Online Appendix:
"Distributional Preferences, Reciprocity-Like Behavior, and Efficiency in Bilateral Exchange"

by Daniel J. Benjamin

Appendix A: Neither Player's Distributional Preferences Are Monotonic

In this appendix, I discuss how the conclusions of the analysis are affected if both FM's and SM's distributional preferences are joint-monotonic but not necessarily monotonic.

Theorem 1 in the main text shows that if one player's distributional preferences are joint-monotonic and the other player's are monotonic, then the set of UPE material payoff pairs is a subset of the set of MPE material payoff pairs. Theorem A1 generalizes to the case where both players' distributional preferences are joint-monotonic. In that case, there are UPE material payoff pairs that are not MPE. To state the result, let the interpersonal indifference curve of FM that goes through SM's favorite transaction $\left(\overline{\overline{\pi}}_1, \overline{\overline{\pi}}_2\right)$ be denoted $\overline{\overline{IC}}_1$, and let the indifference curve of SM that goes through FM's favorite transaction $(\overline{\pi}_1, \overline{\pi}_2)$ be denoted $\overline{IC}_2$.

Theorem A1. *Suppose U_1 and U_2 are joint-monotonic and quasi-concave. FM's and SM's favorite transactions, $(\overline{a}_1, \overline{a}_2)$ and $\left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right)$, exist and are unique. The set of UPE material payoff pairs is a connected set that includes $(\overline{\pi}_1, \overline{\pi}_2)$ and $\left(\overline{\overline{\pi}}_1, \overline{\overline{\pi}}_2\right)$ and lies within the region enclosed by $\overline{\overline{IC}}_1$, $\overline{IC}_2$, and the MPE frontier.*

Figure A1 illustrates the relationship between the set of MPE material payoff pairs and the set of UPE material payoff pairs in the case where neither player has monotonic distributional preferences. The set of UPE material payoff pairs lies within the region enclosed by $\overline{\overline{IC}}_1$, $\overline{IC}_2$, and the MPE frontier because both players prefer any material payoff within that region to any feasible material payoff pair outside that region. A material payoff pair that is UPE either occurs at a tangency point between the players' indifference curves—at a point where both indifference curves are upward sloping (as shown in the figure)—or it occurs on the MPE frontier if the "relevant tangency" lies outside the set of feasible material-payoff pairs.

Theorem A1 specializes to Theorem 1 when at least one player has monotonic distributional preferences. In that case, graphically, there cannot be a tangency between the players' indifference curves because the indifference curves of the player with monotonic distributional preferences are everywhere downward sloping.

Dufwenberg, Heidhues, Kirchsteiger, Riedel, & Sobel (2011) independently prove a different result that is also more general than Theorem 1. They assume that the agents' distributional preferences satisfy a condition they call "social monotonicity," which can be defined as follows:

Definition A1. *U_1 and U_2 are **social-monotonic** if for any (π_1, π_2) and any $\varepsilon > 0$, there is some $(\widehat{\pi}_1, \widehat{\pi}_2)$ such that $0 < \widehat{\pi}_1 - \pi_1 < \varepsilon$, $0 < \widehat{\pi}_2 - \pi_2 < \varepsilon$, $U_1(\widehat{\pi}_1, \widehat{\pi}_2) > U_1(\pi_1, \pi_2)$, and $U_2(\widehat{\pi}_1, \widehat{\pi}_2) > U_2(\pi_1, \pi_2)$.*

The definition differs from joint monotonicity because it requires that for any material payoff pair, there is an arbitrarily close alternative material payoff pair giving more to both players that *both* agents strictly prefer. If the players' preferences satisfy social monotonicity, then both players' preferences are joint-monotonic, but both players' preferences can be joint-monotonic without satisfying social monotonicity. (In comparing social monotonicity with joint monotonicity, Dufwenberg et al mis-state the definition of joint monotonicity to be essentially the same as my statement of social monotonicity.)

Under the same conditions as Theorem 1, except that the players' distributional preferences are assumed to be socially monotonic, Dufwenberg et al prove that the set of UPE material payoff pairs is a subset of the set of MPE material payoff pairs. Their result is more general than Theorem 1 because if one player's distributional preferences are joint-monotonic and the other player's distributional preferences are monotonic, then the players' preferences satisfy social monotonicity.

We now turn from discussing which material payoff pairs are efficient to discussing whether the equilibrium is efficient. Theorem 2 in the main text gives necessary conditions for the equilibrium to be MPE. That theorem applies directly when both players' preferences are joint-monotonic. As discussed above, however, when both players' preferences are joint-monotonic, there may be UPE transactions that are not MPE. Theorem A2 presents necessary conditions for the equilibrium to be UPE. If the equilibrium is MPE, then it is also UPE, but there are also other cases where the equilibrium is UPE but not MPE.

Theorem A2. *Suppose U_1 and U_2 are joint-monotonic and quasi-concave, and both are either twice-continuously differentiable or fairness-kinked. If the equilibrium $(a_1, a_2(a_1))$ is UPE and not MPE, then $(a_1, a_2(a_1))$ is a fairness-rule optimum for SM. If, in addition, $(a_1, a_2(a_1))$ is a strict fairness-rule optimum for SM and any fairness rule is continuously differentiable, then at least one of the following must be true:*

1. *SM's indifference curve for disadvantageously unfair transactions is tangent to SM's fairness rule at $\boldsymbol{\pi}(a_1, a_2(a_1))$.*
2. *U_1 is fairness-kinked, $\boldsymbol{\pi}(a_1, a_2(a_1))$ is on FM's fairness rule, and the respective fairness rules f_1 and f_2 have different slopes at $\boldsymbol{\pi}(a_1, a_2(a_1))$.*

Figure A2a illustrates the Case 2 listed in the theorem, which can be interpreted as a setting where the two agents have different, self-serving ideas about what is fair. However, Figure A2b shows that even if the equilibrium occurs on both players' fairness rules, the equilibrium is not necessarily UPE. A corollary of Theorem A2 is that if both players' interpersonal indifference curves are smooth—thereby ruling out fairness-kinkedness—then the equilibrium is UPE if and only if it is MPE.

Theorems 3 and 4 provide sufficient conditions for the equilibrium to be MPE and UPE, but they assume that FM's preferences are purely self-regarding or monotonic. If FM's preferences are required only to be joint-monotonic, then the conclusions of the theorems may not hold. Even though SM's behavior aligns the *material* incentives of the two players, if FM's preferences are non-monotonic, then she may prefer not to maximize the players' material payoffs. For the case of preferences that satisfy the conditions of Theorem 3—except that FM's distributional preferences are merely joint-monotonic—Figure A3 illustrates an equilibrium that is neither MPE nor UPE.

[A1]

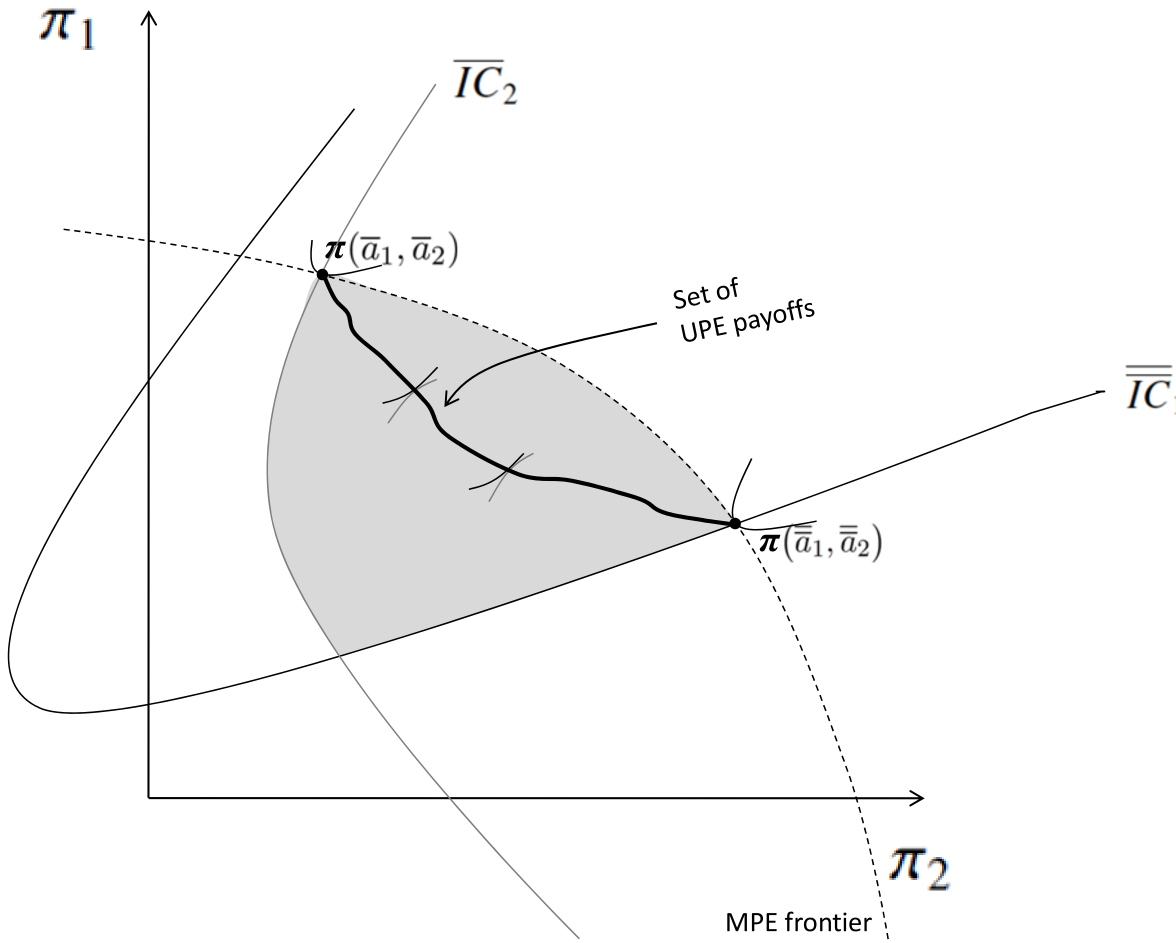

Figure A1. Relationship between utility Pareto efficiency and material Pareto-efficiency. Both players have joint-monotonic preferences. The set of UPE material payoff pairs must lie in the gray region.

[A2a] [A2b]

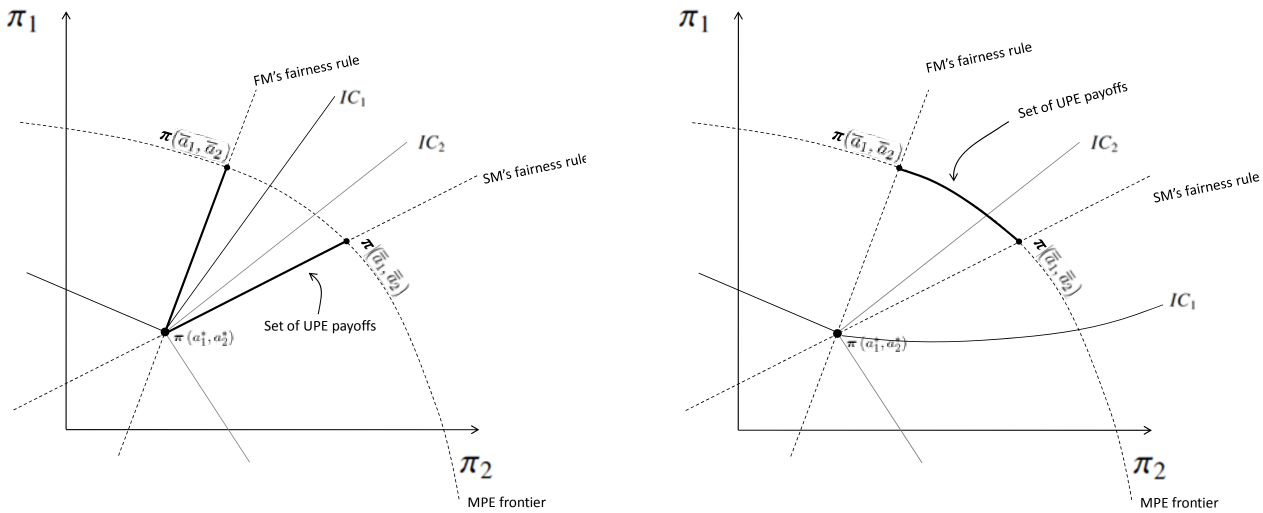

Figure A2. Panel (a): An equilibrium that is UPE but not MPE. Both players' indifference curves are fairness-kinked at the equilibrium, but FM and SM have different fairness rules. Panel (b): A similar situation, except that the equilibrium is neither UPE nor MPE.

