CHARACTERIZING CONSISTENCY WITH MONOMIALS

Peter A. Streufert
University of Western Ontario
pstreuf@uwo.ca
June 26, 2007

Abstract. By definition, an assessment is consistent iff it is the limit of a sequence of full-support assessments, each of which satisfies Bayes Rule. Such sequences are routinely constructed by assigning monomials to actions, and it is important to know that such monomials are not only sufficient, but also necessary, for consistency.

This paper shows that this equivalence can be derived by linear algebra alone. In addition, the paper repairs a nontrivial fallacy in the proofs of the Kreps-Wilson theorems regarding sequential equilibrium. Both of these observations flow naturally from a new perspective on the monomials' exponents: their sums constitute an additive representation of an infinite-relative-probability relation among the nodes.

1. Introduction

1.1. For the Initiated

As defined in Kreps and Wilson (1982, henceforth `KW`), an assessment is consistent iff it is the limit of a sequence of full-support assessments, each of which satisfies Bayes Rule. It is routine to construct such a

I thank Dave Kreps and Bob Wilson for their encouraging comments, and also thank Ian King, Val Lambson, Bart Lipman, Andy McLennan, and seminar participants at Auckland, Sydney, New South Wales, the Australasian Economic Theory Workshop, and the International Conference on Game Theory at Stony Brook. Finally, I thank the Auckland Economics Department for its hospitality during my sabbatical. (This paper subsumes half of Streufert (2006a) and all of Streufert (2006b).)

sequence by first assigning, to each action a, a monomial of the form $c(a)n^{e(a)}$ in which the coefficient $c(a)$ is a positive real and the exponent $e(a)$ is a nonpositive integer (n will index the sequence). The construction is then completed (at each n) by normalizing the monomials to create a strategy profile and by applying Bayes Rule to create a belief system.

It is clear that an assessment is consistent if it is the limit of an assessment sequence that is constructed in this fashion. This statement and its converse appear here as Theorem 2.1, and the chief contribution of this paper is to show that that converse can be derived with nothing more than linear algebra.

The simplicity of this proof comes from a shift in perspective. To find this fresh perspective, first recall from experience that summing the exponents along the paths leading to the nodes in an information set determines the support of the belief at that information set (this abstracts from second-order terms like the n^{-3} in the numerator at node oG in Figure 2.3(b)). In other words, such sums of exponents across actions determine which nodes are infinitely more likely than other nodes. Or, in still other words, such sums across actions constitute an additive integer representation of a binary relation expressing infinite relative probability among the nodes.

From this perspective, the crux of this paper is to show that every consistent assessment admits an additive integer representation of its infinite-relative-probability relation. This task is unexpectedly easy: Section 3.1 does it simply by mimicking the proof of Scott's Theorem, which is a well-known additive-representation theorem from the mathematical-psychology literature (see Appendix A). In particular, Scott's Theorem is proven by linking Farkas' Lemma to cancellation laws. Analogously, Section 3.1 links Farkas' Lemma to similar conditions which follow directly from the definition of consistency (Appendix A explores this analogy further).

In addition, the new perspective reveals a nontrivial fallacy in the proof of `KW` Lemma A1. That proof derives a representation for any consistent assessment's infinite-relative-probability relation. However, it fails to show that that representation is additive. Section 4 repairs the situation by applying Theorem 2.1. This repair is important because `KW` Lemma A1 is an essential component of the `KW` theorems which derive the generic finiteness of the set of sequential-equilibrium outcomes, and also the generic equality of that set to the set of perfect-equilibrium outcomes.

Theorem 2.1 is closely related to Theorem 3.1 of Perea y Monsuwe, Jansen, and Peters (1997, henceforth `PJP`) (see Section 5.1). The proof here is more intuitive than the `PJP` proof in the sense that it mimics `KW` in constructing an infinite-relative-probability relation, recognizes the issue of additive representation, and then mimics Scott's Theorem. In addition, it is more economical in its use of mathematics because it employs Farkas' Lemma from linear algebra rather than the Separating Hyperplane Theorem from analysis.

Finally, and much less significantly, this paper departs from `KW` and `PJP` by expressing its results in terms of the monomial constructions with which economists have grown familiar. And it happens that consistency can be characterized by the monomials themselves, without any reference to the assessment sequences that were constructed in the first paragraph. This tertiary contribution (Theorem 2.1(a$\Leftrightarrow$c)) is welcome computationally (see Figure 2.3's example), and lies behind Subsection 1.2's casual suggestion that the space of monomials be regarded as an extended set of probability numbers that is roughly analogous to the space of complex numbers (the space of monomials is comparatively simple because there is no addition operator and no multiplicative inverse).

Section 2 states and illustrates Theorem 2.1, Section 3 proves it, Section 4 applies it to `KW`, and Section 5 relates it to `PJP` and the remainder of literature (Subsection 5.2 gratefully acknowledges a large hidden debt to Kohlberg and Reny (1997)).

1.2. For the Uninitiated

Imagine that you have identical twin daughters. They're ten years old, they like to giggle, and you are faced with the daunting task of putting them to bed. For some reason, you cannot distract them with songs, stories, or books, and further, you cannot reward them for good behaviour. Instead, all you can do is to notice whether or not there is noise, and then in the event of noise, choose between two punishments, the first of which is harder on the first girl and the second of which is harder on the second girl. Because your twins are identical, you cannot distinguish between their giggles, and thus, you cannot distinguish between the first girl giggling, the second girl giggling, and the two of them giggling simultaneously. All you can do is to notice whether or not there is noise. (This paragraph is a verbal description of the "game form" that will be defined in Figure 2.1(a).)

Nonetheless, you and the girls might come to a mutual understanding under which both of the girls will behave themselves. For example, suppose the girls know that you would believe that the second girl alone is at fault if either or both of them giggle. This belief might then induce you to choose the second punishment, and the threat of the second punishment might be sufficient to induce both girls to behave. (An "assessment" is just a list of strategies and beliefs, and accordingly, this paragraph is a verbal description of the assessment to be defined in Figure 2.1(b).)

Other mutual understandings might also develop. For example, suppose instead that the girls know that you would believe that both girls are giggling simultaneously if either or both of them giggle. This belief might induce you to randomize between the two punishments, and the threat of this randomized punishment might be sufficient to induce both girls to behave. (This alternative assessment does not appear in a figure.)

In the above examples, it is unclear how the parent's belief should accord with the girls' behaviour. Since the girls' behaviour implies zero probability at each of the three possibilities, we might say that any probability distribution over the three possibilities accords with the girls' behaviour. This reasoning would admit both of the above assessments. (Ordinary probability theory can take us no further.)

Alternatively, we might be more restrictive. We might think that both girls giggling is infinitely less likely than either of the two giggling alone, simply because the coincidence of two zero-probability events seems infinitely less likely than either zero-probability event alone. This would rule out the second of the two assessments.

This restriction is incorporated into the concept of consistency which Kreps and Wilson (1982, henceforth `KW`) introduced as part of their path-breaking sequential-equilibrium concept. Their definition states that an assessment is "consistent" iff it is the limit of a sequence of full-support (i.e. positive-valued) assessments which satisfy Bayes Rule (Bayes Rule is part of ordinary probability theory and works well when all probabilities are positive).

This paper's theorem provides an alternative way of understanding `KW`'s concept of consistency. Broadly speaking, the theorem suggests that it is useful to represent each probability with a "monomial," which this paper defines to be an algebraic expression of the form cn^e in which c is a positive real number and e is a nonpositive integer. Monomials are vaguely like complex numbers in the sense that both extend the set of real numbers into a second dimension, and accordingly, a monomial cn^e is a real number when $e=0$ just like a complex number $a+bi$ is a real number when $b=0$. Essentially, monomials with zero exponents express positive probabilities and monomials with negative exponents express different levels of zero probability (lesser, i.e. more negative, exponents express lesser zero probabilities).

After working a number of examples (such as those of Section 2), it becomes intuitive that the definition of consistency is satisfied if a monomial $c(a)n^{e(a)}$ can be assigned to each action a in such a way that (1) the action a is played with probability $c(a)$ if the exponent $e(a)$ is zero and is not played if $e(a)$ is negative, and, (2) the belief at each information set is found, first by calculating the product of the monomials along the path leading to each of the nodes in the information set, second by placing zero probability on every node whose product's exponent is less than that of another node, and finally by assigning positive probability over the remaining nodes in proportion to their products' coefficients.

Theorem 2.1(a⇔c) formalizes that intuition, and much more importantly, includes its converse. In other words, the theorem states that consistency is equivalent to the existence of such monomials. This is important because monomials are more tractable than the sequences that appear in the definition of consistency (a monomial is specified by two numbers while a sequence is specified by an infinity of numbers).

Time-conscious uninitiated readers might peruse Subsection 2.1 with an emphasis on Figure 2.1 (the definition of consistency can be bypassed) and then read Subsection 2.2 with an emphasis on Theorem 2.1(a⇔c) (condition (b) can be bypassed). Subsection 2.3 returns to the story of the twins.

2. Theorem

2.1. Basic Definitions

This section recapitulates some notation and terminology from `KW` while discussing an example which will be used throughout the paper. Every symbol in `KW` means the same thing here, with the exception of $\prec$, which is the game tree's precedence relation in `KW` and which will be an assessment's infinite-relative-probability relation in Section 3.

This paper takes as exogenous a "game form with initial probabilities." By that is meant an extensive form as defined on `KW` page 868, with perfect recall, in which players are identified with information sets, together with a probability distribution over the initial nodes.

