Intentionality detection and ''mindreading'': Why does game form matter?

Kevin A. McCabe[†], Vernon L. Smith[†‡], and Michael LePore[§]

[†]Economic Science Laboratory, McClelland Hall 116, 1130 East Helen, P.O. Box 210108, University of Arizona, Tucson, AZ 85721-0108; and [§]Department of Political Science, University of Rochester, Rochester, NY 14627

Contributed by Vernon L. Smith, January 27, 2000

By around the age of 4 years, children ''can work out what people might know, think or believe'' based on what they say or do. This is called ''mindreading,'' which builds upon the human ability to infer the intentions of others. Game theory makes a strong assumption about what individual A can expect about B's intentions and *vice versa*, viz. that each is a self-interested opponent of the other and will reliably analyze games by using such basic principles as dominance and backward induction, and behave as if the normal form of an extensive form game is equivalent to the latter. But the extensive form allows intentions to be detected from actual sequential play and is therefore not necessarily equivalent psychologically to the normal form. We discuss Baron-Cohen's theory of the mindreading system [Baron-Cohen, S. (1995) *Mindblindness: An Essay on Autism and Theory of Mind* (MIT Press, Cambridge, MA)] to motivate the comparison of behavior in an extensive form game with its corresponding normal form. As in the work of Rapoport [Rapoport, A. (1997) *Int. J. Game Theory* 26, 113–136] and Schotter *et al.* [Schotter, A., Wiegelt, K. & Wilson, C. (1994) *Games Econ. Behav.* 6, 445–468], we find consistent differences in behavior between the normal and extensive forms. In particular, we observe attempts to cooperate, and in some treatments we observe the achievement of cooperation, occurring more frequently in the extensive form. Cooperation in this context requires reciprocity, which is more difficult to achieve by means of intentionality detection in the normal as opposed to the extensive form games we study.

Our objective is to study behavior in the normal form representation of one of two extensive form bargaining games previously examined by using a variety of matching protocols (1). In the current study we limit the matching protocols to single play and to repeat play with the same pairs, and we study comparisons of the normal and extensive forms.

Comparisons of behavior in the normal and extensive forms of various games have been made most notably in refs. 2 and 3. The latter's emphasis is on the rationality principles of iterated dominance and backward induction as factors in individual behavior in addition to whether behavior in a game is invariant to the form of its representation—all fundamental principles in game theory (4). Rapoport (2) and Schotter *et al*. (3) strongly rejected the invariance principle, but explication in terms of game theory was illusive: "where we expected our rationality principles would predict behavioral differences across game forms, either no such differences appeared or they were not what we expected." (ref. 3, pp. 446–447).

Rapoport (2) provides transparent examples of two extensive form versions of the "Battle-of-the Sexes" game and the same game in matrix normal (strategic) form. His examples make clear how order-of-play information provides a principle that can better coordinate player strategies in the extensive forms. That principle, as we would describe it, derives from the human capacity to read another person's thoughts or intentions by placing themselves in the position and information state of the other person. Because of the example's transparency and special character, his experiments are conducted by using a three-person resource dilemma, a public good, and a pure coordination game.

In our games we try to predict certain core features of the variation in behavior with the game form, but only by reaching outside of the traditional rationality principles of game theory to include concepts of reciprocity and "mindreading," from evolutionary psychology. Our ultimate goal is to provide an empirical foundation for modeling this behavior in terms of distributions of types of player—some of whom have a disposition toward noncooperative behavior, whereas others are disposed toward cooperation.

Our wellsprings are hardly new. The essence of our conceptual approach was stated 38 years ago: "A normative theory must produce strategies that are at least as good as what people can do without them. More, it must not deny or expunge details of the game that can demonstrably benefit two or more players and that the players, consequently, should not expunge or ignore in their mutual interest. . . A particular implication of this general point is that the game in 'normal' (mathematically abstract) form is not logically equivalent to the game in 'extensive' (particular) form, once we admit the logic by which rational players concert their expectations of each other." (ref. 5, pp. 98–99).

As so often in the history of ideas, almost no one was ready for this 38 years ago. What has transpired since is (*i*) an increasing discomfort with the relevance of the traditional game theoretic assumptions; (*ii*) new neuroscientific work on how the mind works (see ref. 6 for an accessible summary); (*iii*) growing recognition that mindreading is essential to understanding strategic interaction, and (*iv*) pioneering evolutionary insights into social exchange (7).