[A3]

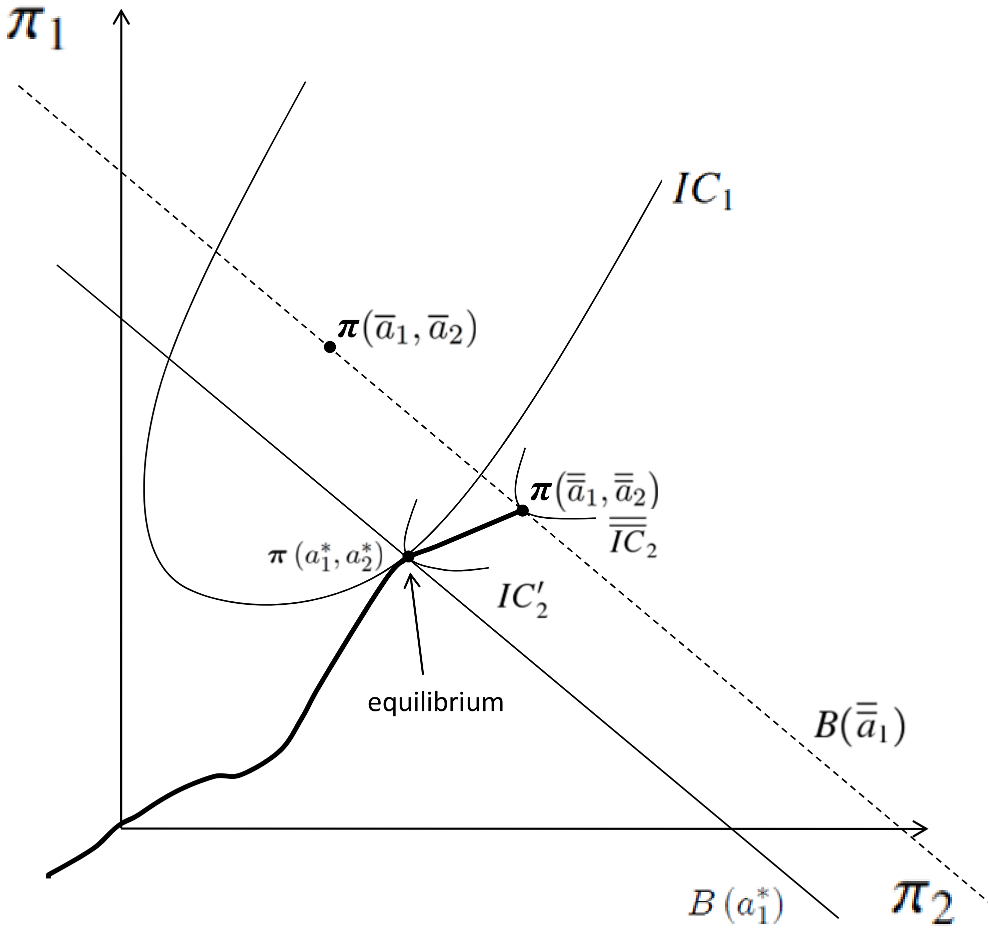

Figure A3. The conditions of Theorem 3 are satisfied, except that FM's distributional preferences are merely joint-monotonic (and not monotonic). Because FM's distributional preferences are joint-monotonic, the equilibrium does not occur at SM's favorite transaction. The dark line shows the path of material payoff pairs that could occur, $\pi(a_1, a_2(a_1))$, given different possible actions by FM. This path increases up to SM's favorite transaction and then goes back down the same path (if FM takes an "inefficiently high" level of her action). The equilibrium material payoff pair occurs at FM's most-preferred point along this path.

Appendix B: Proofs

Before proving the results in the text, we establish a technical lemma.

Technical Lemma. *Suppose* U_1 and U_2 *are joint-monotonic and quasi-concave. Then:*

1. *The set of individually-rational transactions*

$$T \equiv \{(a_1, a_2) \mid U_1(\boldsymbol{\pi}(a_1, a_2)) \geq 0, U_2(\boldsymbol{\pi}(a_1, a_2)) \geq 0\}$$

is non-empty and compact, as is the set of payoff pairs $T_\pi \equiv \{\boldsymbol{\pi}(a_1, a_2) \mid (a_1, a_2) \in T\}$.

2. *Along any graph of the form* $(g(\pi_2), \pi_2)$*, where* g *is a continuous, decreasing, weakly concave function,* U_i *has a unique maximum* π_2^* *and strictly decreases as* π_2 *moves away from this maximum, for* $i = 1, 2$*. Moreover, the MPE frontier and each budget curve* $B(a_1)$ *is such a graph.*

Proof of part 1: The transaction $(a_1, a_2) = (0, 0)$ gives material payoffs $\boldsymbol{\pi}(0, 0) = (0, 0)$ and utilities $U_1(\boldsymbol{\pi}(0, 0)) = U_2(\boldsymbol{\pi}(0, 0)) = 0$, so both sets are non-empty. By TA2, T necessarily lies to the north and east (respectively) of two lines $\pi_1 = \underline{\underline{\pi}}_1 \leq \underline{\pi}_1$ and $\pi_2 = \underline{\underline{\pi}}_2 \leq \underline{\pi}_2$, i.e., $T \subseteq \{(a_1, a_2) \mid \pi_1(a_1, a_2) \geq \underline{\underline{\pi}}_1, \pi_2(a_1, a_2) \geq \underline{\underline{\pi}}_2\}$. Hence T_π is closed and bounded and therefore compact. It follows from A4 that T is also closed and bounded and therefore compact.

Proof of part 2: WLOG, consider U_2. We first show that for any real number k, the set $\{\pi_2 \mid U_2(g(\pi_2), \pi_2) \geq k\}$ is an interval (possibly unbounded). Let $\pi_2' < \pi_2''$ be two values in this set. By construction, $U_2 \geq k$ at $(g(\pi_2'), \pi_2')$ and $(g(\pi_2''), \pi_2'')$. It follows that $U_2 \geq k$ at $(g(\pi_2'), \pi_2'')$. (To see this, let $\overline{y} = \max\{y \in [g(\pi_2''), g(\pi_2')] \mid U_2(y, \pi_2'') \geq k\}$ (the maximum exists by continuity). If $\bar{y} = g(\pi_2')$ then we are done, so assume $\bar{y} < g(\pi_2')$. By joint-monotonicity, we can choose $\hat{y}, \hat{x}$ with $\bar{y} < \hat{y} < g(\pi_2')$ and $\hat{x} > \pi_2''$ so that $U_2(\hat{y}, \hat{x}) > U_2(\bar{y}, \pi_2'') \geq k$. The line segment connecting $(g(\pi_2'), \pi_2')$ and $(\hat{y}, \hat{x})$ meets the line $x = \pi_2''$ at a point with some y-coordinate strictly between $\bar{y}$ and $g(\pi_2')$. By quasi-concavity, the value of U_2 at this point is $\geq k$. This contradicts the maximality of $\bar{y}$.) Now, for any $\pi_2' < \pi_2 < \pi_2''$, the point $(g(\pi_2), \pi_2)$ lies weakly inside the triangle defined by these three points since g is weakly concave. Since U_2 is quasi-concave, $U_2(g(\pi_2), \pi_2) \geq k$ also.

This shows that there cannot be three values $\pi_2' < \pi_2 < \pi_2''$ with $U_2(g(\pi_2'), \pi_2') > U_2(g(\pi_2), \pi_2) < U_2(g(\pi_2''), \pi_2'')$. It follows that on the graph $(g(\pi_2), \pi_2)$, U_2 is either weakly monotonic everywhere, or weakly increasing on $(-\infty, \widetilde{\pi}_2)$ and weakly decreasing on $(\widetilde{\pi}_2, \infty)$ for some $\widetilde{\pi}_2$.

We now show that U_2 cannot be constant on any interval along the graph. Suppose U_2 assumes the constant value k on the interval $[\pi_2', \pi_2'']$. Quasi-concavity implies that U_2 is $\geq k$ at the point $(y_0, x_0) = \left(\frac{g(\pi_2')+g(\pi_2'')}{2}, \frac{\pi_2'+\pi_2''}{2}\right)$. For sufficiently small $\epsilon > 0$, the box $[y_0, y_0+\epsilon] \times [x_0, x_0+\epsilon]$ lies entirely below and to the left of the curve $C = \{(g(\pi_2), \pi_2) \mid \pi_2' < \pi_2 < \pi_2''\}$. Joint-monotonicity ensures that U_2 assumes a value $k' > k$ at some point (y', x') inside this box. Now, let $S = \{(y, x) \mid y \geq y',\ x \geq x',\ U_2(y, x) \geq U_2(y', x')\}$. We know that S does not intersect C because $U_2 \geq k'$ on S, whereas U_2 takes on the constant value k on C, by assumption. S is closed and convex, and must then be bounded (by the lines $y = y', x = x'$, as well as by the curve C since $(y', x') \in S$), so it is compact. Hence we can choose a point $(y, x) \in S$ with $x + y$ maximal. But by joint-monotonicity there exists $y'' > y', x'' > x'$ with $U_2(y'', x'') > U_2(y', x') \geq k'$, contradicting maximality. It follows that U_2 cannot be constant on $[\pi_2', \pi_2'']$ after all.

Next, we rule out that U_2 is monotonic along the entire graph; in particular, we show that for any $(g(\pi_2), \pi_2)$, there are $\pi_2' < \pi_2 < \pi_2''$ such that $U_2(g(\pi_2'), \pi_2') < U_2(g(\pi_2), \pi_2) > U_2(g(\pi_2''), \pi_2'')$. Since the graph is weakly concave, the indifference curve going through $(g(\pi_2), \pi_2)$ is either tangent to the budget curve or by TA2 intersects it at $(g(\pi_2), \pi_2)$ and at some other point $(g(\pi_2'''), \pi_2''')$. In either cases, the claim follows immediately.