Such a game form with initial probabilities is defined by this paragraph and Figure 2.1(a) (this game form resembles Kreps and Ramey (1987, Figure 1)). The example corresponds to Subsection 1.2's story of twins going to sleep if one imagines that the first twin chooses to G (giggle) or S (sleep), that the second twin chooses to g (giggle) or s (sleep), and that the parent chooses between the punishment δ (which is harder on their first daughter) and ε (which is harder on the second). In this example, the set T of nodes t contains the set $X = \{o, oG, oS, oGg, oGs, oSg\}$ of decision nodes x, which in turn contains the set $W = \{o\}$ of initial nodes w. The set W is given the trivial distribution $\rho = (\rho(o)) = (1)$, and the set X is partitioned into the information sets $h \in H = \{\{o\}, \{oG, oS\}, \{oGg, oGs, oSg\}\}$. Further, the set $A = \{G, S, g, s, \delta, \varepsilon\}$ is the set of actions a. As in KW, $H(x)$ is the information set h which contains x, $A(h)$ is the set of actions available from information set h, and $\alpha(t)$ is the last action taken to reach a non-initial node t.

We now turn to the endogenous variables. A strategy profile is a function $\pi{:}A{\rightarrow}[0,1]$ such that $(\forall h)\ \Sigma_{a \in A(h)} \pi(a) = 1$. Let

$$\Pi^x \rho{\cup}\pi = \rho{\circ}p_{\ell(x)}(x) \times \Pi_{k=0}^{\ell(x)-1} \pi{\circ}\alpha{\circ}p_k(x)$$

denote the product of the initial probability and the strategies leading to node x (as in KW, $p_k(x)$ is the kth predecessor of node x, and $\ell(x)$ is the number of its predecessors). Thus the number $\Pi^x \rho{\cup}\pi$ is the probability that node x will be reached (while KW denotes this probability by $P^\pi(x)$, this paper frequently uses pathwise products of the form $\Pi^x f{\cup}g$ as defined in Appendix B).

A belief system is a function $\mu{:}X{\rightarrow}[0,1]$ such that $(\forall h)\ \Sigma_{x \in h} \mu(x) = 1$. An assessment is a strategy-belief pair (π, μ). As on KW page 872, let Ψ^0 consist of those full-support (i.e. positive-valued) assessments

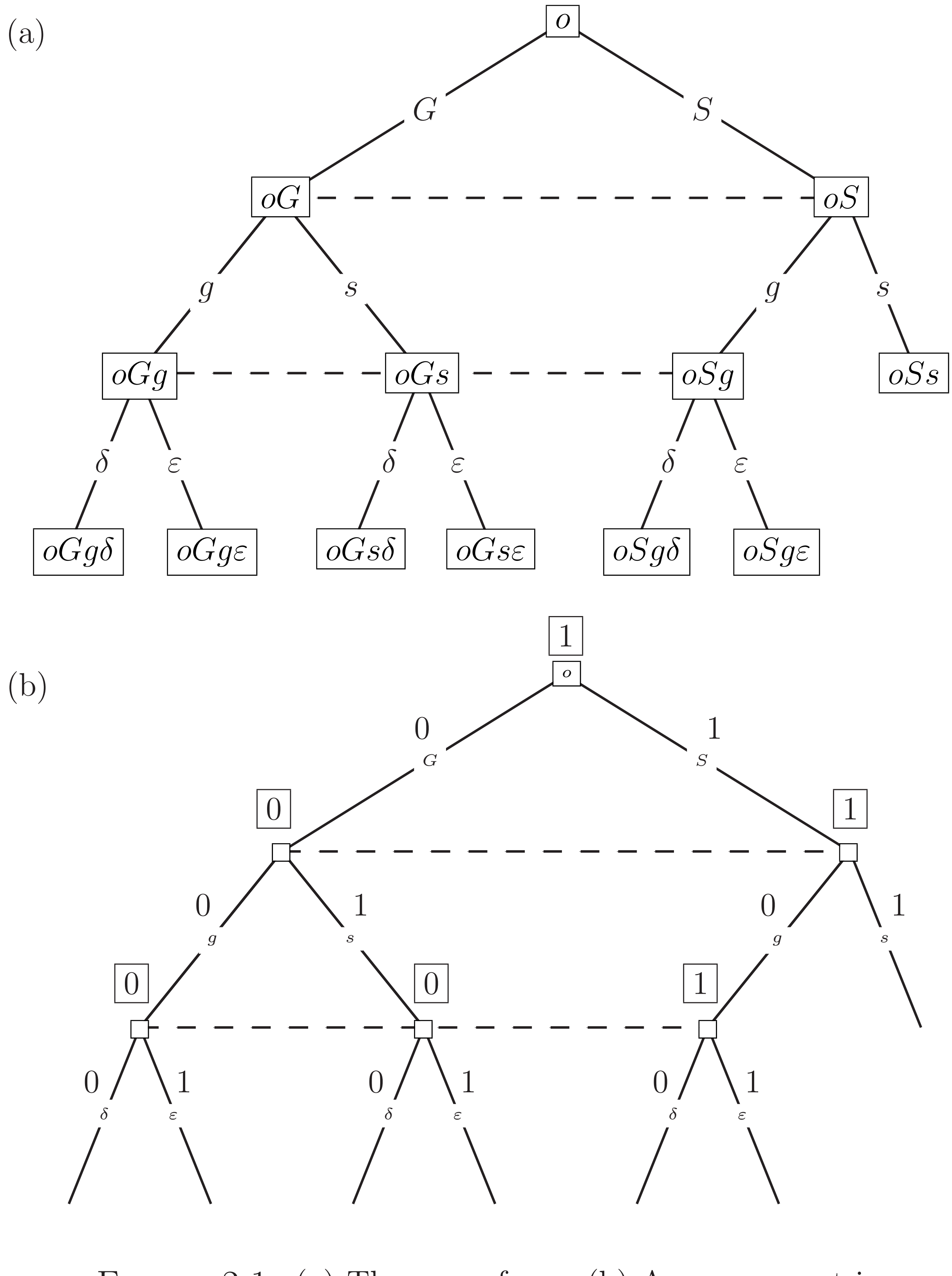

FIGURE 2.1. (a) The game form. (b) An assessment in which both girls sleep and the parent would blame the second girl for any noise (the probabilities on δ and ε are irrelevant for consistency).

(π, μ) for which

$$(1) \qquad (\forall x)\ \mu(x) = \frac{\Pi^x \rho \cup \pi}{\Sigma_{x' \in H(x)} \Pi^{x'} \rho \cup \pi}$$

(this equation is an application of Bayes Rule). Then, an assessment (π, μ) is said to be *consistent* if it is the limit of a sequence $\{(\pi_n, \mu_n)\}_n$ in Ψ^0.

For instance, consider the Figure 2.1(b)'s assessment (this happens to be the first assessment discussed verbally in Subsection 1.2). That assessment is consistent because it is the limit of the sequence of assessments defined in Figure 2.2(b), and because, at any n, Figure 2.2(b)'s assessment has full support and satisfies Bayes Rule (1).

2.2. Theorem Statement

As in Figure 2.2(b), it is routine to verify consistency by means of strategy sequences that have monomials in their numerators. The equivalence of (b) and (c) in the following theorem shows that the existence of such sequences is not only sufficient but also necessary for consistency. Thus more complicated sequences serve no essential role in the concept of consistency.

In addition, the theorem's equivalence between (a) and (c) shows that consistency can be characterized by means of monomials without reference to the strategy sequences defined in (b).

For example, consider Figure 2.2(a), which applies (a) to show the consistency of Figure 2.1(b)'s assessment. The figure's unboxed monomials are the monomials assigned to actions. They appear in (a)'s first equation, which relates the monomials to strategies. Meanwhile, every boxed monomial is the product of the initial probability and unboxed monomials above it. They are used by (a)'s second equation to relate the monomials to beliefs. In particular, the coefficient in the product at any node x is the pathwise product $\Pi^x \rho \cup c$ that appears in (a)'s second equation. Further, the exponent in the product at x is the pathwise sum $\Sigma^x e$ that is used to calculate H^e.

