Discussion Paper No. 1076

Independence for Conditional Probability Systems[1]

by

Jeroen M. Swinkels[2]

First Draft: November 27, 1993

This Draft: January 11, 1994

---

[1] I am grateful to Roger Myerson, Phil Reny, Eddie Dekel and Larry Samuelson for very helpful discussions and commentary. I also thank Bentley Macleod, David Levine and Joseph Greenberg.

[2] J.L. Kellogg Graduate School of Management, Northwestern University, Evanston IL 60208. E-mail: Swinkels@merle.acns.nwu.edu

that all events are positive probability, with events with higher numbers being less likely. Note that whatever the outcome on the player II dimension, the events are ordered in likelihood from top to bottom, and similarly when the outcome is fixed for player I. So, measured in this way, the relative likelihood of events on one coordinate is not affected by events on other coordinates, and so one might be tempted to conclude that the events are independent. Once again, the standard notion of independence runs into difficulty when the measure of size becomes too coarse.

However, the same thought experiment as before now operates. Set one of our randomizing devices to make ULT equally probable to MLB. Then, this device should also make UCT equally probable to MCB and so forth. Exactly as in section 5, this order is not extendible to cover the addition of these events. So, one of these devices must have the property that it equalizes two strategies for a given player conditional on one specification for the other player, but not conditional on some other specification, thus violating the independence condition. One can also prove an analog to our Theorem 3: An order will be extendible only if no cycle yields a chain of comparisons that are all "greater than" or all "less than" with at least one strict term.

REFERENCES

BATTIGALLI, P. [1992], "Strategic Independence, Generally Reasonable Extended Assessments and Consistent Assessments." Manuscript.

BLUME, L., BRANDENBURGER, A., AND E. DEKEL [1990] "Lexicographic Probabilities and Choice Under Uncertainty," *Econometrica*, 59, 61-80.

KOHLBERG, E. AND P. RENY [1992], "An Interpretation of Consistent Assessments." Manuscript.

KREPS, D. AND R. WILSON (1982), "Sequential Equilibria," *Econometrica*, 50, 863–894.

MYERSON, R. (1986), "Multistage Games with Communication," *Econometrica*, 54, 323-358.

Independence for Conditional Probability Systems[1]

JEROEN M. SWINKELS

J.L. Kellogg Graduate School of Management, Northwestern University, Evanston IL 60208
E-mail: Swinkels@merle.acns.nwu.edu

First Draft: November 27, 1993. This Draft: January 14, 1994

It has been argued that the natural notion of independence for conditional probability systems is weaker than sequential equilibrium's consistency condition. Kohlberg and Reny (1992) provide an attractive extra condition on a conditional probability system that implies equivalence to consistency. We provide an alternative condition and argue that it is a natural implication of independence. This condition also implies equivalence to consistency, and so is equivalent to Kohlberg and Reny's. However, the motivation and formulation of the condition are different, providing an additional viewpoint from which to understand consistency.

1. INTRODUCTION

In defining sequential equilibrium, Kreps and Wilson (1982) faced the problem of how to assign beliefs to nodes at information sets that have zero probability under the equilibrium strategy profile. In particular, the notion that players play independently rules out some beliefs. To capture an intuitively plausible set of conditions on beliefs, Kreps and Wilson introduced the notion of a *consistent assessment*: beliefs μ and strategy profile σ are consistent if there is a sequence of completely mixed strategy profiles σ^n converging to σ such that the limit of the Bayesian beliefs implied by σ^n on each information set converges to μ. As an expression of independence, it is hard to argue that this notion is too weak. It is less clear that its full strength is merited.

Perhaps the most natural language to talk about relative probabilities among zero probability events is a *conditional probability system* (Myerson (1986)). A conditional probability system specifies conditional probabilities for all subsets of the state space, even those with zero total probability (a formal definition is below). When the state space has a product structure, there is also a natural analog to the definition of independence for standard probability systems: the relative probability of events in one set of coordinates should not depend on events in the other coordinates. We term this *quasi-independence*. Every consistent assessment generates a quasi-independent conditional probability system. However, there exist quasi-independent conditional probability systems that cannot be generated as the limit of the fully mixed strategy profiles used to define consistency (see Blume, Brandenburger and Dekel (1991), Battigalli (1992), and Kohlberg and Reny (1992)). So, the seemingly natural definition of independence for conditional probability systems is

[1] I am very grateful to Roger Myerson, Phil Reny, Eddie Dekel, and Larry Samuelson for very helpful discussions and commentary. I also thank Bentley Macleod, David Levine, and Joseph Greenberg.