Five Principles of Behavior

We focus on five principles of self-interested behavior that appear to be needed to deal with observed bargaining behavior ranging from ultimatum and dictator games (e.g., see refs. 8 and 9, and the references therein) to structurally richer two-person games such as those in refs. 1 and 3 and in this paper.

The first two are the basic rationality principles of game theory—dominance and backward induction (4)—which, we argue, are relevant for many subjects and must therefore be retained in any extensions in theory. Such rationality need not be the result of conscious cognition. Thus, Baldwin and Meese (10) show that pigs will strategically interact in accordance with the dominance principle without presumed conscious "understanding" of the principle itself. We would argue that the opposite may apply: that the achievement of noncooperative equilibria in certain finitely repeated games, normally thought to require dominance and backward induction, are more likely in situations where agents cannot consciously apply the principles of game theoretic analysis. Thus, in a repeat play game with private payoff information, Brown (11) reports a steady buildup of support for a subgame perfect equilib-

Abbreviations: ID, intentionality detector; EDD, eye direction detector; SAM, a shared-attention mechanism; TOMM, theory-of-mind mechanism; SP, subgame perfect.

[‡]To whom reprint requests should be addressed. E-mail: smith@econlab.arizona.edu.

rium with repetition and subjects in the same role but random rematching, whereas complete payoff information yields a buildup of support for cooperation. In view of the results in refs. 3 and 2, the research reported in ref. 1, and the additional tests reported below for complete payoff information games, we strongly qualify the general principle that the solution of any game be invariant to representation in the extensive and normal forms for all players. This, we hypothesize, is because of the way the human brain naturally functions.

The third principle is the Folk Theorem—that repetition of a constituent game enables cooperation. But various forms of the Folk Theorem all predict many new equilibria without guidance as to how subjects will end up coordinating cooperative ones. The evidence supports the Folk Theorem principle, but the support is much stronger for the extensive than the normal form representation of the game as discussed below.

The fourth principle is reciprocity in which the long term self-interest is served by promoting a reputation in which cheating on cooperative social exchanges is punished (negative reciprocity), and the initiation of cooperative social exchanges is rewarded (positive reciprocity). The reciprocity principle, as we see it being implemented, implies that the normal and extensive forms of a game are inherently different, although for some subjects they might lead to the same outcome, if by different mental processes. Functionally, reciprocity, exchange, and the division of labor are cross-cultural universal characteristic of humans, although their institutional forms vary widely (ref. 12, pp. 137–138). Evolutionary psychologists argue that negative reciprocity is an adaptation in the evolution of our minds; a consequence of 2–3 million years of living in small hunter–gatherer bands in which cooperation was essential in sharing an uncertain harvest in a world with limited food storage and preservation technologies, and no monetary system (7). Consequently, an innate tendency to incur personal cost to punish cheaters on social exchange had high fitness value in promoting gains from exchange. A more complete discussion of the evolutionary psychology perspective and its implications for experimental economics is in ref. 13.

We do not claim that the reciprocity principle is invariant across all institutional circumstances, independent of incentives, as a universal law of behavior. We would expect the rules of interaction (institution) and opportunity cost to qualify reliance on reciprocity in circumstances in which the self-interest would be badly served. Our long-term research program is to better understand these nuances. Thus, in ref. 14 when pairs are matched anonymously in face-to-face Coase bargaining with an asymmetric outside option, 100% of the subjects ignored the opportunity cost of the outside option, $12, and split the pie, $14, equally. But in comparison samples, when the first mover earned the right to be endowed with the outside option, only 30% of the subjects violated individual rationality by splitting the pie equally. Furthermore, in ref. 15, when pairs are run in this constituent game in a 32-person tournament with the first round losers earning $5, the second round $10, the third $25, and fourth, fifth, and final round prizes of $70, $125, and $250, respectively, the incidence of equal split shares fell to only 4%.[¶] But observe that in the Hoffman–Spitzer institutional (nontournament) context, we still have 30% of the population of subjects failing to exhibit individual rationality in circumstances that provide no short-run incentives for being cooperative. Clearly, there exist important phenomena that cannot be comprehended within the traditional game theoretic modeling framework in some institutional contexts. The tournament institution provides sharply distinct rewards to a person or persons who achieve outcomes only slightly larger than their loser counterparts. Such a structure provides strong disincentives for achieving gains from personal exchange through reciprocity.