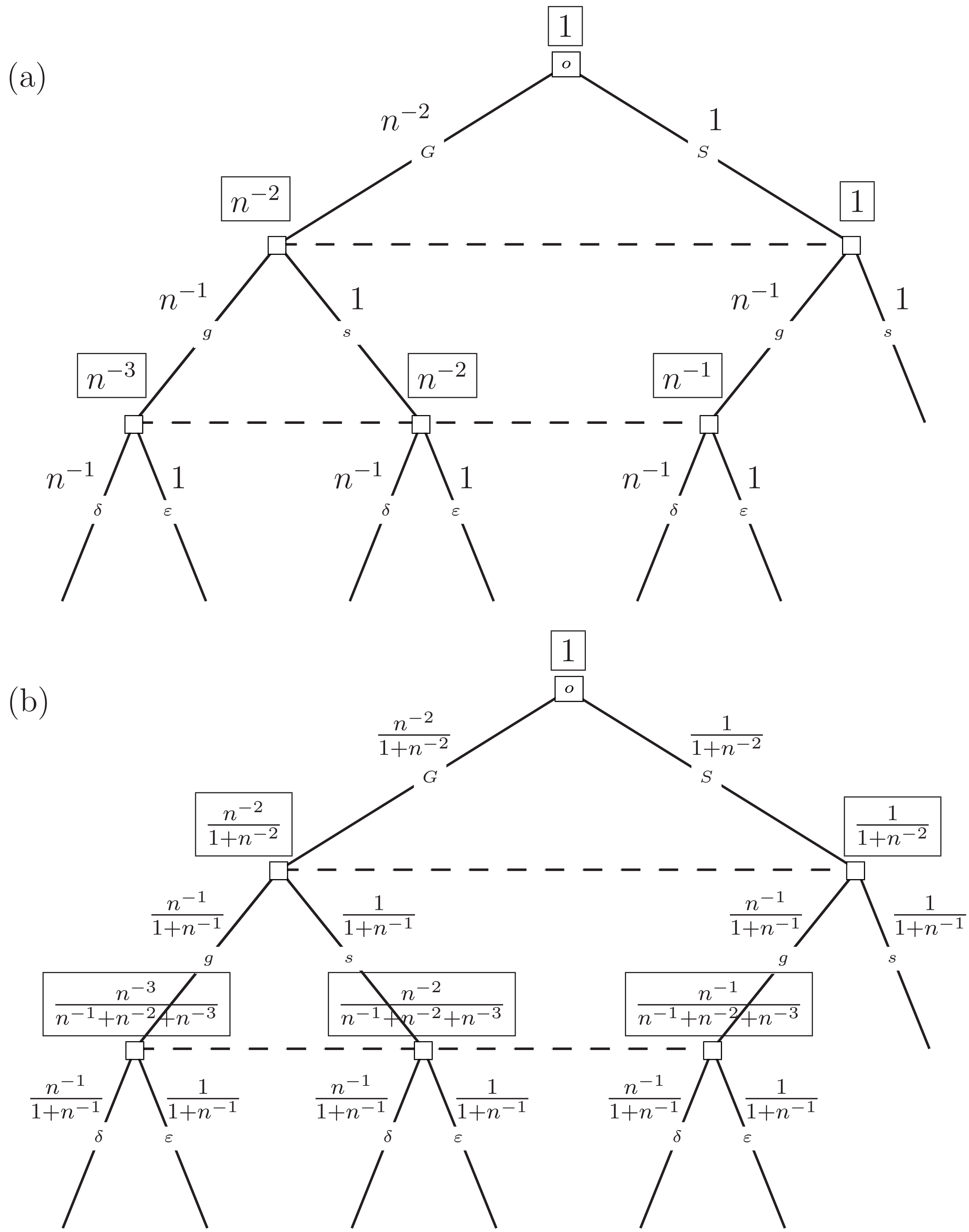

FIGURE 2.2. Characterizations (a) and (b) showing the consistency of Figure 2.1(b)'s assessment.

To be formally explicit, such products of monomials are merely a convenient way to calculate pathwise products of the form

$$\Pi^x \rho{\cup}c = \rho{\circ}p_{\ell(x)}(x) \times \Pi_{k=0}^{\ell(x)-1} c{\circ}\alpha{\circ}p_k(x)$$

and pathwise sums of the form

$$\Sigma^x e = \Sigma_{k=0}^{\ell(x)-1} e{\circ}\alpha{\circ}p_k(x)\ .$$

(Interested readers may wish to see how Appendix B defines arbitrary pathwise products and sums the form $\Pi^x f{\cup}g$ and $\Sigma^x g$.)

Theorem 2.1. *The following are equivalent for any assessment* (μ,π) *in any game form with initial probabilities. (a) There exist* $c{:}A{\to}(0,1]$ *and* $e{:}A{\to}\{...\text{-}2,\text{-}1,0\}$ *such that*

$$(\forall a)\ \pi(a) = \begin{pmatrix} c(a) & \text{if } e(a)=0 \\ 0 & \text{if } e(a)<0 \end{pmatrix} \text{ and}$$

$$(\forall x)\ \mu(x) = \begin{pmatrix} \dfrac{\Pi^x\rho{\cup}c}{\Sigma_{x'\in H^e(x)}\Pi^{x'}\rho{\cup}c} & \text{if } x\in H^e(x) \\ 0 & \text{if } x\notin H^e(x) \end{pmatrix}$$

where $H^e(x) = \text{argmax}\{\,\Sigma^{x'}e \,|\, x'{\in}H(x)\,\}$. *(b) There exist* $c{:}A{\to}(0,1]$ *and* $e{:}A{\to}\{...\text{-}2,\text{-}1,0\}$ *such that* (π,μ) *is the limit of the sequence* $\{(\pi_n,\mu_n)\}_n$ *defined by*

$$(\forall a)\ \pi_n(a) = \frac{c(a)n^{e(a)}}{\Sigma_{a'\in A\circ A^{-1}(a)}\, c(a')n^{e(a')}} \text{ and}$$

$$(\forall x)\ \mu_n(x) = \frac{\Pi^x\rho{\cup}\pi_n}{\Sigma_{x'\in H(x)}\Pi^{x'}\rho{\cup}\pi_n}\ .$$

(c) (π,μ) *is consistent.*

Proof. (b⇒c) This is routine (and requires only that c is positive-valued and e is real-valued). In particular, take any c and e, define the sequence $\{(\pi_n,\mu_n)\}_n$ according to the equations in (b), and assume that (π,μ) is its limit. Then every element (π_n,μ_n) in the sequence is in Ψ^0 because it has full support and because the equation defining μ_n

coincides with Bayes Rule (1). Hence (π, μ) satisfies the definition of consistency.

(c⇒a) See Section 3. (Subsection 3.1's derivation of exponents is the heart of the paper. Subsection 3.2's derivation of coefficients is less interesting.)

(a⇒b) See Appendix C. (This argument is tedious. Intuitively, (b)'s limits isolate the coefficients that correspond to the largest exponents, while (a) does the same thing directly.) □

2.3. The Twins Revisited

Although Section 4's discussion of the `KW` lemmas provides further motivation for the theorem, the story of the twins falling asleep can be used to quickly illustrate the practical importance of Theorem 2.1(c⇒a), which is the difficult part of the theorem.

Subsection 1.2 mentioned that consistency ruled out the assessment in which the parent would believe that both girls were giggling simultaneously. This cannot be deduced easily from the definition of consistency. It does, however, flow easily from Theorem 2.1(c⇒a), which states that any consistent assessment admits monomials which satisfy the two equations in (a). The first of the two equations requires that $e(G) < 0$ and $e(S) = 0$. Thus

$$\Sigma^{oGg} e = e(G) + e(g) < e(S) + e(g) = \Sigma^{oSg} e \ ,$$

thus $oGg \notin H^e(oGg)$, and thus $\mu(oGg) = 0$ by the second of the two equations.

2.4. Monomial Algebra

Figure 2.3 provides a second example. It differs from the first example in that the second girl now moves after having observed her sister's behaviour. Although it has fewer decision nodes, this second example is "less rectangular" in the sense that one of the parent's decision nodes follows one move while the other node follows two moves. Figure 2.3(a) uses characterization (a) to show the consistency of the assessment in

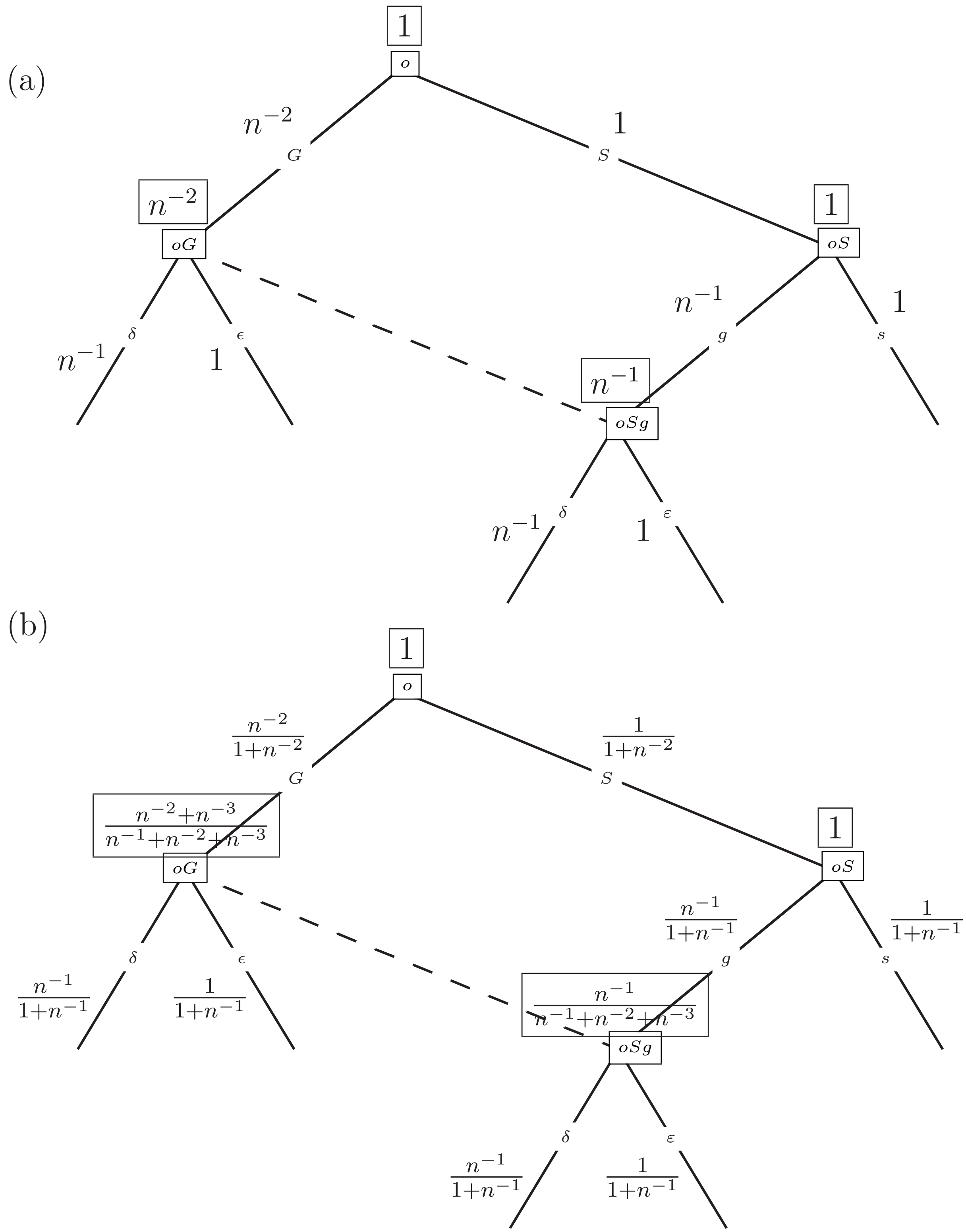

FIGURE 2.3. Characterizations (a) and (b) in a "less rectangular" game form.

which the parent would blame the second girl and both girls sleep. Figure 2.3(b) uses characterization (b) to accomplish the same thing. The next two paragraphs will use this example to extol the computational advantages of characterization (a).