So, using Lemma 5,

$$\prod_{j=1}^{k} \rho(s^j \text{ vs. } t^j) = \prod_{j=1}^{k} \hat{\rho}(T^{D(s^j)}B^{D(t^j)} \text{ vs. } T^{D(t^j)}B^{D(s^j)}).$$

By Lemma 4 and the fact that we defined different devices for each s^j and t^j, this equals

$$\hat{\rho}(T^{D(s^1),\ldots,D(s^k)}B^{D(t^1),\ldots,D(t^k)} \text{ vs. } T^{D(t^1),\ldots,D(t^k)}B^{D(s^1),\ldots,D(s^k)}. \qquad (*)$$

For some $i \in N$ and $s_i \in S_i$, let a be a left hand side device and b a right hand side device corresponding to the same strategy s_i. Then, $\hat{\rho}(T^aB^b \text{ vs. } T^bB^a) = 1$.[11] So, using Lemma 4, (*) is equal to

$$\hat{\rho}(T^{D(s^1),\ldots,D(s^k)\backslash a}B^{D(t^1),\ldots,D(t^k)\backslash b} \text{ vs. } T^{D(t^1),\ldots,D(t^k)\backslash b}B^{D(s^1),\ldots,D(s^k)\backslash a} \cdot \hat{\rho}(T^aB^b \text{ vs. } T^bB^a)$$

$$= \hat{\rho}(T^{D(s^1),\ldots,D(s^k)\backslash a}B^{D(t^1),\ldots,D(t^k)\backslash b} \text{ vs. } T^{D(t^1),\ldots,D(t^k)\backslash b}B^{D(s^1),\ldots,D(s^k)\backslash a}) \cdot 1.$$

Because $((s^1,t^1),\ldots,(s^k,t^k))$ was a cycle, for each $i \in N$ and $s_i \in S_i$ the number of devices for s_i appearing on each side of (*) is the same. Thus, we can repeat this step, pairing off all the elements on each side of (*), and so showing that $\prod_{j=1}^{k} \rho(s^j \text{ vs. } t^j) = 1$. ∎

By Lemma 3 (Kohlberg and Reny's Lemma 3), ρ is thus an independent product. Lemma 2 states that if ρ is an independent product, then it is extendible. We have thus proven:

THEOREM 4: A relative probability ρ on $S = \Pi_{i \in N} S_i$ is extendible if an only if it is an independent product.

We conclude that being an independent product is the appropriate formulation of independence for conditional probability systems.

7. An Analogy

We suggested before that the problem with quasi-consistency arises because the real numbers do not allow one to distinguish different "levels of zero." To see this, return to a world of positive probability events. Consider attempting to understand and formulate the notion of independence with the concepts of greater, less, and equal, but without numbers, so that one cannot measure degrees of inequality. A probability measure is thus an order on events. Consider again Fig. 2, but now assume

---

[11] Note that $\hat{\rho}(s_i^*T^a \text{ vs. } s_iB^a) = \hat{\rho}(s_iB^b \text{ vs. } s_i^*T^b) = 1$. So, $\hat{\rho}(s_i^*T^aB^b \text{ vs. } s_iB^aB^b) \cdot \hat{\rho}(s_iB^aB^b \text{ vs. } s_iB^aB^b) \cdot \hat{\rho}(s_iB^bB^a \text{ vs. } s_i^*T^bB^a) = 1 = \hat{\rho}(T^aB^b \text{ vs. } T^bB^a)$.

weaker than consistency.

Kohlberg and Reny (1992) argue that quasi-independence does not capture the full strength of independence (readers are urged to read the original): Take the viewpoint of an outside observer, and consider the thought experiment of replicating the situation being studied. So, if the product space S $= \Pi_{i\in N}S_i$ is the space of strategy profiles in a game, then one imagines replicating both the players and the game, and running copies of the game in k separate rooms. Within each room, players are unable to observe each other's choices, and the rooms are identical so that choices cannot depend on which room a player is in. Then, Kohlberg and Reny argue, it should be possible to extend the original conditional probability system ρ on S to a conditional probability system ρ^k over joint outcomes in the k rooms such that (1) the marginal of ρ^k for on any given room is ρ independent of what occurs in other rooms, and (2) ρ^k depends only on how many copies of each player i chose each action s_i, and not on the rooms in which the choices were made, or what other players did in those rooms (so in particular, ρ^k is not changed when the room by room choices for copies of a particular player are permuted). Condition (1) can be thought of as capturing the idea that an outside observer is certain about ρ: no matter what the observer sees in rooms 1 through k-1, his beliefs about the outcome in room k remains ρ. Condition (2) in an "exchangability" notion: since the play of different players is independent, the outside observer should view it as equally likely that he sees s_1s_2 in room 1 and t_1t_2 in room 2 as that he sees s_1t_2 in room 1 and t_1s_2 in room 2.

Kohlberg and Reny show that ρ is extendible in this way if and only if it can be generated as the limit of completely mixed strategies profiles and so if and only if the conditional probability system satisfies the conditions implied by consistency. In this note, we present a different notion of "extendibility," and show that it has precisely the same implications as (and so is equivalent to) Kohlberg and Reny's. The interpretation however, is sufficiently different that it may provide a useful additional way of understanding independence for conditional probability systems and so of understanding consistency.

2. Conditional Probability Systems and Independence

We consider conditional probability systems (See Myerson (1986)). We follow Kohlberg and Reny in defining these in terms of *relative probabilities*:

DEFINITION 1: $\rho: S\times S \to \Re^+ \cup \infty$ is a *relative probability* on the finite set S if for any s,t,z in S,

(1.1) $\rho(s \text{ vs. } s) = 1$ and

(1.2) $\rho(s \text{ vs. } z) = \rho(s \text{ vs. } t)\rho(t \text{ vs. } z)$ whenever the right hand side is well defined.