We complete the proof by showing that each budget curve has a graph of the form $(g(\pi_2), \pi_2)$, where g is a continuous, decreasing, weakly concave function; we omit the proof of the same for the MPE frontier, for which the argument is analogous (and is a standard result about the "utility possibility frontier" when utility is purely self-regarding). Fix action $\overline{a}_1$. Let the budget curve $B(\overline{a}_1) \equiv \{\boldsymbol{\pi}(\overline{a}_1, a_2)\}_{a_2 \in \mathbb{R}}$ be parameterized by $\pi_1(\overline{a}_1, a_2) \equiv g(\pi_2(\overline{a}_1, a_2))$; clearly, g is not only continuous but also continuously differentiable with $\left.\frac{d\pi_1}{d\pi_2}\right|_{B(\overline{a}_1)} = \frac{dg}{d\pi_2}$. Differentiating $\pi_1(\overline{a}_1, a_2) \equiv g(\pi_2(\overline{a}_1, a_2))$ with respect to a_2 yields $\frac{\partial \pi_1}{\partial a_2} = \frac{dg}{d\pi_2}\frac{\partial \pi_2}{\partial a_2}$, and therefore $\left.\frac{d\pi_1}{d\pi_2}\right|_{B(\overline{a}_1)} = \frac{dg}{d\pi_2} = \frac{\partial \pi_1/\partial a_2}{\partial \pi_2/\partial a_2} < 0$. Hence g is decreasing. By the chain rule, $\frac{\partial}{\partial a_2}\left(\frac{dg}{d\pi_2}\right) = \frac{d^2 g}{d(\pi_2)^2}\frac{\partial \pi_2}{\partial a_2}$. Rearranging, $\frac{d^2 g}{d(\pi_2)^2} = \frac{\frac{\partial}{\partial a_2}\left(\frac{dg}{d\pi_2}\right)}{\partial \pi_2/\partial a_2} = \frac{\frac{\partial}{\partial a_2}\left(\frac{\partial \pi_1/\partial a_2}{\partial \pi_2/\partial a_2}\right)}{\partial \pi_2/\partial a_2}$. A3 implies that $\frac{\partial}{\partial a_2}\left(\frac{\partial \pi_1/\partial a_2}{\partial \pi_2/\partial a_2}\right) \geq 0$, and A1 implies that $\frac{\partial \pi_2}{\partial a_2} < 0$, so $\frac{d^2 g}{d(\pi_2)^2} \leq 0$. Hence g is weakly concave.

□

Proposition 1 (Rotten kid theorem). *In the equilibrium of the rotten kid game, the child chooses the level of a_1 that maximizes family income.*

Proof: Follows directly from Theorem 3 below, and here we merely check that the assumptions can be verified or appropriately modified. A1-A4 from Section 3 clearly hold, with A2 following from $n'(0) < 1$. TA1 from Section 4 and joint-monotonicity of U_2 are satisfied due to the assumption that $U_2(\pi_1, \pi_2)$ is monotonically increasing in both π_1 and π_2. TA2 from Section 4 is implied by the assumption that there exist $\underline{\pi}_1 < 0$ and $\underline{\pi}_2 < 0$ such that $\lim_{\pi_2 \to \infty} \frac{\partial U_2(\underline{\pi}_1, \pi_2)/\partial \pi_2}{\partial U_2(\underline{\pi}_1, \pi_2)/\partial \pi_1} = 0$ and $\lim_{\pi_1 \to \infty} \frac{\partial U_2(\pi_1, \underline{\pi}_2)/\partial \pi_2}{\partial U_2(\pi_1, \underline{\pi}_2)/\partial \pi_1} = \infty$. We have directly assumed that U_2 is quasi-concave and normal, and U_1 is purely self-regarding. The material payoff functions being quasi-linear implies that they are globally conditionally transferable. There is no assumption that $U_1\left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right) \geq 0$ because neither player has an outside option.

$\square$

Proposition 2. *In the gift-exchange game with a profit-maximizing firm, there exists* $\overline{\sigma} > 0$ *such that if* $\sigma < \overline{\sigma}$ *and* $\rho \geq \frac{1}{2}$, *then the equilibrium transaction is Pareto efficient in terms of the material payoffs.*

Proof: Follows directly from Theorem 4 below, and here we merely check that the assumptions can be verified or appropriately modified. A1-A4 from Section 3 clearly hold, with A2 following from $c'(0) < 1$. TA1 from Section 4 clearly holds. TA2 from Section 4 and the quasi-concavity of U_2 are not needed because the piecewise-linear functional form for U_2, combined with the assumptions regarding the material payoff functions, ensure that an optimal action for the worker exists in response to any a_1. The functional form for U_2 satisfies joint-monotonicity and fairness-kinkedness, and we have directly assumed that U_1 is purely self-regarding. S2 from Section 8.2 clearly holds. S3 can be replaced in the proof of Theorem 4 by the assumption of the piecewise-linear functional form for U_2. S4 and S5 hold but can be dropped as sufficient conditions because FM is purely self-regarding. S1 is satisfied as long as $(1-\sigma) - c'(\widehat{a}_2)\sigma > 0$; or rearranging, $\sigma < \frac{1}{1+c'(\widehat{a}_2)}$. In the next paragraph, we will show that the assumption that $\left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right)$ is a strict fairness-rule optimum is satisfied as long as $\sigma < \frac{1}{2}$ and $\rho > \frac{1}{2}$. The conclusion then follows from setting $\overline{\sigma} = \min\left\{\frac{1}{2}, \frac{1}{1+c'(\widehat{a}_2)}\right\}$ and noting (as explained below) that here $\rho > \frac{1}{2}$ can be weakened to $\rho \geq \frac{1}{2}$.

We now show that $\left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right)$, defined implicitly as the solution to $\pi_1\left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right) = \pi_2\left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right)$ and $c'\left(\overline{\overline{a}}_2\right) = 1$, satisfies $U_1\left(\boldsymbol{\pi}\left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right)\right) = \pi_1\left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right) > 0$ and is a fairness-rule optimum. (Given the assumptions on the material payoff functions and the $c(\cdot)$ function, the solution to these equations exists and is unique.) The conditions on $c(\cdot)$ ensure that the solutions to these equations indeed satisfy $\pi_1\left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right) > 0$. Translated to this gift-exchange game, the two conditions for $\left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right)$ to be a strict fairness-rule optimum (as defined in Section 5) are $(1-\sigma) - c'\left(\overline{\overline{a}}_2\right)\sigma > 0$ and

$(1-\rho) - c'\left(\overline{\overline{a}}_2\right)\rho < 0$. Substituting $c'\left(\overline{\overline{a}}_2\right) = 1$, these two conditions are satisfied as long as $\sigma < \frac{1}{2}$ and $\rho > \frac{1}{2}$, respectively. The latter condition can be weakened to $\rho \geq \frac{1}{2}$ because if $\rho = \frac{1}{2}$, $\overline{\overline{a}}_1$ remains a local optimum for FM since $(1-\rho) - c'(a_2)\rho < 0$ continues to hold for all $a_2 < \overline{\overline{a}}_2$.

□

Lemma 1. *Suppose U_2 is joint-monotonic and quasi-concave. For any a_1, SM has a unique optimal best response, $a_2(a_1)$, that is a continuous function of a_1. Moreover, if U_2 is continuously differentiable at some $(\widehat{a}_1, a_2(\widehat{a}_1))$, then $\frac{\partial U_2}{\partial \pi_1} > 0$ and $\frac{\partial U_2}{\partial \pi_2} > 0$ at $(\widehat{a}_1, a_2(\widehat{a}_1))$.*

Proof: Technical Lemma immediately gives existence and uniqueness of an optimal action $a_2(a_1)$. The Maximum Theorem (e.g., Sundaram 1996, p.235) can now be applied (where we can ignore the compactness requirement on the budget curve since we have already proved existence of an optimal action) to show that $a_2(a_1)$ is an upper-hemicontinuous correspondence. Since $a_2(a_1)$ is single-valued, it is a continuous function.

Since U_2 is continuously differentiable at $\boldsymbol{\pi}(\widehat{a}_1, a_2(\widehat{a}_1))$, SM's unique optimum is characterized by the first-order condition, $\frac{\partial U_2}{\partial a_2}(\boldsymbol{\pi}(\widehat{a}_1, a_2)) = 0$, which after rearranging is $\frac{\partial U_2}{\partial \pi_2} - p(\widehat{a}_1, a_2)\frac{\partial U_2}{\partial \pi_1} = 0$. Joint-monotonicity rules out that both partial derivatives $\frac{\partial U_2}{\partial \pi_1}$ and $\frac{\partial U_2}{\partial \pi_2}$ are negative, and TA1 rules out that they both equal 0. Therefore, the first-order condition implies that both are positive.

□

Proposition 3.

1. *Suppose U_2 is joint-monotonic, quasi-concave, and fairness-kinked. Suppose that $(\widehat{a}_1, a_2(\widehat{a}_1))$ is a strict fairness-rule optimum. Then $(a_1, a_2(a_1))$ is a strict fairness-rule optimum for all a_1 in a neighborhood of $\widehat{a}_1$, and $a_2(a_1)$ is increasing in a_1 at $\widehat{a}_1$. Furthermore, U_2 is locally normal in π_1 and π_2 at $(p(\widehat{a}_1, a_2(\widehat{a}_1)); I(\widehat{a}_1, a_2(\widehat{a}_1)))$.*

2. *Suppose $\frac{\partial}{\partial a_1}\left(\frac{\partial \pi_1/\partial a_2}{\partial \pi_2/\partial a_2}\right) \leq 0$ and U_2 is joint-monotonic and quasi-concave. If U_2 is weakly locally normal at $(p(\widehat{a}_1, a_2(\widehat{a}_1)); I(\widehat{a}_1, a_2(\widehat{a}_1)))$, then $a_2(a_1)$ is increasing in a_1 at $\widehat{a}_1$. Hence if U_2 is weakly normal in π_1, then $a_2(a_1)$ is increasing in a_1.*

Proof of part 1: By definition of $(\widehat{a}_1, a_2(\widehat{a}_1))$ being a strict fairness-rule optimum:

$$\lim_{\boldsymbol{\pi} \to \boldsymbol{\pi}(\widehat{a}_1, a_2(\widehat{a}_1)), \boldsymbol{\pi} \in D_f} \left(\frac{\partial U_2(\boldsymbol{\pi})}{\partial \pi_2} - p(\widehat{a}_1, a_2(\widehat{a}_1))\frac{\partial U_2(\boldsymbol{\pi})}{\partial \pi_1}\right) > 0,$$

and

$$\lim_{\boldsymbol{\pi}\to\boldsymbol{\pi}(\widehat{a}_1,a_2(\widehat{a}_1)),\boldsymbol{\pi}\in A_f}\left(\frac{\partial U_2(\boldsymbol{\pi})}{\partial\pi_2}-p(\widehat{a}_1,a_2(\widehat{a}_1))\frac{\partial U_2(\boldsymbol{\pi})}{\partial\pi_1}\right)<0.$$

Since these inequalities are strict and since $a_2(a_1)$ is a continuous function of a_1 (by Lemma 1), it follows immediately that these inequalities hold for all a_1 in a neighborhood of $\widehat{a}_1$, and thus $(a_1, a_2(a_1))$ is a strict fairness-rule optimum for all a_1 in a neighborhood of $\widehat{a}_1$. Thus, for any a_1 in a neighborhood of $\widehat{a}_1$, SM will choose action $a_2(a_1)$ such that $\boldsymbol{\pi}(a_1, a_2(a_1)) \in \text{graph}(f)$. The fact that the fairness rule is a strictly upward-sloping locus of material payoff pairs, together with A1, implies that $a_2(a_1)$ is increasing in a_1 at $\widehat{a}_1$. Because the above inequalities are strict, they also imply that for any a_1 in a neighborhood of $\widehat{a}_1$,

$$\lim_{\boldsymbol{\pi}\to\boldsymbol{\pi}(a_1,a_2(a_1)),\boldsymbol{\pi}\in D_f}\left(\frac{\partial U_2(\boldsymbol{\pi})}{\partial\pi_2}-p(\widehat{a}_1,a_2(\widehat{a}_1))\frac{\partial U_2(\boldsymbol{\pi})}{\partial\pi_1}\right)>0$$

and

$$\lim_{\boldsymbol{\pi}\to\boldsymbol{\pi}(a_1,a_2(a_1)),\boldsymbol{\pi}\in A_f}\left(\frac{\partial U_2(\boldsymbol{\pi})}{\partial\pi_2}-p(\widehat{a}_1,a_2(\widehat{a}_1))\frac{\partial U_2(\boldsymbol{\pi})}{\partial\pi_1}\right)<0$$

(where note that we are now holding the price fixed at $p(\widehat{a}_1, a_2(\widehat{a}_1))$ as a_1 varies). It follows that U_2 is locally normal in π_1 and π_2 at $(p(\widehat{a}_1, a_2(\widehat{a}_1)); I(\widehat{a}_1, a_2(\widehat{a}_1)))$.