Characterization (a) requires that one calculate the pathwise product $\Pi^x\rho\cup c$ and the pathwise sum $\Sigma^x e$ at every node x. This can be done recursively, just by multiplying monomials, as one moves down the tree. Figure 2.3(a) does this.

Meanwhile, (b) requires that one calculate the probability $\Pi^x\rho\cup\pi_n$ at each node x. Such probabilities typically vary with n and have ugly denominators. Then one must calculate $\mu_n(x)$ at each x, and that formula's denominator is a sum of the aforementioned probabilities, each with its own distinct and ugly denominator. The ensuing algebra is often tedious, and accordingly, the calculations behind Figure 2.3(b) have been mercifully omitted. Further, the algebra often results in second-order nuisance terms like the n^{-3} appearing in the numerator at node oG in Figure 2.3(b). Characterization (a) avoids all this tedium and complexity.

Finally, recall Subsection 1.2's suggestion that a monomial cn^e be interpreted as the positive probability c when e is zero and as a kind of zero probability when e is negative. It was further suggested that lesser, i.e. more negative, exponents specify lesser zero probabilities. Condition (a) bears this out: its equation concerning π is easy to interpret from this perspective, and its definition of H^e accords with the notion that lesser exponents are assigned to nodes that are infinitely less likely.

3. Proof of Theorem 2.1(c⇒a): Deriving Monomials

This section proves Theorem 2.1(c⇒a), that is, that any consistent assessment admits monomials which satisfy (a)'s equations. Accordingly, suppose that (π,μ) is a consistent assessment. In particular, suppose it is the limit of $\{(\pi_n,\mu_n)\}_n$, which is assumed to be a sequence of full-support assessments which satisfy Bayes Rule (1). Subsection 3.1 will derive exponents e, Subsection 3.2 will derive coefficients c, and Subsection 3.3 will show that c and e together satisfy (a)'s equations. Every number and limit in the argument (and in fact the entire paper) is finite.

3.1. Deriving Exponents e

This subsection's derivation of exponents is the heart of the paper. The reader might see Subsection 1.1 (third and fourth paragraphs) for an intuitive introduction to the argument; Subsection 4.2 (last paragraph) for how the argument corrects KW; Subsection 5.1 (last paragraph) for how the argument simplifies PJP; and Appendix A for how the argument mimics Scott's Theorem.

The Infinite-Relative-Probability Relation $\preceq$. The next five paragraphs construct (π,μ)'s infinite-relative-probability relation $\preceq$. The paragraphs will also show that the consistency of (π,μ) implies that $\preceq$ satisfies (6).

Let $x \overset{\mu}{\prec} y$ if x and y belong to the same information set, x is outside the support of μ, and y is inside the support of μ. Further, let $x \overset{\mu}{\approx} y$ if x and y belong to the same information set and both belong to the support of μ. For example, if (π,μ) were Figure 2.1(b)'s assessment, then $\overset{\mu}{\prec}$ is specified by $oG \overset{\mu}{\prec} oS$, $oGg \overset{\mu}{\prec} oSg$ and $oGs \overset{\mu}{\prec} oSg$, and $\overset{\mu}{\approx}$ is empty.

It is well-understood that if x and y are in the same information set and if y is in the support of μ, then

$$\begin{aligned}(2)\qquad & \lim_n \frac{\Pi^x \rho{\cup}\pi_n}{\Pi^y \rho{\cup}\pi_n} \\ &=_1 \lim_n \frac{\Pi^x \rho{\cup}\pi_n / \Sigma_{x' \in H(x)} \Pi^{x'} \rho{\cup}\pi_n}{\Pi^y \rho{\cup}\pi_n / \Sigma_{x' \in H(y)} \Pi^{x'} \rho{\cup}\pi_n} \\ &=_2 \lim_n \frac{\mu_n(x)}{\mu_n(y)}\end{aligned}$$

$$=_3 \frac{\mu(x)}{\mu(y)} \ ,$$

where $=_1$ follows from the assumption that $H(x) = H(y)$, $=_2$ follows from Bayes Rule (1), and $=_3$ follows from consistency and the assumption that $\mu(y) > 0$. Thus

$$(3) \qquad (\forall x \overset{\mu}{\prec} y) \ \lim_n \left(\Pi^x \rho{\cup}\pi_n / \Pi^y \rho{\cup}\pi_n \right) = 0 \text{ and}$$
$$(\forall x \overset{\mu}{\approx} y) \ \lim_n \left(\Pi^x \rho{\cup}\pi_n / \Pi^y \rho{\cup}\pi_n \right) \in (0,\infty) \ .$$

Let $t \overset{\pi}{\prec} u$ if t is an immediate successor of u and the action leading to t from u is played by π with zero probability (the notation (x,y) has been replaced with (t,u) because the (t,u) here need not be a pair of decision nodes). Further, let $t \overset{\pi}{\approx} u$ if t is an immediate successor of u and the action leading to t from u is played by π with positive probability, or symmetrically, if u is an immediate successor of t and the action leading to u from t is played with positive probability. For example, if (π,μ) were Figure 2.1(b)'s assessment, then $\overset{\pi}{\prec}$ is specified by $oG \overset{\pi}{\prec} o$, $oGg \overset{\pi}{\prec} oG$, $oSg \overset{\pi}{\prec} oS$, $oGg\delta \overset{\pi}{\prec} oGg$, $oGs\delta \overset{\pi}{\prec} oGs$, $oSg\delta \overset{\pi}{\prec} oSg$, and $\overset{\pi}{\approx}$ is specified by $oS \overset{\pi}{\approx} o$, $oGs \overset{\pi}{\approx} oG$, $oSs \overset{\pi}{\approx} oS$, $oGg\varepsilon \overset{\pi}{\approx} oGg$, $oGs\varepsilon \overset{\pi}{\approx} oGs$, $oSg\varepsilon \overset{\pi}{\approx} oSg$ together with the converses $o \overset{\pi}{\approx} oS$, $oG \overset{\pi}{\approx} oGs$, $oS \overset{\pi}{\approx} oSs$, $oGg \overset{\pi}{\approx} oGg\varepsilon$, $oGs \overset{\pi}{\approx} oGs\varepsilon$, $oSg \overset{\pi}{\approx} oSg\varepsilon$.

If t immediately succeeds u,

$$(4) \qquad \Pi^t \rho{\cup}\pi_n$$
$$= \rho{\circ}p_{\ell(t)}(t) \times \Sigma_{k=0}^{\ell(t)-1} \pi_n{\circ}\alpha{\circ}p_k(t)$$
$$= \rho{\circ}p_{\ell(t)}(t) \times \Sigma_{k=1}^{\ell(t)-1} \pi_n{\circ}\alpha{\circ}p_k(t) \times \pi_n{\circ}\alpha(t)$$
$$= \rho{\circ}p_{\ell(u)}(u) \times \Sigma_{k=0}^{\ell(u)-1} \pi_n{\circ}\alpha{\circ}p_k(u) \times \pi_n{\circ}\alpha(t)$$
$$= \Pi^u \rho{\cup}\pi_n \times \pi_n{\circ}\alpha(t) \ .$$

Thus if t immediately succeeds u, consistency yields

$$\lim_n \left(\Pi^t \rho{\cup}\pi_n / \Pi^u \rho{\cup}\pi_n \right) = \lim_n \pi_n{\circ}\alpha(t) = \pi{\circ}\alpha(t) \ .$$

Hence the definitions of $\overset{\pi}{\prec}$ and $\overset{\pi}{\approx}$ yield

$$(\forall t \overset{\pi}{\prec} u) \lim_n \left(\Pi^t \rho{\cup}\pi_n / \Pi^u \rho{\cup}\pi_n\right) = 0 \text{ and}$$
$$(\forall t \overset{\pi}{\approx} u) \lim_n \left(\Pi^t \rho{\cup}\pi_n / \Pi^u \rho{\cup}\pi_n\right) \in (0,\infty) \tag{5}$$

(the latter conclusion concerns not only the case where t immediately succeeds u and the limit of the ratio is $\pi{\circ}\alpha(t)$, but also the case where u immediately succeeds t and the limit of the ratio is $1/\pi{\circ}\alpha(u)$).

Define $\prec$ to be the union of $\overset{\mu}{\prec}$ and $\overset{\pi}{\prec}$, and $\approx$ to be the union of $\overset{\mu}{\approx}$ and $\overset{\pi}{\approx}$. By (3) and (5),

$$(\forall t{\prec}u) \lim_n \left(\Pi^t \rho{\cup}\pi_n / \Pi^u \rho{\cup}\pi_n\right) = 0 \text{ and}$$
$$(\forall t{\approx}u) \lim_n \left(\Pi^t \rho{\cup}\pi_n / \Pi^u \rho{\cup}\pi_n\right) \in (0,\infty) \;. \tag{6}$$

Finally, let $\preceq$ be the union of $\prec$ and $\approx$.

An Additive Representation for $\preceq$. In accord with Subsection 1.1 and Appendix A, the next four paragraphs will use Farkas' Lemma to find an additive representation for this infinite-relative-probability relation $\preceq$. It turns out that (6) is all that we need.

For our purposes here, let a *cancelling set* be an indexed set $\{(t_j, u_j)\}_{j=1}^m$ of pairs which obeys

$$\Sigma_{j=1}^m 1^{t_j} = \Sigma_{j=1}^m 1^{u_j} \;, \tag{7}$$

where for any node t the row vector $1^t \in \{0,1\}^A$ is defined by

$$(1^t)_a = \begin{pmatrix} 1 & \text{if } (\exists k{\in}\{0,1,...\,\ell(t){-}1\}) \; a = \alpha{\circ}p_k(t) \\ 0 & \text{otherwise} \end{pmatrix} \;. \tag{8}$$

For example,

$$(t_1, u_1) = (oG, oGs)$$
$$(t_2, u_2) = (oSs, oS)$$
$$(t_3, u_3) = (oSs, oS)$$
$$(t_4, u_4) = (oG, oGs)$$

is a cancelling set of pairs: every action leading to the left-hand node of a pair can be "cancelled" by an action leading to the right-hand node of a (possibly different) pair. (This example is like Appendix A's (31).)