The right hand side fails to be well defined when it involves the multiplication of ∞ and 0.

element of S. Extend ρ to $\hat{\rho}$ by adding for each $r \in \{s^1,\ldots,s^k\} \cup \{t^1,\ldots,t^k\}$, and each i in N a device d such that $\rho(s_i{}^* T^d \text{ vs. } r_i B^d) = 1$. Denote the set of devices added for r by D(r). By construction, the sets D(.) are disjoint. We will use the following two lemmas:

LEMMA 4: Let A', B', C', and D' be disjoint subsets of $N \cup \bigcup_{j=1}^{k}\left(D(s^j) \cup D(t^j)\right)$. Let A specify outcomes for the coordinates in A'.[10] Let B, C and D similarly specify outcomes for the coordinates in B', C', and D'. Then, $\hat{\rho}(A \text{ vs. } B)\hat{\rho}(C \text{ vs. } D) = \hat{\rho}(AC \text{ vs. } BD)$ whenever the left hand side is defined.

PROOF: Since the set of coordinates specified by C is disjoint from A or B, quasi-independence implies that $\hat{\rho}(A \text{ vs. } B) = \hat{\rho}(AC \text{ vs. } BC)$. Similarly, $\hat{\rho}(C \text{ vs. } D) = \hat{\rho}(BC \text{ vs. } BD)$. So, if $\hat{\rho}(A \text{ vs. } B)\hat{\rho}(C \text{ vs. } D)$ is well defined, then $\hat{\rho}(AC \text{ vs. } BC)\hat{\rho}(BC \text{ vs. } BD) = \hat{\rho}(AC \text{ vs. } BD)$ is also well defined, and equals the same thing. ∎

LEMMA 5: For all $j = 1,\ldots,k$, $\rho(s^j \text{ vs. } t^j) = \hat{\rho}(T^{D(s^j)}B^{D(t^j)} \text{ vs. } T^{D(t^j)}B^{D(s^j)})$.

PROOF: Let $r \in \{s^1,\ldots,s^k\} \cup \{t^1,\ldots,t^k\}$. For each $i \in N$ let d_i be the element of D(r) corresponding to r_i. Then, $\hat{\rho}(s_i{}^* T^{d_i} \text{ vs. } r_i B^{d_i}) = 1$ by construction. So, by repeated application of Lemma 4,

$$\rho(s^j \text{ vs. } t^j) = \prod_{i \in N} \hat{\rho}(s_i{}^* T^{d_i} \text{ vs. } r_i B^{d_i}) = \hat{\rho}(s^* T^{D(r)} \text{ vs. } rB^{D(r)}) = 1.$$

$$\begin{aligned}
\text{So, } \hat{\rho}(s^j \text{ vs. } t^j) &= 1 \cdot \hat{\rho}(s^j \text{ vs. } t^j) \cdot 1 \\
&= \hat{\rho}(s^* T^{D(s^j)} \text{ vs. } s^j B^{D(s^j)}) \cdot \hat{\rho}(s^j \text{ vs. } t^j) \cdot \hat{\rho}(t^j B^{D(t^j)} \text{ vs. } s^* T^{D(t^j)}) \\
&= \hat{\rho}(s^* T^{D(s^j)} B^{D(t^j)} \text{ vs. } s^j B^{D(s^j)} B^{D(t^j)}) \\
&\qquad \cdot \hat{\rho}(s^j B^{D(t^j)} B^{D(s^j)} \text{ vs. } t^j B^{D(t^j)} B^{D(s^j)}) \cdot \hat{\rho}(t^j B^{D(t^j)} B^{D(s^j)} \text{ vs. } s^* T^{D(t^j)} B^{D(s^j)}) \\
&= \hat{\rho}(s^* T^{D(s^j)} B^{D(t^j)} \text{ vs. } s^* T^{D(t^j)} B^{D(s^j)}) \\
&= \hat{\rho}(T^{D(s^j)} B^{D(t^j)} \text{ vs. } T^{D(t^j)} B^{D(s^j)}).
\end{aligned}$$
∎

---

[10] That is, let A be an element of $\prod_{i \in A' \cap S} S_i \times \prod_{d \in A' \cap \bigcup_{j=1}^{k}(D(s^j) \cup D(t^j))} \{T,B\}_d$.

For any ρ, there is a sequence $\{p^n\}$ of completely mixed probabilities on S such that

$\rho(r \text{ vs. } s) = \lim_{n\to\infty} \frac{p^n(r)}{p^n(s)}$ for each r and s in S (see Myerson (1986)). But, then for any $\varnothing \subseteq A,B \subseteq$ S, $\rho(A \text{ vs. } B) \equiv \lim_{n\to\infty} \frac{p^n(A)}{p^n(B)}$ is independent of which such sequence $\{p^n\}$ was chosen, and so ρ extends naturally to subsets of S.[2]

Assume that S has a product structure, so that $S = \Pi_{i\in N}S_i$, where N is a finite set. For $\varnothing \subseteq I \subseteq N$, let $S_I \equiv \Pi_{i\in I}S_i$. Let $\varnothing \subseteq A \subseteq S_J$, and $\varnothing \subseteq B \subseteq S_K$, where $\varnothing \subseteq J,K \subseteq N$. Then, define $\rho(A \text{ vs. } B)$ as $\rho(A \times S_{N\setminus J} \text{ vs. } B \times S_{N\setminus K})$. So, when only some coordinates of $\Pi_{i\in N}S_i$ are specified, it is understood that the other coordinates can vary freely.