The fifth principle we will refer to as *intentionality detection*, and it is at the crux of our claim that the extensive and normal forms are psychologically different. People are good at "mindreading," defined as inferring the mental states of others from their words or actions. Thus, in an extensive form game, after player 1 (or 2) has made a choice between two moves, it is natural for player 2 (or 1) to ask (there is no presumption that this is conscious), "What does she intend?" or "What does she want me to think?" Such intentionality detection is difficult in the normal (matrix) form of simultaneous play in a decision tree, and even when repeated is more difficult to interpret in normal form where a strategy choice represents a complex of multiple moves in the corresponding "equivalent" extensive form. For this reason the normal and extensive forms need not be psychologically, and informationally, equivalent, as is evident from ref. 2. We will spell this out more clearly below in the context of the two constituent games we study. Also note that in a tournament institutional structure it is evident to all players what are the intentions of one's counterpart player, who is a foe, and must be treated as an opponent, not a partner. The effect of such a structure on expectations (through intentionality detection) may be much more important than the effect on incentives emphasized above.

Intentionality detection explains why Nash and subgame perfect outcomes in experiments are favored by private as compared with complete payoff information: cooperation becomes infeasible when neither of two bargainers knows the other's payoff and cannot therefore interpret moves in terms of reciprocity-driven intentions (ref. 11; also see ref. 16). Similar considerations apply to experiments that create ingroups or outgroups, or specially recognized "status" groups, which support differential behavior by facilitating the subconscious reading of intentions by bargaining pairs; subconscious, because the different groups are unaware of their differential behavior toward each other (17). Similar results are obtained in the study of natural populations thought to vary in ingroup strength (18).

From the evolutionary perspective, the human mind developed adaptations to its environment across countless generations of experience. That experience was conditioned by extensive form interactions with other humans, and with animals in

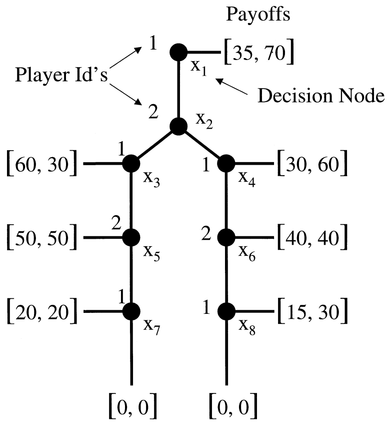

Fig. 1. In this extensive form game players 1 and 2 can, by alternating moves, end up at the outcome (50, 50). However, player 1 has an incentive to play left at decision node x_3, ending the game at (60, 30). Given this incentive, noncooperative game theory predicts player 2 will play right at x_2 and end up at the outcome (40, 40). The theory-of-mind hypothesis discussed in this paper predicts that player 1 will infer from player 2's move left at x_2 that player 2 is trying to reach the mutually beneficial (50, 50) on the left. From this informed inference about player 2's intentions, player 1 will move down at x_3.

[¶] It should be noted that game theory requires only standard reward protocols, not tournament rewards, to predict individually rational outcomes.

Table 1. Matrix normal form

	Player 1									
	1	2	3	4	5	6	7	8	9	10
Player 2	R****	DLR**	DLD*R	DLD*D	DDRL*	DDRD*	DDDLR	DDDLD	DDDDR	DDDDD
1 (Column, row) LL*	(35, 70)	(60, 30)	(60, 30)	(60, 30)	(50, 50)	(50, 50)	(50, 50)	(50, 50)	(50, 50)	(50, 50)
2 (Column, row) LD*	(35, 70)	(60, 30)	(60, 30)	(60, 30)	(20, 20)	(0, 0)	(20, 20)	(20, 20)	(0, 0)	(0, 0)
3 (Column, row) R*R	(35, 70)	(30, 60)	(40, 40)[†]	(40, 40)[†]	(30, 60)	(30, 60)	(40, 40)	(40, 40)	(40, 40)	(40, 40)
4 (Column, row) R*D	(35, 70)[†]	(30, 60)	(15, 30)	(0, 0)	(30, 60)	(30, 60)	(15, 30)	(0, 0)	(15, 30)	(0, 0)

[†] Nash equilibria.

hunting. The hypothesis follows that the resulting strategic reasoning algorithms, providing the adaptations that improved our fitness in the evolutionary environment, would be primed to operate in the extensive form format, and *not* in the normal form format so demonstrably convenient for abstract analysis. Indeed, in teaching game theory a chief pedagogic device is to use lots of extensive form examples, which, under the invariance postulate, we subsequently reduce to normal form. This is because our own intuition and that of the student is best served in the extensive form. (See ref. 19 for a similar argument for the frequency as against the probability format for Bayesian reasoning—in the former people are much better intuitive statisticians than in the latter.)