Equation (6) yields that there cannot be a pair from $\prec$ in any cancelling set of pairs taken from $\preceq$. To see this, take any cancelling set $\{(t_j,u_j)\}_{j=1}^m$ of pairs from $\preceq$. Equation (7) yields that

$$(\forall n)\ \Pi_{j=1}^m \Pi^{t_j}\pi_n = \Pi_{j=1}^m \Pi^{u_j}\pi_n$$

(where the pathwise product $\Pi^t\pi_n$ is defined by $\Pi_{k=0}^{\ell(t)-1}\pi_n{\circ}\alpha{\circ}p_k(t)$ as in Appendix B). This is equivalent to

$$(\forall n)\ \Pi_{j=1}^m \left(\Pi^{t_j}\pi_n/\Pi^{u_j}\pi_n\right) = 1\ ,$$

which obviously yields

$$\lim_n\ \Pi_{j=1}^m \left(\Pi^{t_j}\pi_n/\Pi^{u_j}\pi_n\right) = 1\ ,$$

which yields

$$\lim_n\ \Pi_{j=1}^m \left(\Pi^{t_j}\rho{\cup}\pi_n/\Pi^{u_j}\rho{\cup}\pi_n\right) \in (0,\infty) \tag{9}$$

because the initial probabilities ρ are positive (and invariant with respect to n). Equations (6) and (9) contradict one another if there is a pair (t^j,u^j) from $\prec$.

We now translate to linear algebra: The result of the last paragraph yields that there cannot be column vectors $\beta \in \mathbb{Z}_+^{|\prec|}{\sim}\{0\}$ and $\delta \in \mathbb{Z}^{|\approx|}$ such that $\beta^T B + \delta^T D = 0$, where B and D are the matrices

$$B = [1^t{-}1^u]_{t\prec u} \text{ and}$$
$$D = [1^t{-}1^u]_{t\approx u}$$

whose rows are indexed by the pairs of the binary relations $\prec$ and $\approx$. To see this, suppose that there were such β and δ. By the symmetry of $\approx$, we may define $\hat{\delta} \in \mathbb{Z}_+^{|\approx|}$ by

$$(\forall t{\approx}u)\ \hat{\delta}_{(t,u)} = \begin{pmatrix} \delta_{(t,u)}{-}\delta_{(u,t)} & \text{if } \delta_{(t,u)}{-}\delta_{(u,t)} \geq 0 \\ 0 & \text{otherwise} \end{pmatrix}$$

so that $\delta^T D = \hat{\delta}^T D$, and so that consequently, we have $\beta \in \mathbb{Z}_+^{|\prec|} \sim \{0\}$ and $\hat{\delta} \in \mathbb{Z}_+^{|\approx|}$ such that $\beta^T B + \hat{\delta}^T D = 0$. Now define an indexed set $\{(t_j, u_j)\}_{j=1}^m$ in which every pair (t, u) from $\prec$ appears $\beta_{(t,u)}$ times and every pair (t, u) from $\approx$ appears $\hat{\delta}_{(t,u)}$ times. The equality $\beta^T B + \hat{\delta}^T D = 0$ yields that $\{(t_j, u_j)\}_{j=1}^m$ is a cancelling set, and $\beta \neq 0$ yields that it contains at least one pair from $\prec$. By the last paragraph, this is impossible.

Since the result of the previous paragraph is equivalent to (30), Farkas' Lemma (Fact A.1) now shows that there is a vector $e \in \mathbb{Z}^{|A|}$ such that $Be \ll 0$ and $De = 0$. By the definitions of B and D, this is equivalent to the existence of a function $e{:}A \to \mathbb{Z}$ such that

(10a) $$(\forall t \prec u)\ \Sigma^t e < \Sigma^u e \text{ and}$$

(10b) $$(\forall t \approx u)\ \Sigma^t e = \Sigma^u e\ .$$

Thus e provides an additive representation for (π, μ)'s infinite-relative-probability relation $\preceq$.[1]

Using Additive Representation. Thanks to (10), we now have that e additively represents the infinite-relative-probability relation $\preceq$ which was derived from the assessment (π, μ) at the beginning of this subsection. This implies that e and (π, μ) together satisfy a number of properties. One such property will be derived in each of the next two paragraphs. Later, Subsection 3.3 will use these two facts to show that e defines exponents in the sense of Theorem 2.1(a).

[1]This paper's use of the term "additive representation" departs from the economics literature in two small ways. First, the relation $\preceq$ is defined over the set T of nodes, which is not a Cartesian product (this appears to be a small innovation). Second, an economist (as opposed to a psychologist) would usually say that $\Sigma^{(\cdot)} e$ represents an *extension* of $\preceq$, where that extension $\preceq^*$ would be defined by $(\forall t, u)\ t \preceq^* u$ iff $\Sigma^t e \leq \Sigma^u e$. If one mulls this over, one notices that $\preceq$ is typically neither complete nor transitive: in Figure 2.1(b)'s assessment, both $oGg \preceq oGs$ and $oGg \succeq oGs$ fail (this violates completeness), and further, $oGg \preceq oGs$ fails in spite of the fact that both $oGg \prec oG$ and $oG \approx oGs$ hold (this violates transitivity).

Consider any decision node x. Let h be the information set that owns it. If $\mu(x)=0$, there must be some other node $y \in h$ such that $\mu(y)>0$, and thus the definition of $\prec$ and (10a) yield that $x \notin H^e(x)$ (where H^e is defined in Theorem 2.1). On the other hand, if $\mu(x)>0$, the definitions of $\prec$ and $\approx$ together with (10) yield that $x \in H^e(x)$. Hence,

$$(\forall x)\ \mu(x)>0 \text{ iff } x \in H^e(x)\ . \tag{11}$$

Finally consider any action a. Let t and u be such that a leads to t from u. By algebraic manipulation similar to (4), we have that $\Sigma^t e = \Sigma^u e + e(a)$. Thus (10) yields

$$t \prec u \text{ implies } e(a)<0 \text{ and} \tag{12a}$$

$$t \approx u \text{ implies } e(a)=0\ . \tag{12b}$$

If $\pi(a)=0$, the definition of $\prec$ and (12a) yield $e(a)<0$. If $\pi(a)>0$, the definition of $\approx$ and (12b) yield that $e(a)=0$. Hence,

$$(\forall a)\ e(a) \leq 0 \text{ and} \tag{13a}$$

$$(\forall a)\ \pi(a)>0 \text{ iff } e(a)=0\ . \tag{13b}$$

3.2. Deriving Coefficients c

This subsection is less interesting. It is independent from the previous subsection except that the following paragraph will recycle the notation $\overset{\mu}{\approx}$, the algebra of (2), and the notation 1^x.

As defined in Subsection 3.1's third paragraph, let $x \overset{\mu}{\approx} y$ mean that x and y are both in the support of their common information set. By the well-understood algebra of (2), consistency yields that

$$(\forall x \overset{\mu}{\approx} y)\ \lim_n \frac{\Pi^x \rho \cup \pi_n}{\Pi^y \rho \cup \pi_n} = \frac{\mu(x)}{\mu(y)}\ ,$$

which is equivalent to

$$(\forall x \overset{\mu}{\approx} y)\ \lim_n \frac{\Pi^x \pi_n}{\Pi^y \pi_n} = \frac{\mu(x)}{\mu(y)} \frac{\rho \circ p_{\ell(y)}(y)}{\rho \circ p_{\ell(x)}(x)}\ ,$$

which can be expressed very awkwardly as

$$(14) \qquad (\forall x \overset{\mu}{\approx} y) \ \lim_n (1^x - 1^y)[\ln(\pi_n(a))]_{a \in A} = \ln\left(\tfrac{\mu(x)}{\mu(y)} \tfrac{\rho \circ p_{\ell(y)}(y)}{\rho \circ p_{\ell(x)}(x)}\right) \ ,$$

where 1^x is the row vector defined earlier at (8) and $[\ln(\pi_n(a))]_{a \in A}$ is a column vector in $\mathbb{R}^A$. (Every probability in (14) is positive and every logarithm is finite.)

Further, let $A^+ = \{ \ a^+ \mid \pi(a^+) > 0 \ \}$. Consistency yields that

$$(\forall a^+ \in A^+) \ \lim_n \ \pi_n(a^+) = \pi(a^+) \ ,$$

which can be expressed very awkwardly as

$$(15) \qquad (\forall a^+ \in A^+) \ \lim_n \ (1^{a^+})[\ln(\pi_n(a))]_{a \in A} = \ln(\pi(a^+)) \ ,$$

where 1^{a^+} is the row vector in $\{0,1\}^{|A|}$ which assumes a value of 1 at a^+ and a value of 0 elsewhere. (Every probability in (15) is positive and every logarithm is finite.)