Assume that $S = \Pi_{i\in N}S_i$. If ρ resulted from independent events over the S_i, then knowing what happened on components other than j should not affect the conditional probability of what happened on component j:

DEFINITION 2: The relative probability ρ on $S = \Pi_{i\in N}S_i$ is *individually quasi-independent* if for all $j \in N$ and r_j,t_j in S_j, $\rho(r_js_{N\setminus j} \text{ vs. } t_js_{N\setminus j})$ does not depend on $s_{N\setminus j} \in \Pi_{i\in N\setminus j}S_i$.

This agrees with Kohlberg and Reny's definition of quasi-independence. The "individually" is added to emphasize that only the action on one coordinate is allowed to change. In contrast:

DEFINITION 3: The relative probability ρ on $S = \Pi_{i\in N}S_i$ is *quasi-independent* if for all $\varnothing \subseteq J \subseteq N$ and r_J,t_J in $\Pi_{i\in J}S_i$, $\rho(r_Js_{N\setminus J} \text{ vs. } t_Js_{N\setminus J})$ does not depend on $s_{N\setminus J} \in \Pi_{i\in N\setminus J}S_i$.

---

[2]Let a* be a maximally likely element of A (i.e., $\rho(a \text{ vs. } a^*) \le 1$ for all $a \in A$), and b* of B. Then, $\lim_{n\to\infty} \frac{p^n(A)}{p^n(B)} = \lim_{n\to\infty} \frac{p^n(a^*)}{p^n(b^*)} \cdot \frac{\sum_{a\in A} \frac{p^n(a)}{p^n(a^*)}}{\sum_{b\in B} \frac{p^n(b)}{p^n(b^*)}}$. Now, $|A| > \sum_{a\in A} \lim_{n\to\infty} \frac{p^n(a)}{p^n(a^*)} \ge 1$ where the first inequality is because a* is maximally likely, and the second because $a^* \in B$, and similarly for $\sum_{b\in B} \lim_{n\to\infty} \frac{p^n(b)}{p^n(b^*)}$. So,

$$\lim_{n\to\infty} \frac{p^n(A)}{p^n(B)} = \rho(a^* \text{ vs. } b^*) \cdot \frac{\sum_{a\in A} \rho(a \text{ vs. } a^*)}{\sum_{b\in B} \rho(b \text{ vs. } b^*)}.$$

system in the example. Assume the conditional probability system in Fig. 2 is independently extendible. Let s* = UL. Add four devices denoted M,D,C, and R. Then there is a quasi-independent $\hat{\rho}$ such that $\hat{\rho}(s^*T^M \text{ vs. } MLB^M) = \hat{\rho}(s^*T^D \text{ vs. } DLB^D) = \hat{\rho}(s^*T^C \text{ vs. } UCB^C) = \hat{\rho}(s^*T^R \text{ vs. } URB^R) = 1$. So, the M device measures the relative likelihood of T and M, and the D device the relative probability of T and D. Similarly the C and R devices measure the relative likelihood of C and R versus L. Now, since $\hat{\rho}(\text{ML vs. UC}) = \infty$, it is intuitively reasonable (and can be proven, see Lemma 5 below) that $\hat{\rho}(T^M B^C \text{ vs. } T^C B^M) = \infty$. Similarly, since $\hat{\rho}(\text{UR vs. DL}) = \infty$, $\hat{\rho}(T^R B^D \text{ vs. } T^D B^R) = \infty$. So, $\hat{\rho}(T^M T^R B^C B^D \text{ vs. } T^C T^D B^M B^R) = \infty$ (see Lemma 4 below). But, this implies $\hat{\rho}(\text{MR vs. DC}) = \infty$ a contradiction.

6. Extendibility and Consistency

In this section, we show that a relative probability ρ is independently extendible if and only if it is an independent product, and so generates a consistent assessment. This follows quite directly from the following Definition and Lemma from Kohlberg and Reny:

DEFINITION 6: A sequence $((s^1,t^1),\ldots,(s^k,t^k))$ of pairs of elements of $S = \Pi_{i\in N}S_i$ is a *cycle* if $(s_i^1,\ldots,s_i^k)$ is a permutation of $(t_i^1,\ldots,t_i^k)$ for each i in N.

LEMMA 3 (KOHLBERG AND RENY LEMMA 3): A necessary and sufficient condition for ρ to be an independent product is that for any cycle $((s^1,t^1),\ldots,(s^k,t^k))$, $\prod_{j=1}^k \rho(s^j \text{ vs. } t^j) = 1$ whenever it is defined.

PROOF: Non-trivial. See Kohlberg and Reny (1992).