Proof of part 2: Consider a small increase in FM's action $\widehat{a}_1' > \widehat{a}_1$. Assume (for contradiction) that SM weakly decreases his action, so that SM's material payoff rises while FM's falls. Call A the allocation $\boldsymbol{\pi}(\widehat{a}_1, a_2(\widehat{a}_1))$ and B the allocation $\boldsymbol{\pi}(\widehat{a}_1', a_2(\widehat{a}_1'))$. In the (π_2, π_1) plane, A is northwest of B. Now draw two downward-sloping lines with slopes $-p(\widehat{a}_1, a_2(\widehat{a}_1))$ and $-p(\widehat{a}_1', a_2(\widehat{a}_1')) \leq -p(\widehat{a}_1, a_2(\widehat{a}_1))$ going through A and B, respectively; this inequality is implied by $\frac{\partial p}{\partial a_2} \leq 0$ (by Part 2 of the Technical Lemma) and $-\frac{\partial}{\partial a_1}\left(\frac{\partial\pi_1/\partial a_2}{\partial\pi_2/\partial a_2}\right) = \frac{\partial p}{\partial a_1} \geq 0$ (by hypothesis). If these two slopes are equal, then weak local normality is contradicted. We can therefore assume that $-p(\widehat{a}_1', a_2(\widehat{a}_1')) < -p(\widehat{a}_1, a_2(\widehat{a}_1))$, so that the slope of the line through B is steeper than the slope of the line through A.

The two lines will intersect at some generic point, say C. There are three cases. Case 1 is that C is strictly southeast of both A and B, and Case 2 is that C is strictly southeast of A and northwest of B. The proof in these two cases proceeds identically: The change from A to B can be decomposed into a substitution effect and an income effect. The substitution effect causes a move from A to a point A' weakly northwest of A. Because of weak normality in π_1, the income effect then makes us move from A' to B, where B needs to be weakly north of A'—and therefore weakly north of A. But B actually lies strictly south of A, a contradiction.

Case 3 is that C is strictly northwest of both A and B. The change from A to B can again be decomposed into a substitution effect and an income effect, where the substitution effect causes a move from A to a point A' weakly northwest of A. Because of weak normality in π_2, the income effect then makes us move from A' to B, where B needs to be weakly west of A'—and therefore weakly west of A. But B actually lies strictly east of A, a contradiction.

□

Theorem 1. *Suppose U_1 is monotonic and quasi-concave, and suppose U_2 is joint-monotonic and quasi-concave. FM's and SM's favorite transactions, $(\overline{a}_1, \overline{a}_2)$ and $\left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right)$, exist and are unique. The set of UPE material payoff pairs coincides exactly with the set of material payoff pairs on the MPE frontier between $(\overline{\pi}_1, \overline{\pi}_2)$ and $\left(\overline{\overline{\pi}}_1, \overline{\overline{\pi}}_2\right)$.*

Proof: We will prove that SM's favorite transaction exists, and deduce the result for FM by symmetry. If SM's favorite transaction exists, then joint-monotonicity implies that it must lie on the MPE frontier. Technical Lemma implies that there does in fact exist a maximizing material payoff pair on the MPE frontier, and it is unique. Since this payoff pair is on the MPE frontier, there is in turn exactly one transaction $\left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right)$ that achieves these payoffs. To see that, we will work in the (a_1, a_2) plane and study the *material* indifference curves for FM and SM. At a MPE action pair, we must have a tangency between the material indifference curves: $-\frac{\partial \pi_1/\partial a_1}{\partial \pi_1/\partial a_2} = \frac{da_2}{da_1}\Big|_{\pi_1=\overline{\overline{\pi}}_1} = \frac{da_2}{da_1}\Big|_{\pi_2=\overline{\overline{\pi}}_2} = -\frac{\partial \pi_2/\partial a_1}{\partial \pi_2/\partial a_2}$. By A3, $\frac{d^2a_2}{d(a_1)^2}\Big|_{\pi_1=\overline{\overline{\pi}}_1} \geq 0$ and $\frac{d^2a_2}{d(a_1)^2}\Big|_{\pi_2=\overline{\overline{\pi}}_2} \leq 0$ with at least one of these equalities strict. It follows that SM's favorite transaction is unique.

FM's favorite material payoff pair $(\overline{\pi}_1, \overline{\pi}_2)$ is UPE because there is no alternative feasible material payoff pair that FM prefers. Analogously, SM's favorite material payoffs pair $\left(\overline{\overline{\pi}}_1, \overline{\overline{\pi}}_2\right)$ is UPE because there is no alternative feasible material payoff pair that SM prefers.

Note that no material payoff pair (π_1', π_2') that is strictly within the materially-feasible set can be UPE; by joint-monotonicity of U_2, there is some feasible material payoff pair $(\pi_1'', \pi_2'') \gg (\pi_1', \pi_2')$ that SM prefers, and FM also prefers (π_1'', π_2'') by monotonicity.

Finally, any material payoff pair (π_1', π_2') on the MPE frontier between $(\overline{\pi}_1, \overline{\pi}_2)$ and $\left(\overline{\overline{\pi}}_1, \overline{\overline{\pi}}_2\right)$ is UPE. For contradiction, suppose (π_1', π_2') is not UPE. Then there exists another material payoff pair (π_1'', π_2'') giving at least equally high utility to both players. We may assume (π_1'', π_2'') to be MPE; if not, then by joint-monotonicity, there exists an MPE material payoff pair giving yet higher utility to both players that we can use instead. Suppose $(\overline{\pi}_1, \overline{\pi}_2)$ is northwest of $\left(\overline{\overline{\pi}}_1, \overline{\overline{\pi}}_2\right)$ on the MPE frontier; the argument is analogous if the positioning is reversed. If (π_1'', π_2'') is northwest of (π_1', π_2') on the MPE frontier, then $(\pi_1'', \pi_2''), (\pi_1', \pi_2'), (\overline{\overline{\pi}}_1, \overline{\overline{\pi}}_2)$ lie in that order along the MPE frontier, and $U_2(\pi_1'', \pi_2'') \geq U_2(\pi_1', \pi_2') < U_2(\overline{\overline{\pi}}_1, \overline{\overline{\pi}}_2)$; but this contradicts the Technical Lemma. On the other hand, if (π_1'', π_2'') is southeast of (π_1', π_2') on the MPE frontier, then $(\overline{\pi}_1, \overline{\pi}_2), (\pi_1', \pi_2'), (\pi_1'', \pi_2'')$ lie in that order along the MPE frontier, and $U_1(\overline{\pi}_1, \overline{\pi}_2) > U_1(\pi_1', \pi_2') \leq U_1(\pi_1'', \pi_2'')$; but this also contradicts the Technical Lemma.

□

Lemma 2. *Suppose U_2 is joint-monotonic and quasi-concave. Then there exists a unique $\widehat{a}_1$ such that the resulting transaction $(\widehat{a}_1, a_2(\widehat{a}_1))$ is MPE. This transaction is SM's favorite transaction $(\overline{\overline{a}}_1, \overline{\overline{a}}_2)$, and it is UPE.*

Proof: We will prove that given any action $\widehat{a}_1$, the transaction $(\widehat{a}_1, a_2(\widehat{a}_1))$ resulting from the unique best-response $a_2(\widehat{a}_1)$ is MPE if and only if $(\widehat{a}_1, a_2(\widehat{a}_1))$ is SM's favorite transaction. The "if" direction follows immediately from the fact that SM's favorite transaction is MPE (Theorem 1), so we focus on the "only if" direction. Suppose $(\widehat{a}_1, a_2(\widehat{a}_1))$ is MPE but is not SM's favorite transaction $(\overline{\overline{a}}_1, \overline{\overline{a}}_2)$. Every point on the MPE frontier $\boldsymbol{\pi}(a_1, a_2)$ touches exactly one budget curve, $B(a_1)$; the transaction (a_1, a_2) satisfies the MPE condition $\frac{\partial\pi_1/\partial a_1}{\partial\pi_1/\partial a_2} = \frac{\partial\pi_2/\partial a_1}{\partial\pi_2/\partial a_2}$, which implies $\left.\frac{d\pi_1}{d\pi_2}\right|_{MPE} = \frac{\partial\pi_1/\partial a_1}{\partial\pi_2/\partial a_1} = \frac{\partial\pi_1/\partial a_2}{\partial\pi_2/\partial a_2} = \left.\frac{d\pi_1}{d\pi_2}\right|_{B(a_1)}$, and therefore the budget curve is tangent to the MPE frontier at $\boldsymbol{\pi}(a_1, a_2)$. Hence SM's indifference curve passing through $\boldsymbol{\pi}(\widehat{a}_1, a_2(\widehat{a}_1))$ is tangent to the MPE frontier at $\boldsymbol{\pi}(\widehat{a}_1, a_2(\widehat{a}_1))$. So there is some $\boldsymbol{\pi}(a_1', a_2')$ on the MPE frontier between $\boldsymbol{\pi}(\widehat{a}_1, a_2(\widehat{a}_1))$ and $\boldsymbol{\pi}(\overline{\overline{a}}_1, \overline{\overline{a}}_2)$, sufficiently close to $\boldsymbol{\pi}(\widehat{a}_1, a_2(\widehat{a}_1))$, such that $U_2(\boldsymbol{\pi}(a_1', a_2')) < U_2(\boldsymbol{\pi}(\widehat{a}_1, a_2(\widehat{a}_1)))$. But this contradicts the fact that U_2 is strictly decreasing as we move away from $\boldsymbol{\pi}(\overline{\overline{a}}_1, \overline{\overline{a}}_2)$ along the MPE frontier (as stated in Technical Lemma).

Finally, Theorem 1 states that SM's favorite transaction is UPE.

□

Lemma 3. *An equilibrium exists. Moreover, if $U_1(\overline{\overline{\pi}}_1, \overline{\overline{\pi}}_2) \geq 0$, then an equilibrium exists in which the players exchange rather than taking their outside options.*

Proof: From Lemma 2, if FM chooses action $\overline{\overline{a}}_1$, SM will choose action $\overline{\overline{a}}_2$. The facts that $U_2(\boldsymbol{\pi}(0,0)) = 0$ and $(\overline{\overline{a}}_1, \overline{\overline{a}}_2)$ is SM's favorite transaction imply that $U_2(\boldsymbol{\pi}(\overline{\overline{a}}_1, \overline{\overline{a}}_2)) \geq 0$. Since some action other than $\overline{\overline{a}}_1$ may give FM an even higher utility than $U_1(\boldsymbol{\pi}(\overline{\overline{a}}_1, \overline{\overline{a}}_2)) \geq 0$, 0 is a lower bound on FM's equilibrium utility. From Technical Lemma, the set of individually-rational transactions T is compact. Since $U_1(\boldsymbol{\pi}(a_1, a_2(a_1)))$ is continuous, there exists an optimal action a_1 in T. The result follows.