Equations (14) and (15) can be stacked together as the vector equation

$$\lim_n \begin{pmatrix} [1^x - 1^y]_{x \overset{\mu}{\approx} y} \\ \text{- - - - - - - -} \\ [1^{a^+}]_{a^+ \in A^+} \end{pmatrix} \Big[\ln(\pi_n(a))\Big]_{a \in A} = \begin{pmatrix} [\ln(\tfrac{\mu(x)}{\mu(y)} \tfrac{\rho \circ p_{\ell(y)}(y)}{\rho \circ p_{\ell(x)}(x)})]_{x \overset{\mu}{\approx} y} \\ \text{- - - - - - - - - - - - - - -} \\ [\ln(\pi(a^+))]_{a^+ \in A^+} \end{pmatrix} .$$

Since every $\pi_n(a)$ lies within $(0,1]$, this equation implies that the column vector

$$\hat{b} = \begin{pmatrix} [\ln(\tfrac{\mu(x)}{\mu(y)} \tfrac{\rho \circ p_{\ell(y)}(y)}{\rho \circ p_{\ell(x)}(x)})]_{x \overset{\mu}{\approx} y} \\ \text{- - - - - - - - - - - - - - -} \\ [\ln(\pi(a^+))]_{a^+ \in A^+} \end{pmatrix}$$

is in the closure of half-cone consisting of the nonpositive multiples of the columns of the matrix

$$\begin{pmatrix} [1^x - 1^y]_{x \overset{\mu}{\approx} y} \\ \text{- - - - - - - - -} \\ [1^{a^+}]_{a^+ \in A^+} \end{pmatrix} .$$

Thus, since the half-cone is closed, it contains $\hat{b}$. Hence there is some $c{:}A{\to}(0,1]$ such that

$$(16)\quad \begin{pmatrix} [1^x{-}1^y]_{x\overset{\mu}{\approx}y} \\ \text{- - - - - - - - -} \\ [1^{a^+}]_{a^+\in A^+} \end{pmatrix} \Big[\ln(c(a))\Big]_{a\in A} = \begin{pmatrix} [\ln(\frac{\mu(x)}{\mu(y)}\frac{\rho{\circ}p_{\ell(y)}(y)}{\rho{\circ}p_{\ell(x)}(x)})]_{x\overset{\mu}{\approx}y} \\ \text{- - - - - - - - - - - - - - -} \\ [\ln(\pi(a^+))]_{a^+\in A^+} \end{pmatrix} .$$

Because c is positive-valued (and every probability is positive), this vector equality (16) is equivalent to the combination of

$$(\forall x\overset{\mu}{\approx}y)\ \frac{\Pi^x\rho{\cup}c}{\Pi^y\rho{\cup}c} = \frac{\mu(x)}{\mu(y)} \text{ and}$$
$$(\forall a^+{\in}A^+)\ c(a^+) = \pi(a^+) .$$

Hence, the definition of $\overset{\mu}{\approx}$ yields

$$(17)\quad \begin{array}{c}(\forall h)(\forall\{x,y\}{\subseteq}h)\ \mu(x) > 0 \text{ and } \mu(y) > 0 \text{ implies} \\ \dfrac{\Pi^x\rho{\cup}c}{\Pi^y\rho{\cup}c} = \dfrac{\mu(x)}{\mu(y)} ,\end{array}$$

and the definition of A^+ yields

$$(18)\quad (\forall a)\ \pi(a) > 0 \text{ implies } c(a) = \pi(a) .$$

3.3. Conclusion (of Deriving Monomials)

The exponents e were taken to be integers (by the sentence containing (10)) and the coefficients c were taken from $(0,1]$ (by the sentence containing (16)). We can now derive the nonpositivity of e from (13a), the first equation in Theorem 2.1(a) from (13b) and (18), and the second equation in Theorem 2.1(a) from (11) and (17).

4. Application to KW

4.1. Overview

`KW` contains three fundamental theorems in addition to its path-breaking definition of sequential equilibrium.

Two of the theorems concern the set of sequential-equilibrium outcomes (that is, the projection of the set of sequential-equilibrium assessments on the terminal nodes). KW Theorem 2 shows that this set is finite, and KW Theorem 3 shows that it coincides with the set of perfect-equilibrium outcomes (both results assume generic payoffs).

These two theorems are derived from KW Theorem 1, which shows how the set of sequential-equilibrium assessments can be partitioned into a finite collection of tractable subsets. That theorem is in turn based on KW Lemma 2, which shows that the set of consistent assessments can be partitioned into a finite collection of manifolds. And finally, that lemma is based on Lemmas A1 and A2 in the KW Appendix.

Although it appears otherwise, KW Lemma A1 is close to an additive-representation theorem. In particular, the lemma's derivation of a "labelling" for every "consistent basis" is close to the statement that every consistent assessment admits an additive representation of its infinite-relative-probability relation (Subsection 1.1 notes that this is the crux of Theorem 2.1). However, the KW lemma's proof fails to establish the additivity of its representation. Subsection 4.2 explains this fallacy, and then Subsection 4.3 proves not only KW Lemma A1, but also KW Lemma A2, by means of Theorem 2.1.

4.2. A Fallacy in the Proof of KW Lemma A1

KW page 872 defines Ψ to be the set of consistent assessments, that is, the closure of the set Ψ^0 defined in Subsection 2.1. Page 880 then partitions Ψ into subsets of the form

$$\Psi_b = \{\ (\pi,\mu) \in \Psi \mid (\forall a)\ \pi(a){>}0 \text{ iff } a{\in}b \text{ and } (\forall x)\ \mu(x){>}0 \text{ iff } x{\in}b\ \}\ , \tag{19}$$

where b is a subset of $A{\cup}X$. KW page 872 also defines a *basis* b to be any subset of $A{\cup}X$, and defines a *consistent* basis to be a basis b for which Ψ_b is nonempty.

As on KW page 887, say that a basis can be *labelled* if there is a nonnegative-integer-valued function $K{:}A{\rightarrow}\mathbb{Z}_+$ satisfying

$$(\forall h)(\exists a{\in}A(h))\ K(a) = 0 \tag{20a}$$

$$(\forall a)\ a \in b \text{ iff } K(a) = 0 \tag{20b}$$

$$(\forall x)\ x \in b \text{ iff } x \in \text{argmin}\{J_K(x')|x'{\in}H(x)\}\ , \tag{20c}$$

where $J_K{:}X{\rightarrow}\mathbb{Z}_+$ is defined by

$$J_K(x) = \Sigma_{k=0}^{\ell(x)-1} K{\circ}\alpha{\circ}p_k(x)\ . \tag{21}$$

KW Lemma A1 observes that a basis is consistent iff it can be labelled. Yet its proof on page 887 is fallacious. In particular, the proof's second paragraph seeks to establish that any consistent basis can be labelled. It takes an arbitrary consistent basis, derives a binary relation $\dot{<}$ over the set of nodes, derives a function J which represents $\dot{<}$, and then derives a function K over the set of actions. The last line on page 887 claims but does not demonstrate that $J = J_K$. In fact, Streufert (2006b, Subsection 3.2) shows by counterexample that this equation does not follow from their construction.

This matter can be understood intuitively from the vantage point of additive representation that was introduced in Subsection 1.1. The KW ordering $\dot{>}$ is an extension of the infinite-relative-probability relation $\prec$ that was discussed informally in Subsection 1.1 and defined formally at the start of Subsection 3.1's proof. The offending equation $J = J_K$ would claim that the KW representation J is additive (note the definition of J_K at (21)).

4.3. Proofs of KW Lemmas A1 and A2

KW Lemmas A1 and A2 are weaker than Theorem 2.1(a⇔c). Proving this requires tedious notational gymnastics. To ease the task, we begin with three simple lemmas which will be used repeatedly. To get oriented, note that any two of (22), (24), and (25) imply the third.

LEMMA 4.1. *$(\pi,\mu) \in \Psi_b$ iff (π,μ) is consistent,*

$$(\forall a)\ \pi(a) > 0 \text{ iff } a \in b\ , \text{ and} \tag{22a}$$

$$(\forall x)\ \mu(x) > 0 \text{ iff } x \in b\ . \tag{22b}$$

Proof. This follows from the definition (19) of Ψ_b. □

LEMMA 4.2. *b can be labelled iff there exists $e{:}A{\rightarrow}\mathbb{Z}_-$ such that*

$$(\forall h)(\exists a{\in}A(h))\ e(a) = 0\ , \tag{23}$$

$$(\forall a)\ a \in b \text{ iff } e(a) = 0\ , \text{ and} \tag{24a}$$

$$(\forall x)\ x \in b \text{ iff } x \in H^e(x)\ . \tag{24b}$$

Proof. (20) is equivalent to the combination of (23) and (24) after J_K has been substituted out, Theorem 2.1's H^e has been substituted in, and K and $-e$ have been identified. □

LEMMA 4.3. *If (π,μ) and (c,e) satisfy Theorem 2.1(a)'s equations, then*

$$(\forall a)\ \pi(a) > 0 \text{ iff } e(a) = 0 \text{ and} \tag{25a}$$

$$(\forall x)\ \mu(x) > 0 \text{ iff } x \in H^e(x)\ . \tag{25b}$$

Proof. Obvious. □

PROPOSITION 4.4 (KW Lemma A1). *A basis is consistent iff it can be labelled.*

Proof. Suppose b is consistent. This means $\Psi_b \neq \varnothing$, and thus Lemma 4.1 yields the existence of a consistent assessment (π,μ) satisfying (22). Because (π,μ) is consistent, Theorem 2.1(c⇒a) yields (c,e) satisfying Theorem 2.1(a)'s equations. By Lemma 4.3, these equations yield (25), which together with (22) yields (24). In addition, Theorem 2.1(a)'s first equation and the well-definition of π yield (23). Hence, by Lemma 4.2, b can be labelled. (This paragraph was the difficult part.)

Conversely, suppose that b can be labelled. Then by Lemma 4.2 there exists some e which satisfies (23) and (24). Define c by

$$c(a) = 1/|\{\ a'{\in}A{\circ}A^{-1}(a)\ |\ e(a'){=}0\ \}|\ .$$

Because of (23) and the normalization in the definition of c, we can construct π and μ to satisfy Theorem 2.1(a)'s equations. Then by Theorem 2.1(a⇒c), (π,μ) is consistent. Further, by Lemma 4.3, Theorem 2.1(a)'s equations yield (25), which together with (24) yields (22). Hence, by Lemma 4.1, $(\pi,\mu) \in \Psi_b$. Hence b is consistent. □

As on KW page 888, define

$$\Xi_b = \{\ c{:}A{\to}(0,\infty)\ |\ (\forall h)\ \Sigma_{a\in b\cap A(h)}c(a) = 1\ \}\ ,$$

let π^b map $c \in \Xi_b$ to

$$\pi^b(c)(a) = \begin{pmatrix} c(a) & \text{if } a \in b \\ 0 & \text{if } a \notin b \end{pmatrix}\ , \tag{26a}$$

and let μ^b map $c \in \Xi_b$ to

$$\mu^b(c)(x) = \begin{pmatrix} \dfrac{\Pi^x\rho{\cup}c}{\Sigma_{x'\in b\cap H(x)}\Pi^{x'}\rho{\cup}c} & \text{if } x \in b \\ 0 & \text{if } x \notin b \end{pmatrix} \tag{26b}$$

(the KW symbol ξ has been replaced by c, the KW multinomials m^x have been substituted out, and the restriction $(\forall w)\ \rho(w) = 1/|W|$ arbitrarily imposed at the start of KW Subection A.1 has been relaxed).