We now prove:

THEOREM 3: If ρ is independently extendible, then for any cycle $((s^1,t^1),\ldots,(s^k,t^k))$, $\prod_{j=1}^k \rho(s^j \text{ vs. } t^j) = 1$ whenever it is defined.

PROOF: Fix a cycle $((s^1,t^1),\ldots,(s^k,t^k))$ such that $\prod_{j=1}^k \rho(s^j \text{ vs. } t^j)$ is defined. Let s* be an arbitrary

REMARK 1: To avoid ambiguity, we will explicitly write down the product space with respect to which ρ is quasi-independent. So, when we say that ρ is quasi-independent on $S \times T$, we will mean precisely that for all r and s in S, ρ(rt vs. st) does not depend on t in T, and for all u and v in T, ρ(su vs. sv) does not depend on s in S.[3] Any product structure of S or T is ignored. So, if S itself has a product structure, say $S = \Pi_{i\in N}S_i$, quasi-independence on $S \times T$ does not mean that for example $\rho(r_J s_{N\setminus J}t \text{ vs. } t_J s_{N\setminus J}t)$ cannot depend on $s_{N\setminus J}t$ in $\Pi_{i\in N\setminus J}S_i \times T$. When we intend that ρ is quasi-independent also with respect to the coordinates of S, we will write that ρ is quasi-independent on $\Pi_{i\in N}S_i \times T$.

Quasi-independence agrees with Battigalli's notion of independence. An equivalent definition of quasi-independence is that for all $\varnothing \subseteq J \subseteq N$, r_J, t_J in $\Pi_{i\in J}S_i$, and $s_{N\setminus J} \in \Pi_{i\in N\setminus J}S_i$, $\rho(r_J s_{N\setminus J} \text{ vs. } t_J s_{N\setminus J}) = \rho(r_J \text{ vs. } t_J)$. For standard probability theory, if ρ is individually quasi-independent then it is quasi-independent (as in particular, both agree with independence). For conditional probability systems, this need not hold:[4]

II

I		L	R
	T	z^1	z^2
	B	z^3	z^4

U

III

II

I		L	R
	T	z^5	z^7
	B	z^6	z^8

D

Fig. 1: An individually quasi-independent but not quasi-independent conditional probability system.

Define a relative probability on the events indicated in Fig. 1 by setting $\rho(z^i \text{ vs. } z^j)=0$ whenever $i>j$. This conditional probability system is individually quasi-independent, because regardless of the actions of the other players, I always chooses T arbitrarily more often than B, II L over R, and III U over D. However, it is not quasi-independent, because the likelihood of BL vs. TR by I and II depends on whether III chooses U or D.

We work with quasi-independence, and so in this dimension, our conditions are stronger than Kohlberg and Reny's. However, it seems likely that any motivation for individual quasi-independence would also be a motivation for quasi-independence.[5]

---

[3]As Kohlberg and Reny note, these two conditions are not redundant for conditional probability systems.

[4]I am grateful to Roger Myerson for pointing this out and providing the example.

[5]It is interesting to consider the notion of independence put forward by Blume, Brandenburger and Dekel (1991). Say that ρ is BBD-independent if for each $i \in N$,

$\rho(s_{N\setminus i}s_i \text{ vs. } t_{N\setminus i}s_i)$

So, non-extendibility relative to a set $\{(s^j,t^j)\}_{j\in K}$ does not reflect any incoherence in the thought experiment, but rather the inability to move from quasi-independence on $S\times\Pi_{j\in K}\{T,B\}_j$ to quasi-independence on $\Pi_{i\in N}S_i\times\Pi_{j\in K}\{T,B\}_j$. By Lemma 1, this is precisely the statement that the independence condition fails somewhere.

Extendibility will play the same role in this paper as Kohlberg and Reny's requirement that ρ can be extended to a quasi-independent k-fold product of ρ satisfying coordinate wise exchangability. In fact, since we will show that for a relative probability ρ to be quasi-independent and independently extendible is equivalent to ρ being an independent product, and since the ability to extend ρ to k-fold products satisfying coordinate wise exchangability has precisely the same implications, it must be that the two conditions are equivalent.[9] So, one can view the purpose of this paper as purely to provide a different way of understanding and motivating the requirement that ρ can be extended to an exchangeable k-fold product, not to provide radically different axioms or an independent proof.

LEMMA 2: If ρ is an independent product, then ρ is independently extendible.

PROOF: Obvious from the construction in the proof of Theorem 2, noting that since ρ is an independent product, $\{p^n\}$ can itself be taken as a sequence of product probabilities. ■

Note that we have in fact proven something stronger: If ρ is an independent product, then it can always be extended, with the extension *also* being an independent product. So, an alternative motivation for extendibility is simply as a coherence condition on an independence definition: Whatever the definition of independence, a system that is independent under this definition should be extendible to include arbitrary extra dimensions also independent under this definition. The results of this paper say that product independence is coherent in this sense, while any definition of independence that is at least as strong as quasi-independence and coherent in this sense must in fact be at least as strong as product independence.