□

Proposition 4. *Suppose U_1 and U_2 are joint-monotonic and quasi-concave. If the equilibrium $(a_1, a_2(a_1))$ is MPE, then $(a_1, a_2(a_1))$ is SM's favorite transaction, and $U_1(\boldsymbol{\pi}(a_1, a_2(a_1))) \geq 0$.*

Proof: The fact that $(a_1, a_2(a_1))$ is SM's favorite transaction follows directly from Lemma 2. Suppose that $U_1(\boldsymbol{\pi}(a_1, a_2(a_1))) < 0$. Then FM would choose her outside option rather than taking action a_1, so $(a_1, a_2(a_1))$ is not an equilibrium. But this is a contradiction.

□

Theorem 2. *Suppose U_1 and U_2 are joint-monotonic and quasi-concave, and U_2 is either twice-continuously differentiable or fairness-kinked. If the equilibrium $(a_1, a_2(a_1))$ is MPE, then at least one of the following must be true:*

1. *$(a_1, a_2(a_1))$ is FM's favorite transaction.*
2. $\frac{dp(a_1, a_2(a_1))}{da_1} = 0$.
3. *U_2 is fairness-kinked, and $(a_1, a_2(a_1))$ is a fairness-rule optimum.*

Proof: SM's best-response function $a_2(a_1)$ solves the problem of choosing SM's most-preferred material payoff pair along the budget curve $B(a_1)$:

$$(\pi_1^*(a_1, a_2(a_1)), \pi_2^*(a_1, a_2(a_1))) = \arg\max_{\boldsymbol{\pi}} U_2(\boldsymbol{\pi}) \text{ subject to } \boldsymbol{\pi} \in B(a_1). \tag{1}$$

As described in the text and illustrated in Figure 2, the solution to this problem, $\boldsymbol{\pi}$, is the same as the solution to the standard consumer optimization where the budget line is the linear approximation to the budget curve at the solution $\boldsymbol{\pi}^*(a_1, a_2(a_1))$ to the problem (1):

$$(\widetilde{\pi}_1(p, I), \widetilde{\pi}_2(p, I)) = \arg\max_{\boldsymbol{\pi}} U_2(\boldsymbol{\pi}) \text{ subject to } \pi_1 + p\pi_2 = I, \tag{2}$$

where $p = p(a_1, a_2(a_1)) = -\left.\frac{d\pi_1}{d\pi_2}\right|_{B(a_1)}$ and $I = \pi_1^*(a_1, a_2(a_1)) + p(a_1, a_2(a_1))\, \pi_2^*(a_1, a_2(a_1))$. Since U_2 is either twice-continuously differentiable or fairness-kinked, $p(a_1, a_2(a_1))$, $I(a_1, a_2(a_1))$, $\widetilde{\pi}_1(p, I)$, and $\widetilde{\pi}_2(p, I)$ are all continuously differentiable functions. Now, there are two possible cases, depending on whether the change in FM's action leads to a change in p.

Case 1: $\frac{dp(a_1,a_2(a_1))}{da_1} \neq 0$. The Slutsky equation can be applied to find the effects on $\widetilde{\pi}_1$ and $\widetilde{\pi}_2$:

$$\begin{aligned} \frac{d}{da_1}\widetilde{\pi}_1(p,\, I) &= \frac{d\widetilde{\pi}_1(p,\, \widetilde{\pi}_1 + p\widetilde{\pi}_2)}{dp} + \frac{\partial \widetilde{\pi}_1(p,\, I)}{\partial I}(\omega_1 - \pi_1^*) \\ \frac{d}{da_1}\widetilde{\pi}_2(p,\, I) &= \underbrace{\frac{d\widetilde{\pi}_2(p,\, \widetilde{\pi}_1 + p\widetilde{\pi}_2)}{dp}}_{\text{substitution effect}} + \underbrace{\frac{\partial \widetilde{\pi}_2(p,\, I)}{\partial I}(\omega_2 - \pi_2^*)}_{\text{income effect}} \end{aligned}$$

where π_1^* and π_2^* are the solutions from (1), (ω_1, ω_2) is the material payoff pair where the original budget line intersects with the new budget line (in standard consumer theory, this intersection point would be interpreted as the endowment consumption bundle), and we omit writing the dependence of p and I on $(a_1, a_2(a_1))$ to avoid cluttering notation.

To calculate the income effect, we begin by finding (ω_1, ω_2). We suppress dependence on $a_2(a_1)$ by writing the equation for the budget line as $\pi_1(a_1) = I(a_1) - p(a_1)\pi_2(a_1)$. Since (ω_1, ω_2) is the intersection of the old budget line and the new budget line, it satisfies $\omega_1 = I(a_1) - p(a_1)\omega_2$ and $\omega_1 = I(a_1 + \Delta a_1) - p(a_1 + \Delta a_1)\omega_2$. Solving these two equations simultaneously gives $\omega_2 = \frac{I(a_1+\Delta a_1) - I(a_1)}{p(a_1+\Delta a_1) - p(a_1)} = \frac{\frac{I(a_1+\Delta a_1)-I(a_1)}{\Delta a_1}}{\frac{p(a_1+\Delta a_1)-p(a_1)}{\Delta a_1}}$, so for small Δa_1,

$$\omega_2 = \frac{dI(a_1)/da_1}{dp(a_1)/da_1} \text{ and } \omega_1 = I(a_1) - p(a_1)\omega_2.$$

We now calculate $(\omega_1 - \pi_1^*)$ and $(\omega_2 - \pi_2^*)$. Using the definition of I, $\frac{dI(a_1)}{da_1} = \frac{dp(a_1)}{da_1}\pi_2^*(a_1) + p(a_1)\frac{d\pi_2^*(a_1)}{da_1} + \frac{d\pi_1^*(a_1)}{da_1}$. Substituting and simplifying gives

$$\begin{aligned} (\omega_2 - \pi_2^*) &= \frac{p(a_1, a_2(a_1))\frac{d\pi_2^*(a_1,a_2(a_1))}{da_1} + \frac{d\pi_1^*(a_1,a_2(a_1))}{da_1}}{\frac{dp(a_1,a_2(a_1))}{da_1}} \\ &= \frac{p(a_1, a_2(a_1))\frac{\partial\pi_2^*(a_1,a_2(a_1))}{\partial a_1} + \frac{\partial\pi_1^*(a_1,a_2(a_1))}{\partial a_1}}{\frac{dp(a_1,a_2(a_1))}{da_1}} = 0. \end{aligned}$$

The second equality can be intepreted as an envelope condition: the indirect effect through a_2, $p(a_1, a_2(a_1))\frac{\partial\pi_2^*(a_1,a_2(a_1))}{\partial a_2} + \frac{\partial\pi_1^*(a_1,a_2(a_1))}{\partial a_2} = 0$, equals zero because, at a fixed $p = p(a_1, a_2(a_1))$, SM has maximized "income" by choosing the material payoff pair on the MPE frontier. The third equality follows from $p \equiv -\left.\frac{d\pi_1}{d\pi_2}\right|_{B(a_1)} = -\frac{\partial\pi_1(a_1,a_2)/\partial a_2}{\partial\pi_2(a_1,a_2)/\partial a_2}$ and the MPE condition, $\frac{\partial\pi_1^*(a_1,a_2)/\partial a_2}{\partial\pi_2^*(a_1,a_2)/\partial a_2} = \frac{\partial\pi_1^*(a_1,a_2)/\partial a_1}{\partial\pi_2^*(a_1,a_2)/\partial a_1}$. Now, substituting $\omega_2 = \pi_2^*$ into the equation for ω_1 gives $\omega_1 = I(a_1) - p(a_1)\pi_2^*$, but since this expression equals π_1^*, $(\omega_1 - \pi_1^*) = 0$. Therefore, starting from an MPE transaction, the income effect from a change in FM's action equals zero.

To calculate the substitution effect, we define $\widetilde{I}(p) = p\widetilde{\pi}_2 + \widetilde{\pi}_1$ and use the implicit function theorem on the first-order condition for problem (2), $\frac{\partial U_2(\widetilde{I} - p\widetilde{\pi}_2, \widetilde{\pi}_2)}{\partial \pi_2} - p\frac{\partial U_2(\widetilde{I} - p\widetilde{\pi}_2, \widetilde{\pi}_2)}{\partial \pi_1} = 0$:

$$\frac{d\widetilde{\pi}_2(p,\,\widetilde{\pi}_1+p\widetilde{\pi}_2)}{dp} = -\frac{\frac{\partial^2 U_2}{\partial\pi_1\partial\pi_2}\left(\frac{d\widetilde{I}(p)}{dp}-\widetilde{\pi}_2\right)-\frac{\partial U_2}{\partial\pi_1}-p\frac{\partial^2 U_2}{\partial(\pi_1)^2}\left(\frac{d\widetilde{I}(p)}{dp}-\widetilde{\pi}_2\right)}{\frac{\partial^2 U_2}{\partial(\pi_2)^2}-2p\frac{\partial^2 U_2}{\partial\pi_1\partial\pi_2}+p^2\frac{\partial^2 U_2}{\partial(\pi_1)^2}}$$

$$= -\frac{-\left(\frac{\partial U_2}{\partial\pi_1}\right)^3}{\frac{\partial^2 U_2}{\partial(\pi_2)^2}\left(\frac{\partial U_2}{\partial\pi_1}\right)^2-2\frac{\partial U_2}{\partial\pi_1}\frac{\partial U_2}{\partial\pi_2}\frac{\partial^2 U_2}{\partial\pi_1\partial\pi_2}+\left(\frac{\partial U_2}{\partial\pi_2}\right)^2\frac{\partial^2 U_2}{\partial(\pi_1)^2}} = \frac{1}{-\left.\frac{d^2\pi_2}{d(\pi_1)^2}\right|_{U_2(\pi_1^*,\pi_2^*)}},$$

where the second equality follows from $\frac{d\widetilde{I}(p)}{dp}=\widetilde{\pi}_2$ and substituting SM's first-order condition for problem (2). A similar calculation yields $\frac{d\widetilde{\pi}_1(p,p\widetilde{\pi}_1+\widetilde{\pi}_2)}{dp}=\frac{p}{\left.d^2\pi_2/d(\pi_1)^2\right|_{U_2(\pi_1^*,\pi_2^*)}}$.

An interior equilibrium transaction satisfies FM's first-order condition, which can be written in terms of the budget lines: $\frac{d}{da_1}U_1(\widetilde{\pi}_1(p,\,I),\,\widetilde{\pi}_2(p,\,I))=0$. (A4 combined with joint-monotonicity of U_2 ensures that SM's favorite transaction is indeed interior.) Using the income and substitution effects derived above,

$$\frac{d}{da_1}U_1(\widetilde{\pi}_1(p,\,I),\,\widetilde{\pi}_2(p,\,I)) = \frac{\partial U_1}{\partial\pi_1}\frac{d}{da_1}\widetilde{\pi}_1(p,\,I)+\frac{\partial U_1}{\partial\pi_2}\frac{d}{da_1}\widetilde{\pi}_2(p,\,I)$$
$$= \left(\frac{\partial U_1}{\partial\pi_2}-p\frac{\partial U_1}{\partial\pi_1}\right)\cdot\frac{1}{-\left.\frac{d^2\pi_2}{d(\pi_1)^2}\right|_{U_2(\pi_1^*,\pi_2^*)}}.$$

Recall that $\frac{\partial U_1}{\partial\pi_1}>0$ and $\frac{\partial U_1}{\partial\pi_2}>0$ (from Lemma 1). Hence FM's first-order condition is satisfied only if (A) SM's indifference curve is kinked at (π_1^*,π_2^*), i.e., $\left.d^2\pi_2/d(\pi_1)^2\right|_{U_2(\pi_1^*,\pi_2^*)}=-\infty$; or (B) FM's favorite transaction is (π_1^*,π_2^*), i.e., $\frac{\partial U_1/\partial\pi_2}{\partial U_1/\partial\pi_1}=p$ at (π_1^*,π_2^*), which is also SM's favorite transaction.