Proposition 4.5 (KW Lemma A2). *For any consistent b, Ψ_b is the image of Ξ_b under the mapping (π^b,μ^b).*

Proof. Take any assessment (π,μ) in Ψ_b. By Lemma 4.1, we have (22). Further, since (π,μ) is consistent, Theorem 2.1(c⇒a) yields the existence of (c,e) which satisfy Theorem 2.1(a)'s equations (Theorem 2.1(c⇒a)'s result that c is bounded by one is not needed). By Lemma 4.3, these equations yield (25), which together with (22) yields (24). We now assemble three facts. [a] $c \in \Xi_b$ by Theorem 2.1(a)'s first equation, the well-definition of π, and (24a). [b] $\pi = \pi^b(c)$ by Theorem 2.1(a)'s first equation, definition (26a), and (24a). [c] $\mu = \mu^b(c)$ by Theorem 2.1(a)'s second equation, definition (26b), (24b), and by the fact that $(\forall x)\ b \cap H(x) = H^e(x)$ by (24b). By these three facts, (π, μ) is in the image of Ξ_b under (π^b, μ^b).

Conversely, take any consistent b and any $c \in \Xi_b$. By Proposition 4.4 and Lemma 4.2, there is some e which satisfies (24). Then, $\pi^b(c)$ satisfies Theorem 2.1(a)'s first equation by definition (26a) and (24a). Further, (24b) yields $(\forall x)\ b \cap H(x) = H^e(x)$, and thus, $\mu^b(c)$ satisfies Theorem 2.1(a)'s second equation by definition (26b) and (24b). Therefore, since $(\pi^b(c), \mu^b(c))$ satisfies Theorem 2.1(a)'s equations, Theorem 2.1(a⇒c) yields that $(\pi^b(c), \mu^b(c))$ is consistent [tiny detail: the c here has not been bounded by one, but any positive c can be accommodated by the first sentence in the proof of b⇒c (Subsection 2.2) and the first paragraph in the proof of a⇒b (Appendix C)]. Therefore, since definition (26) yields that $(\pi^b(c), \mu^b(c))$ satisfies (22), Lemma 4.1 yields $(\pi^b(c), \mu^b(c)) \in \Psi_b$. □

5. Literature

5.1. KW and PJP

Theorem 2.1's precedents are KW Lemmas A1 and A2 and PJP Theorem 3.1 (PJP abbreviates Perea y Monsuwe, Jansen, and Peters (1997), whose proofs also appear in Perea (2001, pages 76-81)).

Statements. First, let us consider the statements of the results (independently of their proofs). The following two paragraphs describe how Theorem 2.1 is slightly stronger than translated versions of the results in KW and PJP. Thus, with regard to theorem statements only, Theorem 2.1 essentially contributes its convenient formulation of monomials.

KW Lemma A1's assignment of labels K is equivalent to the negative of an assignment of exponents e. Further, KW Lemma A2's assignment of numbers ξ is comparable to an assignment of coefficients c: the only distinction is that this paper's coefficients are bounded by one and no such restriction is imposed in KW. Although these analogies are easily seen, it is difficult to translate formally between Theorem 2.1 and the

KW lemmas: for example, Subsection 4.3 takes almost three pages just to show that Theorem 2.1(a⇔c) implies the KW lemmas. In the end, Theorem 2.1's characterization of consistency is sharper to the extent that it derives coefficients that are not bigger than one.

PJP's completely mixed pseudo behaviour strategy profile $\hat{\sigma}$ is equivalent to an assignment of coefficients c. Further, the logarithms of the PJP mistake probabilities ε are comparable to exponents e: the main distinction is that exponents e are required to be integers as opposed to reals (by the way, both e and ε increase with the chance of *not* making a mistake). Given these translations, PJP Theorem 3.1 corresponds to Theorem 2.1(a⇔c), and PJP Corollary 3.3 corresponds to Theorem 2.1(b⇔c). In the end, Theorem 2.1's characterization is sharper to the extent that it derives integer rather than real exponents.

Proofs. The fundamental insight of this paper is the fresh perspective of seeking an additive representation for an assessment's infinite-relative-probability relation (Subsection 1.1).

That fresh perspective reveals a nontrivial fallacy in the KW proofs (Subsection 4.2) which can be repaired with the help of Theorem 2.1 (Subsection 4.3). Although the KW proof was flawed, it was intuitive in the sense that it studied the infinite-relative-probability relation that is induced by a consistent assessment.

The new perspective also provides an alternative to the PJP proof. This alternative (see Subsection 3.1) is more intuitive in the sense that it follows the KW proof in constructing an infinite-relative-probability relation and then mimics Scott's Theorem to find an additive representation. Further, it is more economical in its use of mathematics because it employs Farkas' Lemma from linear algebra rather than the Separating Hyperplane Theorem from analysis. In hindsight, this simplification rings true because the Separating Hyperplane Theorem can be usefully regarded as an analytic generalization of Farkas' Lemma (e.g., Ziegler (1995, page 40)).

5.2. Other Papers

Theorem 2.1 also bears some resemblance to Theorem 2.4 of Govindan and Klumpp (2002), which uses polynomials to characterize perfection. The proofs of the two results are quite different: the proof here uses only linear algebra while the proof there uses algebraic topology. A formal link between the results themselves has yet to be established.

Finally, Streufert (2006a) complements Theorem 2.1's monomial characterization with a second characterization of consistency that is formulated in terms of a new concept of producthood for relative probability. Both that paper and this one owe a large hidden debt to Kohlberg and Reny (1997).

Appendix A. Scott's Theorem

In social choice theory, Suzumura (1976, Theorem 3) showed that a binary relation $\preceq$ over a finite set Z can be extended to an ordering iff for any $\{z^i\}_{i=1}^k$

$$\begin{pmatrix} z^1 & \preceq & z^2 \\ z^2 & \preceq & z^3 \\ & \cdots & \\ z^{k-1} & \preceq & z^k \\ z^k & \preceq & z^1 \end{pmatrix} \text{ implies } \begin{pmatrix} z^1 & \approx & z^2 \\ z^2 & \approx & z^3 \\ & \cdots & \\ z^{k-1} & \approx & z^k \\ z^k & \approx & z^1 \end{pmatrix} . \tag{27}$$

Notice that the bracketted set of ordered pairs exhibits an elementary sort of cancelling: every left-hand term is cancelled by a right-hand term.

Analogously, Scott (1964, Theorem 3.1) showed that a binary relation $\preceq$ over a finite Cartesian product like $\{A,B,C\}\times\{a,b,c\}$ can be extended to an ordering that has an additive representation iff it satisfies many "cancellation laws" like

$$\begin{pmatrix} Aa \preceq Ab \\ Bb \preceq Ba \end{pmatrix} \text{ implies } \begin{pmatrix} Aa \approx Ab \\ Bb \approx Ba \end{pmatrix} . \tag{28}$$

Notice that (28) differs from (27) in that the dimensions are cancelled separately. For example, every A in the first dimension of a left-hand term can be matched with an A in the first dimension of a (possibly different) right-hand term. (Scott's Theorem is well-known in mathematical psychology like Debreu (1960) and Gorman (1968) are well-known in mathematical economics. It is relatively unfamiliar to economists because it concerns relations with discrete rather than continuous domains.)

Scott's Theorem is derived from linear algebra. In particular, Krantz, Luce, Suppes, and Tversky (1971, Subsection 9.2) shows that it and many similar results can be derived from Fact A.1, which appears in their book as Theorem 2.7 (include the top of their page 63 and replace their $[\alpha_i]_{i=1}^{m'}$ with $-B$ and their $[\beta_i]_{i=1}^{m''}$ with D). This result is a version of Farkas' Lemma, which goes by a half-dozen other names including the Theorem of the Alternative, and the Duality Theorem for linear programming (see Ziegler (1995, pages 39–40)). Here $\mathbb{Q}$ denotes the set of rationals, $\mathbb{Z}$ denotes the set of integers, and $Bx \ll 0$ means that every element of the vector Bx is negative.

Fact A.1 (Farkas' Lemma for Rational Matrices). *For any matrices $B \in \mathbb{Q}^{bk}$ and $D \in \mathbb{Q}^{dk}$, the following are equivalent.*

$$(\exists x{\in}\mathbb{Z}^k)\ Bx \ll 0 \text{ and } Dx = 0\ . \tag{29}$$

$$\text{Not } (\exists \beta{\in}\mathbb{Z}^b_+{\sim}\{0\})(\exists \delta{\in}\mathbb{Z}^d)\ \beta^T B + \delta^T D = 0\ . \tag{30}$$

This paper cannot apply Scott's Theorem directly because an infinite-relative-probability relation $\preceq$ is defined over the set of nodes, which is not a Cartesian product. Nonetheless, Scott's insight can be applied by mimicking his use of Farkas' Lemma.

Essentially, Farkas' Lemma can be used to show that a relation $\preceq$ over nodes has a representation that is additive over actions in the sense of (10) if it satisfies conditions like

$$\begin{pmatrix} oG \preceq oGs \\ oSs \preceq oS \end{pmatrix} \text{ implies } \begin{pmatrix} oG \approx oGs \\ oSs \approx oS \end{pmatrix} . \tag{31}$$

(this particular condition resembles (28) if one sets $A = oG$, $B = oS$, $b = s$, and drops a altogether). Thus the crux of Subsection 3.1's proof is to show that conditions like (31) are satisfied by a consistent assessment's infinite-relative-probability relation $\preceq$.