5. THE EXAMPLE REVISITED

We begin by showing how the notion of extendibility rules out the conditional probability

[9]This equivalence becomes even clearer if instead of introducing binary randomizing devices, we introduced devices with an arbitrary (finite) number of outcomes, in which the relative likelihood of successive pairs of outcomes are determined by pairs of events in S. Then, one could effectively make devices into copies of players and Kohlberg and Reny's exchangability would follow directly from extendibility. We think it is somewhat simpler given our interpretation to think about binary devices, and we like the fact that our proof goes directly to Kohlberg and Reny's Lemma 3, which we consider more fundamental.

Finally,

DEFINITION 4: The relative probability ρ on $S = \Pi_{i \in N} S_i$ is an *independent product* if there is a sequence $\{p_i^n\}$ of strictly positive probabilities on each S_i such that for all $r,s \in S$,

$$\rho(r \text{ vs. } s) = \lim_{n \to \infty} \frac{\prod_{i \in N} p_i^n(r_i)}{\prod_{i \in N} p_i^n(s_i)}.$$

The key to the connection between consistency and conditional probability systems is that consistent assessments and independent products are essentially the same thing:

THEOREM 1: Let Γ be an extensive form game, and S its normal form. Every independent product on $\Pi_{i \in N} S_i$ induces a consistent assessment on Γ. Every consistent assessment on Γ is induced by an independent product on $\Pi_{i \in N} S_i$.

We sketch a proof. Let ρ be an independent product on S. Consider any terminal node x in Γ. Nature, who is assumed to move only at the beginning of the game, has one move consistent with reaching x. Let $\upsilon(x)$ be the probability nature assigns to that move. Associated with x will be a subset S(x) of S, such that x is reached if and only if players choose from S(x) and nature chooses appropriately. So, ρ induces a conditional probability system on terminal nodes given by $\frac{P(x \text{ reached})}{P(y \text{ reached})} = \rho(S(x) \text{ vs. } S(y)) \cdot \frac{\nu(x)}{\nu(y)}$. This extends in the standard way to beliefs at information sets and distributions over actions at each node. Since ρ was an independent product, this conditional probability system can be seen as coming from the limit of completely mixed strategy profiles, and so by Kuhn's theorem, from a sequence of completely mixed behavior strategy profiles. Thus, the distribution of actions is the same at all nodes in a given information set. So, ρ induces a well defined behavior strategy and beliefs on Γ, and these beliefs and strategy are consistent.

Conversely, consider any consistent assessment (μ,σ), and the generating sequence of completely mixed strategy profiles. This generates a sequence $\{p^n\}$ of completely mixed strategy

---

does not depend on s_i. In contrast to the distinction between individual quasi-independence and quasi-independence, this restriction has a nice interpretation: player i does not find his own strategy choice to be informative about the choice which other players will make. So, if BBD-independence were weaker than quasi-independence, it would be so in an interesting way. In fact, however, BBD-independence and quasi-independence are equivalent: Consider $\rho(s_J s_{N \setminus J} \text{ vs. } t_J s_{N \setminus J})$, and let $t_{N \setminus J} \in S_{N \setminus J}$. Observe that one can move from $s_{N \setminus J}$ to $t_{N \setminus J}$ in $|N \setminus J|$ steps, where at each step, one player i in $N \setminus J$ changes his action from s_i to t_i. For each of these steps, BBD's condition implies that ρ doesn't change. So, $\rho(s_J s_{N \setminus J} \text{ vs. } t_J s_{N \setminus J}) = \rho(s_J t_{N \setminus J} \text{ vs. } t_J t_{N \setminus J})$.

So, if the independence condition is to hold whatever the devices we add, it must be that ρ is *independently extendible* in the following sense:

DEFINITION 5: A relative probability ρ on $S = \Pi_{i\in N}S_i$ is *independently extendible* if for any finite set $\{(s^j,t^j)\}_{j\in K}$ of pairs from S, there is a quasi-independent relative probability $\hat{\rho}$ on $\Pi_{i\in N}S_i \times \Pi_{j\in K}\{T,B\}_j$ such that

(5.1) for all $j\in K$, $\hat{\rho}(s^jT^j \text{ vs. } t^jB^j) = 1$, and,

(5.2) for all s,t in S, $\hat{\rho}(s \text{ vs. } t) = \rho(s \text{ vs. } t)$.

That is, ρ is independently extendible if for any finite set of extra dimensions, and pre-specified relative probabilities on those dimensions, the measure can be extended to those dimensions while retaining quasi-independence. If ρ is not independently extendible, then if we could build the randomizing devices described, the independence condition would fail at least once. So, a non-extendible conditional probability system is not truly consistent with independence. Conversely, if ρ is independently extendible, then whatever set of such randomizing devices we imagine, a conditional probability system consistent with these devices is guaranteed to exist.