Case 2: $\frac{dp(a_1,a_2(a_1))}{da_1}=0$. Since there is no substitution effect, the new and old budget lines do not intersect at an "endowment" (ω_1,ω_2). In this case, the Slutsky equation is:

$$\frac{d}{da_1}\widetilde{\pi}_1(p,\,I) = \frac{\partial\widetilde{\pi}_1(p,\,I)}{\partial I}\frac{dI(a_1,a_2(a_1))}{da_1}$$
$$\frac{d}{da_1}\widetilde{\pi}_2(p,\,I) = \underbrace{\frac{\partial\widetilde{\pi}_2(p,\,I)}{\partial I}\frac{dI(a_1,a_2(a_1))}{da_1}}_{\text{income effect}}.$$

Differentiating $I(a_1,a_2(a_1))=p(a_1,a_2(a_1))\,\pi_2(a_1,a_2(a_1))+\pi_1(a_1,a_2(a_1))$ gives

$$\frac{dI(a_1,a_2(a_1))}{da_1} = \left(\frac{\partial\pi_1}{\partial a_1}+p\frac{\partial\pi_2}{\partial a_1}\right)+\left(\frac{\partial\pi_1}{\partial a_2}+p\frac{\partial\pi_2}{\partial a_2}\right)\frac{da_2(a_1)}{da_1}+\frac{dp(a_1,a_2(a_1))}{da_1}\pi_2$$
$$= \frac{\partial\pi_1}{\partial a_1}+p\frac{\partial\pi_2}{\partial a_1}=0$$

In the first line, the third term is zero by hypothesis, and the second term is zero using the envelope theorem as above. The third equality follows from an analogous envelope observation:

for fixed $p = p(a_1, a_2(a_1))$, FM's action a_1 maximizes income since $(a_1, a_2(a_1))$ is MPE. Since the income effect is zero, FM's first-order condition is clearly satisfied: $\frac{d}{da_1} U_1(\widetilde{\pi}_1(p, I), \widetilde{\pi}_2(p, I)) = \frac{\partial U_1}{\partial \pi_1} \frac{d}{da_1} \widetilde{\pi}_1(p, I) + \frac{\partial U_1}{\partial \pi_2} \frac{d}{da_1} \widetilde{\pi}_2(p, I) = 0$.

□

Theorem 3. *Suppose U_2 is joint-monotonic, quasi-concave, and normal. Suppose the material payoff functions are globally conditionally transferable. If U_1 is monotonic or purely self-regarding, and if $U_1\left(\boldsymbol{\pi}\left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right)\right) \geq 0$, then the unique equilibrium transaction is the efficient transaction $\left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right)$.*

Proof: Since the material payoff functions are globally conditionally transferable, the budget curves are all parallel lines with slope $-p \equiv -\left.\frac{d\pi_1}{d\pi_2}\right|_{B(a_1)} = \frac{\partial \pi_2(a_1,a_2)/\partial a_2}{\partial \pi_1(a_1,a_2)/\partial a_2} = -k$ for some $k > 0$. Because U_2 is normal, SM's best-response function $a_2(a_1)$ ensures that π_1 and π_2 are both strictly increasing in $I(a_1)$. Since U_1 is monotonic or purely self-regarding, FM maximizes her utility by taking the action $\widetilde{a}_1$ that maximizes $I(a_1)$. This is the action $\widetilde{a}_1 = \overline{\overline{a}}_1$ that induces SM's favorite transaction because that is the unique action that induces an MPE transaction (by Lemma 2). Since $U_1\left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right) \geq 0$, this action gives FM at least as high utility as her outside option and is therefore the unique equilibrium.

□

Theorem 4. *Suppose U_2 is joint-monotonic, quasi-concave, and fairness-kinked. Assume S1-S5. If U_1 is monotonic or purely self-regarding, if $\left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right)$ is a strict fairness-rule optimum, and if $U_1\left(\boldsymbol{\pi}\left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right)\right) \geq 0$, then the unique equilibrium transaction is the efficient transaction $\left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right)$.*

Proof: We first show that $(\widehat{a}_1, \widehat{a}_2)$ exists and is the unique transaction satisfying $\pi_1(\widehat{a}_1, \widehat{a}_2) = \pi_1\left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right)$, $U_2(\boldsymbol{\pi}(\widehat{a}_1, \widehat{a}_2)) = 0$, and $\widehat{a}_1 < \overline{\overline{a}}_1$. Given A1, A3, and A4, clearly there is a unique material payoff pair on SM's $U_2 = 0$ indifference curve such that $\pi_1 = \pi_1\left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right)$, so $(\widehat{a}_1, \widehat{a}_2)$ exists. Call that material payoff pair $\left(\overline{\overline{\pi}}_1, \widehat{\pi}_2\right)$. Clearly $\widehat{\pi}_2 < \overline{\overline{\pi}}_2$ (since all feasible material payoff pairs $(\pi_1, \pi_2) \neq \left(\overline{\overline{\pi}}_1, \overline{\overline{\pi}}_2\right)$ with $\pi_1 = \overline{\overline{\pi}}_1$ have $\pi_2 < \overline{\overline{\pi}}_2$). In the remainder of this paragraph, we show that $(\widehat{a}_1, \widehat{a}_2)$ is unique and satisfies $(\widehat{a}_1, \widehat{a}_2) \ll \left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right)$. Define $\widetilde{a}_2(a_1)$ implicitly by $\pi_1(a_1, \widetilde{a}_2(a_1)) = \overline{\overline{\pi}}_1$, which is a continuously differentiable, strictly increasing function (by A1): $\frac{d\widetilde{a}_2(a_1)}{da_1} = -\frac{\partial \pi_1/\partial a_1}{\partial \pi_1/\partial a_2} > 0$.

It is also weakly convex:

$$\frac{d^2\widetilde{a}_2(a_1)}{d(a_1)^2} = \frac{-\frac{\partial^2\pi_1(a_1,\widetilde{a}_2(a_1))}{\partial(a_1)^2}\frac{\partial\pi_1}{\partial a_2} - \frac{\partial^2\pi_1(a_1,\widetilde{a}_2(a_1))}{\partial a_1\partial a_2}\frac{d\widetilde{a}_2(a_1)}{da_1}\frac{\partial\pi_1}{\partial a_2} + \frac{\partial\pi_1}{\partial a_1}\left(\frac{\partial^2\pi_1(a_1,\widetilde{a}_2(a_1))}{\partial a_1\partial a_2} + \frac{\partial^2\pi_1}{\partial(a_2)^2}\frac{d\widetilde{a}_2(a_1)}{da_1}\right)}{\left(\frac{\partial\pi_1}{\partial a_2}\right)^2}$$

$$= \frac{-\frac{\partial^2\pi_1}{\partial(a_1)^2}\frac{\partial\pi_1}{\partial a_2} + \frac{\partial\pi_1}{\partial a_1}\frac{\partial^2\pi_1}{\partial(a_2)^2}\frac{d\widetilde{a}_2(a_1)}{da_1}}{\left(\frac{\partial\pi_1}{\partial a_2}\right)^2} \geq 0,$$

where the second equality follows from substituting $\frac{d\widetilde{a}_2(a_1)}{da_1} = -\frac{\partial\pi_1/\partial a_1}{\partial\pi_1/\partial a_2}$, and the inequality follows from A1, $\frac{\partial^2\pi_1}{\partial(a_1)^2} \leq 0$ (due to A3), and $\frac{d\widetilde{a}_2(a_1)}{da_1} > 0$. Define $\widetilde{\widetilde{a}}_2(a_1,\pi_2)$ implicitly by $\pi_2(a_1,\widetilde{\widetilde{a}}_2) = \pi_2$, which is a continuously differentiable function, strictly increasing in a_1, strictly decreasing in π_2, and (due to A3) weakly concave in a_1. By A3, we also know that $\widetilde{a}_2(a_1)$ is strictly convex or $\widetilde{\widetilde{a}}_2(a_1,\pi_2)$ is strictly concave in a_1 (or both). From Theorem 1, we know there exists a unique a_1 such that $\widetilde{a}_2(a_1) = \widetilde{\widetilde{a}}_2\left(a_1,\overline{\overline{\pi}}_2\right)$, which is $\overline{\overline{a}}_1$. In the (a_1,a_2) plane, draw the graph of $\widetilde{a}_2(a_1)$ as an increasing, convex curve and the graph of $\widetilde{\widetilde{a}}_2\left(a_1,\overline{\overline{\pi}}_2\right)$ as an increasing, concave curve. These curves are tangent at $\overline{\overline{a}}_1$. Since $\widehat{\pi}_2 < \overline{\overline{\pi}}_2$ and $\widetilde{\widetilde{a}}_2(a_1,\pi_2)$ is decreasing in π_2, draw the graph of $\widetilde{\widetilde{a}}_2(a_1,\widehat{\pi}_2)$ as an upward shift of the graph of $\widetilde{\widetilde{a}}_2\left(a_1,\overline{\overline{\pi}}_2\right)$. There are two intersections of the graphs of $\widetilde{a}_2(a_1)$ and $\widetilde{\widetilde{a}}_2(a_1,\widehat{\pi}_2)$, one with $(a_1,a_2) \gg \left(\overline{\overline{a}}_1,\overline{\overline{a}}_2\right)$ and one with $(a_1,a_2) \ll \left(\overline{\overline{a}}_1,\overline{\overline{a}}_2\right)$. The latter is $(\widehat{a}_1,\widehat{a}_2)$.

We next show that $\overline{\overline{a}}_1$ is a local optimum for FM. By hypothesis, $\left(\overline{\overline{a}}_1,\overline{\overline{a}}_2\right)$ is a strict fairness-rule optimum. Therefore, Part 1 of Proposition 3 implies that $(a_1,a_2(a_1))$ is a strict fairness-rule optimum for all a_1 in a neighborhood of $\overline{\overline{a}}_1$. It follows that $\overline{\overline{a}}_1$ is a *local* optimum for FM, regardless of whether her distributional preferences are purely self-regarding or monotonic.

In the remainder of the proof, we show that $\overline{\overline{a}}_1$ is a *global* optimum for FM. To avoid cluttering notation with limits, we define the function U_2^D, which fully characterizes SM's preferences in the region of disadvantageous unfairness but is everywhere twice-continuously differentiable: $U_2^D(\boldsymbol{\pi}) \equiv U_2(\boldsymbol{\pi})$ for all $\boldsymbol{\pi} \in D_f$, and $\left(\frac{\partial U_2^D(\boldsymbol{\pi})}{\partial\pi_1}, \frac{\partial U_2^D(\boldsymbol{\pi})}{\partial\pi_2}\right) = \lim_{\boldsymbol{\pi}'\to\boldsymbol{\pi},\boldsymbol{\pi}'\in D_f}\left(\frac{\partial U_2(\boldsymbol{\pi}')}{\partial\pi_1}, \frac{\partial U_2(\boldsymbol{\pi}')}{\partial\pi_2}\right)$ for all $\boldsymbol{\pi} \in \text{graph}(f)$. (We do not constrain U_2^D in the region of advantageous unfairness because we will not use it there.) Now, it will be helpful in what follows to prove a preparatory claim.

