitly assume simple belief hierarchies and one-theory-per-choice, respectively, as correctness of beliefs properties. Therefore, canonical correlated equilibrium is conceptually closer to Nash equilibrium than correlated equilibrium is to Nash equilibrium, independent of whether the common prior assumption exhibits any correctness of beliefs flavour, or not.

Ex Ante and Ex Post. From an ex ante perspective before any reasoning or decision-making takes place, correlated equilibrium and canonical correlated equilibrium induce the same probability measure on the players' choice combinations. While a canonical correlated equilibrium $\rho \in \Delta(\times_{i\in I} C_i)$ directly specifies such a probability measure, the induced such measure in Aumann structures – being based on the common prior and the choice functions – is given by $\pi\big(\{\omega \in \Omega : \sigma_i(\omega) = c_i \text{ for all } i \in I\}\big) \in \Delta(\times_{i\in I} C_i)$. Thus, equivalence ex ante is formally expressed by $\rho\big((c_i)_{i\in I}\big) := \pi\big(\{\omega \in \Omega : \sigma_i(\omega) = c_i \text{ for all } i \in I\}\big)$ for all $(c_i)_{i\in I} \in \times_{i\in I} C_i$ such that ρ and $(\sigma_i)_{i\in I}$ constitute a canonical correlated equilibrium and correlated equilibrium, respectively, of the same under-

lying game. If $(\sigma_i)_{i \in I}$ constitutes a correlated equilibrium, then simply define $\rho\big((c_i)_{i \in I}\big) := \pi\big(\{\omega \in \Omega : \sigma_i(\omega) = c_i \text{ for all } i \in I\}\big)$ for all $(c_i)_{i \in I} \in \times_{i \in I} C_i$ and it then follows by the proof of Aumann (1987, Main Theorem) that ρ constitutes a canonical correlated equilibrium. Conversely, if $\rho \in \Delta(\times_{i \in I} C_i)$ constitutes a canonical correlated equilibrium, observe that the constructed correlated equilibrium in the paragraph just before Remark 1 exhibits the property that $\pi\big(\{\omega \in \Omega : \sigma_i(\omega) = c_i \text{ for all } i \in I\}\big) = \rho\big((c_i)_{i \in I}\big)$ for all $(c_i)_{i \in I} \in \times_{i \in I} C_i$.

While the equivalence of correlated equilibrium and canonical correlated equilibrium in terms of the induced probability measure on the players' choice combinations a priori is well-known, such an ex ante equivalence is only of limited interest for reasoning and decision-making in games. Indeed, the posterior beliefs and the optimal choices in line with these posterior beliefs are the relevant objects for reasoning and decision-making. The two solution concepts have been shown here to differ in terms of both their possible posterior beliefs (Remark 3) as well as their optimal choices (Remark 4), i.e. in terms of both dimensions significant for reasoning and decision-making.

Two Distinct Solution Concepts. The epistemic characterizations of correlated equilibrium (Theorem 2) and canonical correlated equilibrium (Theorem 3) show that the two solution concepts are actually distinct. In addition to common belief in rationality and a common prior, canonical correlated equilibrium also requires a correctness of beliefs assumption in form of the one-theory-per-choice condition and thus makes stronger epistemic assumption than correlated equilibrium. Intuitively, in a correlated equilibrium a player can justify an opponent's choice with two different first-order beliefs in his reasoning, but not in canonical correlated equilibrium. In classical terms, correlated equilibrium and its simplified variant differ, because two information cells can induce the same choice yet different conditional beliefs for a given player via his choice function in a correlated equilibrium, while two different conditioning events, i.e. two distinct choices, always induce different choices in a canonical correlated equilibrium, as the conditioning events in a canonical correlated equilibrium coincide with those choices that receive positive weight by the probability measure on the players' choice combinations. Hence, canonical correlated equilibrium can be viewed as a special case of correlated equilibrium, where different information cells prescribe different choices. To support a particular first-order belief in a correlated equilibrium it may be crucial to use two information cells inducing the same choice for a given player. There generally thus exists more flexibility to build beliefs in a correlated equilibrium, and to consequently also make choices optimal. To conclude, correlated equiilbrium and canonical correlated equilibrium form two distinct solution concepts for games based on the idea of correlation.

References

Aumann, R. J. (1974): Subjectivity and Correlation in Randomized Strategies. *Journal of Mathematical Economics* 1, 67–96.

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

Aumann, R. J. and Brandenburger, A. (1995): Epistemic Conditions for Nash Equilibrium. *Econometrica* 63, 1161–1180.

Aumann, R. J. and Dreze, J. H. (2008): Rational Expectations in Games. *American Economic Review* 98, 72–86.

Bach, C. W. and Tsakas, E. (2014): Pairwise Epistemic Conditions for Nash Equilibrium. *Games and Economic Behavior* 85, 48–59.

Barelli, P. (2009): Consistency of Beliefs and Epistemic Conditions for Nash and Correlated Equilibria. *Games and Economic Behavior* 67, 363–375.

Bonanno, G. (2017): Behavior and Deliberation in Perfect-Information Games: Nash Equilibrium and Backward Induction. Mimeo.

Brandenburger, A. and Dekel, E. (1987): Rationalizability and Correlated Equilibria. *Econometrica* 55, 1391–1402.

Dekel, E. and Siniscalchi, M. (2015): Epistemic Game Theory. In *Handbook of Game Theory with Economic Applications*, Vol. 4, 619–702.

Forges, F. (1990): Universal Mechanisms. *Econometrica*, 59, 1341–1364.

Harsanyi, J. C. (1967-68): Games of Incomplete Information played by "Bayesian Players". Part I, II, III. *Management Science* 14, 159–182, 320–334, 486–502.

Perea, A. (2007): A One-Person Doxastic Characterization of Nash Strategies. *Synthese*, 158, 1251–1271.

Perea, A. (2012): *Epistemic Game Theory: Reasoning and Choice*. Cambridge University Press.