The reader might worry that the non-extendibility of ρ might indicate not that ρ is not independent, but rather that the thought experiment of adding independent calibrated devices to S was somehow incoherent. We thus show:

THEOREM 2: Let ρ be a relative probability on S and let $\{(s^j,t^j)\}_{j\in K}$ be an arbitrary finite set of pairs of elements of S. Then, there is a quasi-independent relative probability $\hat{\rho}$ on $S \times \Pi_{j\in K}\{T,B\}_j$ (but not necessarily on $\Pi_{i\in N}S_i \times \Pi_{j\in K}\{T,B\}_j$) such that

(1) for all $j\in K$, $\hat{\rho}(s^jT^j \text{ vs. } t^jB^j) = 1$, and,

(2) for all s,t in S, $\hat{\rho}(s \text{ vs. } t) = \rho(s \text{ vs. } t)$.

PROOF: Let $\{p^n\}$ be a sequence of completely mixed probabilities on S generating ρ. For $j \in K$, define q_j^n by $\frac{q_j^n}{1-q_j^n} \cdot \frac{p^n(t^j)}{p^n(s^j)} = 1$. For each $s \in S$ and $r \in \Pi_{j\in K}\{T,B\}_j$, define $\hat{p}^n(s,r) =$

$p^n(s) \cdot \prod_{j\in K|r_j=T} q_j^n \cdot \prod_{j\in K|r_j=B} (1-q_j^n)$, and let $\hat{\rho}$ be the conditional probability system generated by $\{\hat{p}^n\}$. Then, $\hat{\rho}$ is an independent product on $S \times \Pi_{j\in K}\{T,B\}_j$, and so certainly quasi-independent. The verifications of (1) and (2) are simple algebraic exercises. ∎

profiles on S. Now, $\lim_{n\to\infty} \frac{p^n(r)}{p^n(s)}$ may not be well defined for all r and s in S, but $\lim_{n\to\infty} \frac{p^n(S(x))}{p^n(S(y))}$ will be for all terminal nodes x and y. Taking an appropriate subsequence of $\{p^n\}$ thus generates a product measure that induces (μ,σ).

So, justifying consistency amounts to justifying the condition that conditional probability systems should be independent products.

3. An Example

Consider Fig. 2. A conditional probability system for the space of strategy profiles is given by the entries z^1-z^9, with the interpretation that $\rho(z^i \text{ vs. } z^j)=0$ whenever $i>j$.[6]

		II L	II C	II R
I	U	z^1	z^3	z^4
I	M	z^2	z^5	z^8
I	D	z^6	z^7	z^9

Fig. 2: A quasi-independent but inconsistent belief system.

To see that ρ is quasi-independent, note that however one conditions, U is arbitrarily more likely than M which is arbitrarily more likely than D, and similarly for player II. However, ρ is not an independent product. In particular, it is inconsistent that ρ (ML vs. UC) = ρ (DC vs. MR) = ∞, but ρ (DL vs. UR) =0.[7] Battigalli (1992) and Kohlberg and Reny (1992) provide examples of 3 player games with this general structure in which ρ (ML vs. UC), ρ (DC vs. MR), and ρ (DL vs. UR) correspond to player III's beliefs about players I and II at different information sets.

---

[6]This example is very similar to one in Battigalli (1992), who credits Reny. Similar examples appear in Blume, Brandenburger and Dekel (1991), who credit Myerson, and in Kohlberg and Reny (1992).

[7]Assume the system is an independent product. Then $\lim_{n\to\infty} \frac{\prod_{i\in N} p_1^n(M)p_2^n(L)}{\prod_{i\in N} p_1^n(U)p_2^n(C)} = \rho(\text{ML vs. UC}) = \infty$, and $\lim_{n\to\infty} \frac{\prod_{i\in N} p_1^n(D)p_2^n(C)}{\prod_{i\in N} p_1^n(M)p_1^n(R)} = \rho(\text{DC vs. MR}) = \infty$. So, $\lim_{n\to\infty} \frac{\prod_{i\in N} p_1^n(D)p_2^n(L)}{\prod_{i\in N} p_1^n(U)p_2^n(R)} = \infty$, contradicting $\rho(\text{DL vs. UR}) = 0$.

However, it does seem valid to ask: *for what types of conditional probability system do we know even without building such devices that if we could, they would expose a problem?*

To answer this question, we must be more formal about our thought experiment. Let ρ be a conditional probability system on $S = \Pi_{i \in N} S_i$. Consider adding one idealized randomizing device for each element of a finite set K, and consider the conditional probability system $\hat{\rho}$ describing relative probabilities over $S \times \Pi_{j \in K}\{T,B\}_j$. Regardless of how the devices are set, $\hat{\rho}$ must satisfy three conditions. First, since the devices are independent of S and of each other, and since we have taken quasi-independence as a necessary condition for independence, $\hat{\rho}$ must be quasi-independent on $S \times \Pi_{j \in K}\{T,B\}_j$. Second is the key condition described above:

INDEPENDENCE CONDITION: Let $M \subseteq N$, let $s_M, t_M \in \Pi_{i \in M} S_i$ and let $u,v \in \Pi_{j \in K}\{T,B\}_j$. Then, $\hat{\rho}(s_M s_{N\backslash M} u$ vs. $t_M s_{N\backslash M} v)$ does not depend on $s_{N\backslash M} \in \Pi_{i \in N\backslash M} S_i$.