Preparatory claim: We claim that

$$\frac{\partial U_2^D}{\partial\pi_2} - p(\overline{\overline{a}}_1,\widehat{a}_2)\frac{\partial U_2^D}{\partial\pi_1} > 0$$

at *all* individually-rational transactions (a_1,a_2) such that $\pi_1(a_1,a_2) > \pi_1\left(\overline{\overline{a}}_1,\overline{\overline{a}}_2\right)$.

To prove it, suppose to the contrary there were some $\boldsymbol{\pi}(a_1,a_2)$ at which $\frac{\partial U_2^D}{\partial\pi_2} - p(\overline{\overline{a}}_1,\widehat{a}_2)\frac{\partial U_2^D}{\partial\pi_1} \leq 0$; we will first show that $\frac{\partial U_2^D}{\partial\pi_1} \geq 0$. There are two cases. When $\frac{\partial U_2^D}{\partial\pi_2} - p(\overline{\overline{a}}_1,\widehat{a}_2)\frac{\partial U_2^D}{\partial\pi_1} = 0$, $\frac{\partial U_2^D}{\partial\pi_1}$ and $\frac{\partial U_2^D}{\partial\pi_2}$

have the same sign since $p(\overline{\overline{a}}_1, \widehat{a}_2) > 0$, and so $\frac{\partial U_2^D}{\partial \pi_1} \geq 0$ there (else joint-monotonicity is violated). And when $\frac{\partial U_2^D}{\partial \pi_2} - p(\overline{\overline{a}}_1, \widehat{a}_2)\frac{\partial U_2^D}{\partial \pi_1} < 0$, we must again have $\frac{\partial U_2^D}{\partial \pi_1} \geq 0$ (else $\frac{\partial U_2^D}{\partial \pi_2} < 0$, violating joint-monotonicity). Now that we have established that $\frac{\partial U_2^D}{\partial \pi_1} \geq 0$, we know that by choosing a value k slightly larger than $p(\overline{\overline{a}}_1, \widehat{a}_2)$, we must have

$$\frac{\partial U_2^D}{\partial \pi_2} - k\frac{\partial U_2^D}{\partial \pi_1} < 0$$

at $\boldsymbol{\pi}(a_1, a_2)$. Since k is very close to $p(\overline{\overline{a}}_1, \widehat{a}_2)$, using S1, we also know that

$$\frac{\partial U_2^D}{\partial \pi_2} - k\frac{\partial U_2^D}{\partial \pi_1} > 0$$

at $\boldsymbol{\pi}\left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right)$. Drawing budget lines l, l' each with slope $-k$ passing through the two points $\boldsymbol{\pi}\left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right)$ and $\boldsymbol{\pi}(a_1, a_2)$, respectively, the above inequalities imply that SM's most-preferred point on l is below $\pi_1\left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right)$ and his most-preferred point on l' is above $\pi_1(a_1, a_2)$. By assumption, $\pi_1(a_1, a_2) > \pi_1\left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right)$. Since $\boldsymbol{\pi}\left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right)$ lies on the MPE frontier, which is downward sloping and concave, l is to the right of l'. So S3 (the normality assumption) is violated; a contradiction. This proves the Preparatory Claim.

We will prove that $\overline{\overline{a}}_1$ is the global optimum for FM in two cases, but before proceeding, we note three useful facts.

First, at any individually-rational transaction such that $\pi_1(a_1, a_2) = \pi_1(\widehat{a}_1, \widehat{a}_2)$ and $\pi_2(a_1, a_2) > \pi_2(\widehat{a}_1, \widehat{a}_2)$ we must have $(a_1, a_2) \gg (\widehat{a}_1, \widehat{a}_2)$. Suppose not. In that case, since A1 rules out $a_1 \geq \widehat{a}_1$ and $a_2 \leq \widehat{a}_2$ or vice-versa, it must be that $(a_1, a_2) \ll (\widehat{a}_1, \widehat{a}_2)$. Assuming for now that $\left.\frac{da_2}{da_1}\right|_{\pi_1=\pi_1(\widehat{a}_1,\widehat{a}_2)} < \left.\frac{da_2}{da_1}\right|_{\pi_2=\pi_2(\widehat{a}_1,\widehat{a}_2)}$, then by A1 and weak concavity of π_2 (from A3), $\pi_2(a_1, a_2) < \pi_2(\widehat{a}_1, \widehat{a}_2)$; a contradiction. We now show that $\left.\frac{da_2}{da_1}\right|_{\pi_1=\pi_1(\widehat{a}_1,\widehat{a}_2)} < \left.\frac{da_2}{da_1}\right|_{\pi_2=\pi_2(\widehat{a}_1,\widehat{a}_2)}$. Recall from the argument in the first paragraph of this proof that $(\widehat{a}_1, \widehat{a}_2)$ is the unique intersection of the graphs of $\widetilde{a}_2(a_1)$ and $\widetilde{\widetilde{a}}_2(a_1, \widehat{\pi}_2)$ such that $(\widehat{a}_1, \widehat{a}_2) \ll \left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right)$. Since the graphs of $\widetilde{a}_2(a_1)$ and $\widetilde{\widetilde{a}}_2\left(a_1, \overline{\overline{\pi}}_2\right)$ are tangent at $\overline{\overline{a}}_1$, $\frac{d\widetilde{a}_2(\overline{\overline{a}}_1)}{da_1} = \frac{\partial \widetilde{\widetilde{a}}_2(\overline{\overline{a}}_1, \overline{\overline{\pi}}_2)}{\partial a_1}$. Since $\widetilde{a}_2(a_1)$ is increasing and convex, $\frac{d\widetilde{a}_2(\widehat{a}_1)}{da_1} < \frac{d\widetilde{a}_2(\overline{\overline{a}}_1)}{da_1}$. Due to S2 (in particular, SM's material payoff function being additively separable), $\widetilde{\widetilde{a}}_2(a_1, \pi_2)$ is additively separable, and since it is also increasing and concave in a_1, $\frac{\partial \widetilde{\widetilde{a}}_2(\widehat{a}_1, \pi_2(\widehat{a}_1, \widehat{a}_2))}{\partial a_1} = \frac{\partial \widetilde{\widetilde{a}}_2(\widehat{a}_1, \overline{\overline{\pi}}_2)}{\partial a_1} > \frac{\partial \widetilde{\widetilde{a}}_2(\overline{\overline{a}}_1, \overline{\overline{\pi}}_2)}{\partial a_1}$. Combining these observations and noting that $\frac{d\widetilde{a}_2(\widehat{a}_1)}{da_1} = \left.\frac{da_2}{da_1}\right|_{\pi_1=\pi_1(\widehat{a}_1,\widehat{a}_2)}$ and $\frac{\partial \widetilde{\widetilde{a}}_2(\widehat{a}_1, \pi_2(\widehat{a}_1, \widehat{a}_2))}{\partial a_1} = \left.\frac{da_2}{da_1}\right|_{\pi_2=\pi_2(\widehat{a}_1,\widehat{a}_2)}$, the needed inequality follows.

Second, due to S2 (the material payoff functions being additively separable), the slope of any budget curve, $p(a_1, a_2)$, does not depend on a_1. Therefore, $\frac{\partial p(a_1,a_2)}{\partial a_1} = 0$, and it thus follows from the Technical Lemma that $\frac{\partial p(a_1,a_2)}{\partial a_2} \leq 0$.

Third, since $\boldsymbol{\pi}\left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right) \in \text{graph}(f)$, and since SM's fairness rule is strictly increasing, any (π_1, π_2) with $\pi_1 \geq \pi_1\left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right)$ is in the region of disadvantageous unfairness, and thus we can use U_2^D.

Case 1: FM is purely self-regarding. To prove that $\overline{\overline{a}}_1$ is the unique global optimum for FM, it is sufficient to show that there does not exist any individually-rational transaction $(a_1, a_2(a_1)) \neq \left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right)$ that satisfies $\pi_1(a_1, a_2(a_1)) \geq \pi_1\left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right)$. Suppose to the contrary that there exists an individually-rational transaction $(a_1', a_2(a_1')) \neq \left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right)$ such that $\pi_1(a_1', a_2(a_1')) \geq \pi_1\left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right)$.

We first show that without loss of generality, we can assume that $(a_1', a_2(a_1')) \gg \left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right)$. By A1, the only other possibility is $(a_1', a_2(a_1')) \ll \left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right)$. But in that case, there exists $(a_1'', a_2'') \gg \left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right)$ such that $\boldsymbol{\pi}(a_1'', a_2'') = \boldsymbol{\pi}(a_1', a_2(a_1'))$ (this follows from an argument similar to that in the first paragraph of this proof). Since $a_2'' > a_2(a_1')$ and $\frac{\partial p(a_1,a_2)}{\partial a_2} \leq 0$, the change from the budget line through $(a_1', a_2(a_1'))$ to the budget line through (a_1'', a_2'') is a Slutsky-compensated decrease in the price of FM's material payoff, and thus SM chooses a higher material payoff for FM: $\pi_1(a_1'', a_2(a_1'')) \geq \pi_1(a_1'', a_2'')$. Hence in the remainder of the proof, we can simply use $(a_1'', a_2(a_1''))$ in place of $(a_1', a_2(a_1'))$ and relabel it as $(a_1', a_2(a_1'))$.

Because $\boldsymbol{\pi}(a_1', a_2(a_1'))$ lies strictly in the interior of the region of disadvantageous unfairness and $a_2(a_1')$ is a best response, $\frac{\partial U_2^D}{\partial \pi_2} - p(a_1', a_2(a_1'))\frac{\partial U_2^D}{\partial \pi_1} = 0$ at $\boldsymbol{\pi}(a_1', a_2(a_1'))$. We now show that $\pi_1(a_1', a_2(a_1')) = \pi_1\left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right)$ leads to a contradiction (and then turn in the next paragraph to the case $\pi_1(a_1', a_2(a_1')) > \pi_1\left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right)$). Clearly $\pi_2(a_1', a_2(a_1')) < \pi_2\left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right)$. Now consider $(a_1''', a_2''') \gg (a_1', a_2(a_1'))$ such that $\pi_1(a_1''', a_2''') = \pi_1(a_1', a_2(a_1'))$, $\pi_2(a_1', a_2(a_1')) < \pi_2(a_1''', a_2''') < \pi_2\left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right)$ (the existence of such a transaction follows from an argument similar to that in the first paragraph of this proof). We know that

$$
\begin{aligned}
& \frac{\partial U_2^D\left(\boldsymbol{\pi}(a_1', a_2(a_1'))\right)}{\partial \pi_2} - p(a_1', a_2(a_1'))\frac{\partial U_2^D\left(\boldsymbol{\pi}(a_1', a_2(a_1'))\right)}{\partial \pi_1} \\
> \quad & \frac{\partial U_2^D\left(\boldsymbol{\pi}(a_1''', a_2''')\right)}{\partial \pi_2} - p(a_1', a_2(a_1'))\frac{\partial U_2^D\left(\boldsymbol{\pi}(a_1''', a_2''')\right)}{\partial \pi_1} \\
\geq \quad & \frac{\partial U_2^D\left(\boldsymbol{\pi}(a_1''', a_2''')\right)}{\partial \pi_2} - p(\overline{\overline{a}}_1, \widehat{a}_2)\frac{\partial U_2^D\left(\boldsymbol{\pi}(a_1''', a_2''')\right)}{\partial \pi_1},
\end{aligned}
$$

where the first inequality follows from S3 (the normality of U_2), and the second inequality follows from $a_2(a_1') > \widehat{a}_2$, $\frac{\partial p(a_1,a_2)}{\partial a_2} \leq 0$, and $\frac{\partial p(a_1,a_2)}{\partial a_1} = 0$. Therefore,

$$
\frac{\partial U_2^D\left(\boldsymbol{\pi}(a_1''', a_2''')\right)}{\partial \pi_2} - p(\overline{\overline{a}}_1, \widehat{a}_2)\frac{\partial U_2^D\left(\boldsymbol{\pi}(a_1''', a_2''')\right)}{\partial \pi_1} < 0,
$$

but this is a contradiction because the Preparatory Claim implies that the left-hand side is ≥ 0.