To be precise, the satisfaction of all conditions like (31) is specified within Subsection 3.1's proof by the statement that "every cancelling set of pairs from $\preceq$ contains no elements of $\prec$." The paragraph containing (9) derives this statement from equation (6), which in turn follows directly from the definition of consistency.

Appendix B. Pathwise Products and Sums

Let $f \in \mathbb{R}^W$ denote an arbitrary function from the set W of initial nodes, and let $g \in \mathbb{R}^A$ denote an arbitrary function from the set A of actions. Then let $\Pi{:}X \times \mathbb{R}^{W \cup A} \rightarrow \mathbb{R}$ be defined by

$$\Pi^x f \cup g = f \circ p_{\ell(x)}(x) \times \Pi_{k=0}^{\ell(x)-1} g \circ \alpha \circ p_k(x) ,$$

and let $\Sigma{:}X \times \mathbb{R}^{W \cup A} \rightarrow \mathbb{R}$ be defined by

$$\Sigma^x f \cup g = f \circ p_{\ell(x)}(x) + \Sigma_{k=0}^{\ell(x)-1} g \circ \alpha \circ p_k(x) .$$

For example, $\Pi^x \rho \cup \pi$ is the product of the initial probability and the strategies on the path leading to node x.

Further, let $\Pi^x g$ abbreviate $\Pi^x 1 \cup g$, and similarly, let $\Sigma^x g$ abbreviate $\Sigma^x 0 \cup g$. For example, $\Pi^x \pi$ is the product of the strategies on the path leading to node x, and similarly, $\Sigma^x e$ is the sum of the exponents assigned to the actions on the path leading to node x.

Appendix C. Proof of Theorem 2.1(a⇒b)

This appendix contains a tedious portion of Theorem 2.1's proof. It shows that something slightly stronger than Theorem 2.1(a⇒b) holds for any positive-valued c and any real-valued e. In particular, take any positive-valued c and any real-valued e, assume that (π, μ) satisfies the two equations in (a), and define the sequence $\{(\pi_n, \mu_n)\}_n$ according to the two equations in (b). The following will show that $\{(\pi_n, \mu_n)\}_n$ converges to (π, μ).

The convergence of $\{\pi_n\}_n$ *to* π. The first equation in (a) and the well-definition of π yield that

$$(\forall h) \lim_n \Sigma_{a \in A(h)} c(a) n^{e(a)} = 1 \ . \tag{32}$$

Then

$$\begin{aligned}(\forall a) \ \pi(a) &=_1 \lim_n c(a) n^{e(a)} \\ &=_2 \frac{\lim_n c(a) n^{e(a)}}{\lim_n \Sigma_{a' \in A \circ A^{-1}(a)} c(a') n^{e(a')}} \\ &=_3 \lim_n \frac{c(a) n^{e(a)}}{\Sigma_{a' \in A \circ A^{-1}(a)} c(a') n^{e(a')}} \\ &=_4 \lim_n \pi_n(a) \ ,\end{aligned}$$

where $=_1$ holds by the first equation in (a), $=_2$ holds by (32), $=_3$ holds because the numerator's limit is real by the nonpositivity of $e(a)$ and because the denominator's limit is nonzero by (32), and $=_4$ holds by the definition of π_n in (b).

The convergence of $\{\mu_n\}_n$ *to* μ. Fix a node x. We proceed in three steps. First, define $e^* = \max\{ \Sigma^{x'} e \mid x' {\in} H(x) \}$ and note that

$$(\forall x' {\in} H(x)) \lim_n n^{-e^*} \times \Pi^{x'} c n^e \in [0, \infty) \text{ and} \tag{33a}$$

$$(\exists x' {\in} H(x)) \lim_n n^{-e^*} \times \Pi^{x'} c n^e \in (0, \infty) \ . \tag{33b}$$

Second, note that

$$\begin{aligned}(34) \quad (\forall x' {\in} H(x)) \ &\lim_n n^{-e^*} \times \Pi^{x'} \rho {\cup} c n^e \\ =_1 \ &\lim_n n^{-e^*} \times \rho_{p_{\ell(x')}(x')} \times \Pi_{k=0}^{\ell(x')-1} c {\circ} \alpha {\circ} p_k(x') n^{e \circ \alpha \circ p_k(x')}\end{aligned}$$

$$
\begin{aligned}
&=_2 \ \frac{\lim_n \ n^{-e^*} \times \rho_{p_{\ell(x')}(x')} \times \Pi_{k=0}^{\ell(x')-1} c \circ \alpha \circ p_k(x') n^{e \circ \alpha \circ p_k(x')}}{\Pi_{k=0}^{\ell(x')-1} \ \lim_n \ \Sigma_{a \in A \circ A^{-1}(\alpha \circ p_k(x'))} c(a) n^{e(a)}} \\
&=_3 \ \lim_n \ \frac{n^{-e^*} \times \rho_{p_{\ell(x')}(x')} \times \Pi_{k=0}^{\ell(x')-1} c \circ \alpha \circ p_k(x') n^{e \circ \alpha \circ p_k(x')}}{\Pi_{k=0}^{\ell(x')-1} \ \Sigma_{a \in A \circ A^{-1}(\alpha \circ p_k(x'))} c(a) n^{e(a)}} \\
&=_4 \ \lim_n \ n^{-e^*} \times \rho_{p_{\ell(x')}(x')} \times \Pi_{k=0}^{\ell(x')-1} \pi_n \circ \alpha \circ p_k(x') \\
&=_5 \ \lim_n \ n^{-e^*} \times \Pi^{x'} \rho \cup \pi_n \ ,
\end{aligned}
$$

where $=_1$ holds by the definition of a pathwise product (Appendix B), $=_2$ holds because each limit in the denominator is one by (32), $=_3$ holds because each limit in the denominator is one and because the limit in the numerator is real by (33a), and $=_4$ holds because of the definition of π_n in (b), and $=_5$ holds by the definition of a pathwise product. Third,

$$
\begin{aligned}
\mu(x) \ &=_1 \ \lim_n \ \frac{\Pi^x \rho \cup c n^e}{\Sigma_{x' \in H(x)} \ \Pi^{x'} \rho \cup c n^e} \\
&=_2 \ \lim_n \ \frac{n^{-e^*} \times \Pi^x \rho \cup c n^e}{\Sigma_{x' \in H(x)} \ n^{-e^*} \times \Pi^{x'} \rho \cup c n^e} \\
&=_3 \ \frac{\lim_n \ n^{-e^*} \times \Pi^x \rho \cup c n^e}{\Sigma_{x' \in H(x)} \ \lim_n \ n^{-e^*} \times \Pi^{x'} \rho \cup c n^e} \\
&=_4 \ \frac{\lim_n \ n^{-e^*} \times \Pi^x \rho \cup \pi_n}{\Sigma_{x' \in H(x)} \ \lim_n \ n^{-e^*} \times \Pi^{x'} \rho \cup \pi_n} \\
&=_5 \ \lim_n \ \frac{n^{-e^*} \times \Pi^x \rho \cup \pi_n}{\Sigma_{x' \in H(x)} \ n^{-e^*} \times \Pi^{x'} \rho \cup \pi_n} \\
&=_6 \ \lim_n \ \frac{\Pi^x \rho \cup \pi_n}{\Sigma_{x' \in H(x)} \ \Pi^{x'} \rho \cup \pi_n} \\
&=_7 \ \lim_n \ \mu_n(x) \ .
\end{aligned}
$$

where $=_1$ holds by the second half of (a), $=_2$ holds by algebra, $=_3$ holds by (33), $=_4$ holds by (34), $=_5$ holds by (33) and (34), $=_6$ holds by algebra, and $=_7$ holds by the definition of $\mu_n(x)$ in (b). This entire equality states that $\{\mu_n(x)\}_n$ converges to $\mu(x)$. Since x was chosen arbitrarily at the start of the paragraph, it must be that $\{\mu_n\}_n$ converges to μ.

References

Debreu, G. (1960): "Topological Methods in Cardinal Utility Theory," in *Mathematical Methods in the Social Sciences, 1959*, ed. by K. J. Arrow, S. Karlin, and P. Suppes, pp. 16–26. Stanford University Press, reprinted in Debreu (1983).

——— (1983): *Mathematical Economics: Twenty Papers of Gerard Debreu*. Cambridge University Press.

Gorman, W. M. (1968): "The Structure of Utility Functions," *Review of Economic Studies*, 35, 376–390.

Govindan, S., and T. Klumpp (2002): "Perfect Equilibrium and Lexicographic Beliefs," *International Journal of Game Theory*, 31, 229–243.

Kohlberg, E., and P. J. Reny (1997): "Independence on Relative Probability Spaces and Consistent Assessments in Game Trees," *Journal of Economic Theory*, 75, 280–313.

Krantz, D. H., R. D. Luce, P. Suppes, and A. Tversky (1971): *Foundations of Measurement, Volume I: Additive and Polynomial Representations*. Academic Press.

Kreps, D. M., and G. Ramey (1987): "Structural Consistency, Consistency, and Sequential Rationality," *Econometrica*, 55, 1331–1348.

Kreps, D. M., and R. Wilson (1982): "Sequential Equilibria," *Econometrica*, 50, 863–894.

Perea, A. (2001): *Rationality in Extensive Form Games*. Kluwer.

Perea y Monsuwe, A., M. Jansen, and H. Peters (1997): "Characterization of Consistent Assessments in Extensive Form Games," *Games and Economic Behavior*, 21, 238–252.

Scott, D. (1964): "Measurement Structures and Linear Inequalities," *Journal of Mathematical Psychology*, 1, 233–247.

Streufert, P. A. (2006a): "Characterizing Consistency by Monomials and by Product Dispersions," University of Western Ontario, Economics Department Research Report 2006-02, 27 pages.

——— (2006b): "A Comment on 'Sequential Equilibria'," University of Western Ontario, Economics Department Research Report 2006-03, 10 pages.

Suzumura, K. (1976): "Remarks on the Theory of Collective Choice," *Economica*, 43, 381–390.

Ziegler, G. M. (1995): *Lectures on Polytopes*. Springer-Verlag.