If $\hat{\rho}$ fails the independence condition, then knowing $s_{N\backslash M}$ gives information about the relative likelihood of s_M vs. t_M (albeit information that may be lost by ρ), and so we conclude that the outcomes on M and N\M are not independent.

Finally, for any fixed outcome of the devices, the induced conditional probability system on S must be ρ.

LEMMA 1: Quasi-independence of $\hat{\rho}$ on $S \times \Pi_{j \in K}\{T,B\}_j$ plus the independence condition is equivalent to quasi-independence of $\hat{\rho}$ on $\Pi_{i \in N} S_i \times \Pi_{j \in K}\{T,B\}_j$. (Recall Remark 1.)

PROOF: Let $\varnothing \subseteq I \subseteq K$, and let $\varnothing \subseteq M \subseteq N$. Let s_M and t_M be elements of $\Pi_{i \in M} S_i$, and similarly for $s_{N\backslash M}$, $t_{N\backslash M}$, t_I, s_I, $s_{K\backslash I}$ and $t_{K\backslash I}$. We need to show that

$$\hat{\rho}(t_M t_I t_{N\backslash M} t_{K\backslash I} \text{ vs. } s_M s_I t_{N\backslash M} t_{K\backslash I}) = \hat{\rho}(t_M t_I s_{N\backslash M} s_{K\backslash I} \text{ vs. } s_M s_I s_{N\backslash M} s_{K\backslash I}).$$

But, since $\hat{\rho}$ is quasi-independent on $S \times \Pi_{k \in K}\{T,B\}_j$,

$$\hat{\rho}(t_M t_I t_{N\backslash M} t_{K\backslash I} \text{ vs. } s_M s_I t_{N\backslash M} t_{K\backslash I}) = \hat{\rho}(t_M t_I t_{N\backslash M} s_{K\backslash I} \text{ vs. } s_M s_I t_{N\backslash M} s_{K\backslash I}),$$

and, by the independence condition, the right hand side is equal to $\hat{\rho}(t_M t_I s_{N\backslash M} s_{K\backslash I}$ vs. $s_M s_I s_{N\backslash M} s_{K\backslash I})$. The converse is obvious. ■

So, specify any set of pairs $\{(s^j,t^j)\}_{j \in K}$. Denote by T^j the event that the outcome on coordinate j is T, and B^j the event that the outcome is B. Calibrate each device j such that $\hat{\rho}(s^j T^j$ vs. $t^j B^j) = 1$. If $\hat{\rho}$ satisfies the independence condition, then $\hat{\rho}$ is a quasi-independent measure on $\Pi_{i \in N} S_i \times \Pi_{j \in K}\{T,B\}_j$ where for each j, $\hat{\rho}(s^j T^j$ vs. $t^j B^j) = 1$, and for all s and t in S, $\hat{\rho}(s$ vs. $t) = \rho(s$ vs. $t)$.

So, consistency imposes restrictions on information sets not implied by quasi-independence.

4. Independence Revisited

Either quasi-independence does not capture all the natural restrictions of independence, or consistency lacks a coherent foundation in independence. In this section, we argue that there are natural implications of independence not captured by quasi-independence. In Figure 2, ρ(Mx vs. Ux) = 0 for x = L, M, and R. None-the-less, we are not satisfied that player I chooses U vs. M independently of what II does. Heuristically, we have no way of knowing if the zero's are the "same size." Of course, one zero being smaller than another is meaningless in the context of the real numbers. However, note that for example ρ(DL vs. UL) is also zero, but on an intuitive level this zero is "smaller" than the zero corresponding to ρ(ML vs. UL) since ρ(ML vs. DL) is infinite. To formalize this, we need to develop a way to "measure" the zeros.

We do this by the thought experiment of introducing idealized randomizing devices with outcome space {T,B}, and with the property that the outcomes of the devices are independent of the outcomes on S and each other.[8] In Fig. 2 for example, U was arbitrarily more likely than M regardless of whether one conditioned on L, C, or R. So, consider adding one of these devices, and comparing the relative probability of the outcome of the game being Ux and the device generating T vs. the outcome of the game being Mx and the device generating B. *It seems to us that a clear necessary condition for the independence of I's choices from II's choices is that this relative probability should not depend on whether x is L, C or R.*

Assume that for any r and s in S, the device can be calibrated to generate T and B in such a way that r and T is equally likely to s and B. So, the device "measures" the relative likelihood of r and s and thus gives us the ability to test whether the relative likelihoods discussed are really the same: Say that when the device is set to make URT equally likely to MRB it turns out that UCT is twice as likely as MCB. Then it seems clear that player I chooses M half as often when player II chooses C as he does when player II chooses L. So, player II's choice of L vs. C contains information about player I's action and I and II are not truly playing independently. However, this information is too fine to survive the language of a conditional probability system: even though player I chooses M half as often when player II chooses C as when she chooses L, it remains the case that given either L or C, M is arbitrarily less likely that U, but arbitrarily more likely than D. Essentially, ρ(MC vs. UC) = $\frac{1}{2} \cdot 0 = 0 = \rho$(ML vs. UL).

Of course, actually building such devices and conducting this test may be problematic.

---

[8]We will take this to mean at least quasi-independence.