We now show that $\pi_1(a_1', a_2(a_1')) > \pi_1\left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right)$ also leads to a contradiction. By A1, there is a unique transaction a_2' that satisfies $\pi_1(a_1', a_2') = \pi_1\left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right)$, where $a_2' < a_2(a_1')$. Draw budget lines

m, m' with respective slopes $-p(a_1', a_2')$ and $-p(a_1', a_2(a_1'))$ passing through the two points $\boldsymbol{\pi}(a_1', a_2')$ and $\boldsymbol{\pi}(a_1', a_2(a_1'))$. We know that at $\boldsymbol{\pi}(a_1', a_2')$, $\frac{\partial U_2^D}{\partial \pi_2} - p(a_1', a_2') \frac{\partial U_2^D}{\partial \pi_1} \geq \frac{\partial U_2^D}{\partial \pi_2} - p(\overline{\overline{a}}_1, \widehat{a}_2) \frac{\partial U_2^D}{\partial \pi_1} \geq 0$, where the first inequality follows from $a_2' > \widehat{a}_2$, $\frac{\partial p(a_1,a_2)}{\partial a_2} \leq 0$, and $\frac{\partial p(a_1,a_2)}{\partial a_1} = 0$, and the second inequality follows from the Preparatory Claim. Therefore, SM's most-preferred point on line m yields a material payoff for FM that is weakly smaller than $\pi_1(a_1', a_2')$. By construction, SM's most-preferred point on line m' is $\boldsymbol{\pi}(a_1', a_2(a_1'))$. Now, draw a third line m'' with slope $-p(a_1', a_2') \leq -p(a_1', a_2(a_1'))$ going through $\boldsymbol{\pi}(a_1', a_2(a_1'))$. Since moving from m' to m'' can be thought of as a Slutsky-compensated price change, SM's most-preferred point on line m'' must yield a material payoff for FM that is at least as large as $\pi_1(a_1', a_2')$. But comparing FM's material payoff when moving from m'' to m reveals a violation of U_2 being normal in π_1 (the assumption S3); a contradiction.

Case 2: FM's distributional preferences are strictly monotonic, and S4 and S5 hold. We claim that there is no $a_1' \neq \overline{\overline{a}}_1$ such that $U_1(a_1', a_2(a_1')) \geq U_1\left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right)$. We showed in Case 1 that there is no $a_1' \neq \overline{\overline{a}}_1$ such that $\pi_1(a_1', a_2(a_1')) \geq \pi_1\left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right)$. The result then follows from the observation that, since U_1 is monotonic, $\pi_1(\overline{a}_1, \overline{a}_2) > \pi_1\left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right)$ (from S4), and U_1 is weakly quasi-concave (from S5), the region enclosed by the upper-contour set of FM's $U_1 = U_1\left(\boldsymbol{\pi}\left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right)\right)$ indifference curve and MPE frontier contains only material payoff pairs satisfying $\pi_1(a_1, a_2) > \pi_1\left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right)$. This completes the proof.

□

Theorem A1. *Suppose U_1 and U_2 are joint-monotonic and quasi-concave. FM's and SM's favorite transactions, $(\overline{a}_1, \overline{a}_2)$ and $\left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right)$, exist and are unique. The set of UPE material payoff pairs is a connected set that includes $(\overline{\pi}_1, \overline{\pi}_2)$ and $\left(\overline{\overline{\pi}}_1, \overline{\overline{\pi}}_2\right)$ and lies within the region enclosed by $\overline{\overline{IC}}_1$, $\overline{IC}_2$, and the MPE frontier.*

Proof: The proofs that $(\overline{a}_1, \overline{a}_2)$ and $\left(\overline{\overline{a}}_1, \overline{\overline{a}}_2\right)$ exist, are unique, and are UPE are the same as in the proof of Theorem 1.

To see that the set of UPE material pairs is a connected set, consider the problem $\boldsymbol{\pi}\left(\overline{U}_2\right) \in \arg\max_{\{\boldsymbol{\pi}:\boldsymbol{\pi}\in T_\pi, U_2(\boldsymbol{\pi})=\overline{U}_2\}} U_1(\boldsymbol{\pi})$. The Maximum Theorem (e.g., Sundaram 1996, p.235) implies that $\boldsymbol{\pi}\left(\overline{U}_2\right)$ is an upper-hemicontinuous correspondence. It follows that $\left\{\boldsymbol{\pi}\left(\overline{U}_2\right)\right\}_{\overline{U}_2 \in \left[0, U_2\left(\overline{\overline{\pi}}_1, \overline{\overline{\pi}}_2\right)\right]}$ is a connected set. But $\left\{\boldsymbol{\pi}\left(\overline{U}_2\right)\right\}_{\overline{U}_2 \in \left[0, U_2\left(\overline{\overline{\pi}}_1, \overline{\overline{\pi}}_2\right)\right]}$ is exactly the set of UPE material payoff pairs.

There does not exist a UPE material payoff pair $(\widehat{\pi}_1, \widehat{\pi}_2)$ outside of the region enclosed by $\overline{\overline{IC}}_1$, $\overline{IC}_2$, and the MPE frontier because by construction $(\widehat{\pi}_1, \widehat{\pi}_2)$ is worse than $(\overline{\pi}_1, \overline{\pi}_2)$ or $\left(\overline{\overline{\pi}}_1, \overline{\overline{\pi}}_2\right)$ for *both* FM and SM.

$\square$

Theorem A2. *Suppose U_1 and U_2 are joint-monotonic and quasi-concave, and both are either twice-continuously differentiable or fairness-kinked. If the equilibrium $(a_1, a_2(a_1))$ is UPE and not MPE, then $(a_1, a_2(a_1))$ is a fairness-rule optimum for SM. If, in addition, $(a_1, a_2(a_1))$ is a strict fairness-rule optimum for SM and any fairness rule is continuously differentiable, then at least one of the following must be true: $\boldsymbol{\pi}(a_1', a_2(a_1'))$*

1. *SM's indifference curve for disadvantageously unfair transactions is tangent to SM's fairness rule at $\boldsymbol{\pi}(a_1, a_2(a_1))$.*

2. *U_1 is fairness-kinked, $\boldsymbol{\pi}(a_1, a_2(a_1))$ is on FM's fairness rule, and the respective fairness rules f_1 and f_2 have different slopes at $\boldsymbol{\pi}(a_1, a_2(a_1))$.*

Proof: We begin with the first claim: if the equilibrium $(a_1, a_2(a_1))$ is UPE and not MPE, then $(a_1, a_2(a_1))$ is a fairness-rule optimum for SM. We will prove that if the equilibrium $(a_1, a_2(a_1))$ is UPE and if U_2 is continuously differentiable at $\boldsymbol{\pi}(a_1, a_2(a_1))$, then $\boldsymbol{\pi}(a_1, a_2(a_1))$ is also MPE. Suppose not. Then $\boldsymbol{\pi}(a_1, a_2(a_1))$ is in the interior of the materially-feasible set. Lemma 1 implies that SM's distributional preferences are (locally) monotonic in a neighborhood of $\boldsymbol{\pi}(a_1, a_2(a_1))$. Since FM's distributional preferences are joint-monotonic, there is an alternative material payoff pair giving higher material payoff to both players that both players prefer. This contradicts UPE.

From now on, we assume that the equilibrium $(a_1, a_2(a_1))$ is UPE, is not MPE, and is a strict fairness-rule optimum for SM, and we assume that any fairness rule is continuously differentiable.

We next show that if U_1 is continuously differentiable at $\boldsymbol{\pi}(a_1, a_2(a_1))$, then SM's indifference curve for disadvantagously unfair transactions is tangent to SM's fairness rule at $\boldsymbol{\pi}(a_1, a_2(a_1))$. For contradiction, suppose that SM's indifference curve for disadvantagously-unfair transactions is not tangent to SM's fairness rule at $\boldsymbol{\pi}(a_1, a_2(a_1))$: $\left.\frac{d\pi_1}{d\pi_2}\right|_{U_2^D = U_2^D(\boldsymbol{\pi}(a_1, a_2(a_1)))} \neq f'(\pi_2)$. By a similar argument to that in the previous paragraph, SM's distributional preferences cannot be locally monotonic; therefore, SM's interpersonal indifference curve at $\boldsymbol{\pi}(a_1, a_2(a_1))$ is upward-sloping. Since the indifference curve also lies in the region of disadvantageous unfairness, it must be that $\left.\frac{d\pi_1}{d\pi_2}\right|_{U_2^D = U_2^D(\boldsymbol{\pi}(a_1, a_2(a_1)))} > f'(\pi_2)$. Since $\boldsymbol{\pi}(a_1, a_2(a_1))$ is not MPE, we know that it is in the interior of the materially-feasible set, and this, together with $\boldsymbol{\pi}(a_1, a_2(a_1))$ being UPE, implies that $\left.\frac{d\pi_1}{d\pi_2}\right|_{U_1 = U_1(\boldsymbol{\pi}(a_1, a_2(a_1)))} \geq \left.\frac{d\pi_1}{d\pi_2}\right|_{U_2^D = U_2^D(\boldsymbol{\pi}(a_1, a_2(a_1)))}$. Since $(a_1, a_2(a_1))$ is a strict fairness-rule optimum for SM, Part 1 of Proposition 3 implies that there exists a slight deviation for FM such that SM's optimal response would yield a material-payoff pair slightly southwest on SM's fairness rule. But the above inequalities imply that $\left.\frac{d\pi_1}{d\pi_2}\right|_{U_1=U_1(\boldsymbol{\pi}(a_1,a_2(a_1)))} > f'(\pi_2)$ at $\boldsymbol{\pi}(a_1, a_2(a_1))$, meaning that FM would prefer this alternative material-payoff pair, contradicting that $(a_1, a_2(a_1))$ is an equilibrium.

Finally, we show that if U_1 is fairness-kinked and $\boldsymbol{\pi}(a_1, a_2(a_1))$ is on FM's fairness rule, then FM's and SM's respective fairness rules f_1 and f_2 have different slopes at $\boldsymbol{\pi}(a_1, a_2(a_1))$. For contradiction, suppose instead that f_1 and f_2 have the same slope at $\boldsymbol{\pi}(a_1, a_2(a_1))$. Since $(a_1, a_2(a_1))$ is a strict fairness-rule optimum for SM, Part 1 of Proposition 3 implies that there exists a slight deviation for FM such that SM's optimal response would yield a material-payoff pair slightly northeast along SM's fairness rule. We know that FM would prefer a sufficiently small northeast movement along SM's fairness rule, contradicting that $(a_1, a_2(a_1))$ is an equilibrium.

□

References

[1] Rangarajan K. Sundaram. *A First Course in Optimization Theory.* Cambridge University Press, Cambridge, UK, 1996.