A Mental Anatomy of Mindreading

"By the end of the first year of life, normal infants, according to the evidence presented in the last chapter, can tell that they and someone else are attending to the same thing, and can read people's actions as directed at goals and as driven by desires. As toddlers, they can pretend and understand pretense. And by the time they begin school, around 4, they can work out what people might know, think and believe." (ref. 20, pp. 59–60).

Based in part on the pattern of evidence in people with mental process/sensor disorders such as autism and blindness, Baron-Cohen (20) has proposed four mental modules that constitute separate components of the human mindreading system: (*i*) an intentionality detector (ID), (*ii*) an eye direction detector (EDD), (*iii*) a shared-attention mechanism (SAM), and (*iv*) a theory-of-mind mechanism (TOMM).

As individuals interact in our experimental setting we hypothesize that they use their ID to generate dyadic models of their counterpart's intentions. For example, "he is acting in his self-interest," or "she is trying to cooperate." But shared attention requires individuals to form triadic expectations of the form "he knows that I am trying to cooperate and he knows that I know this." Using one's SAM requires additional information, which we hypothesize is found in recognizing play that leads to mutual gains supported by reciprocity.

The blind child who is otherwise normal lacks only EDD. The child, when told, "Make it so mommy can't see the car," responds by putting the toy in his pocket and nonchalantly relaxing his arms at his side. Such blind children are aware of mental phenomena in others, of what "seeing" is in sighted people, and will say, "See, it's in my lap." Since most bargaining experiments are conducted anonymously to control for "social effects," bargaining games using blind subjects would be predicted by this theory to yield results indistinguishable from those of sighted subjects. But face-to-face bargaining might be another matter if eye contact is important.

Table 2. Experimental design: Number of subjects/pairs/observations

	Game form	
Treatments	Extensive form	Matrix normal form
Single	52/26/26	48/24/24
Same	46/23/460	48/24/480

Children with autism fall into two groups: (*i*) those who lack both SAM and TOMM, and (*ii*) those for whom only TOMM is impaired. Such children are unable to understand pretense and cannot understand that someone might hold a false belief. But their ID and EDD capabilities are intact. Autistic adults partially overcome their handicap by simply learning to associate (memorize) certain reactions in others to certain stimuli, and to adjust accordingly. They carry in their minds vast libraries of such "how to behave" prescriptions. But they are abnormal in understanding allusions, innuendo, metaphors, irony, and jokes (20). Autism occurs in twins and families as one would expect of a genetic disorder, and quite disproportionately affects males relative to females.

The Constituent Game: Reciprocity Interpretations

The Extensive Form Game: Sequential Play. The constituent game we study is shown in Fig. 1.[‖] A round of play begins with player 1 choosing between the outside option, right at x_1, yielding (35, 70) for (player 1, player 2), or down. Note that a move right at x_1 is predicted if player 1 is altruistic, and obtains satisfaction from giving money to player 2 at low cost to player 1. If the move is down, then player 2 chooses between right branch play and left branch play at node x_2. A round continues until a move terminates the game at a payoff outcome. The right branch of the game tree contains the unique subgame perfect (SP) equilibrium at (40, 40), achieved by applying backward induction and eliminating moves that end in dominated outcomes for each player. A better outcome is the symmetric joint maximum at (50, 50) on the left, but without cooperation it cannot be achieved. Thus, assume that both players are myopically self-interested and each believes that the other is myopically self-interested. Then if player 2 moves left at x_2, player 1 will move left at x_3, yielding (60, 30), which is better for player 1 than (50, 50), the self-interested choice of player 2 if player 1 moves down at x_3. Thus, it is in player 1's interest to defect from cooperation at node x_3. Player 2, using backward induction, should therefore conclude that her best move at node x_2 is right, because (40, 40) is better than (60, 30) for player 2. Consequently, for a single round of play through game 1, noncooperative theory predicts down at x_1, right at x_2, down at x_4 and right at x_6, ending at (40, 40). Notice in particular that this standard game theoretic analysis hardwires each player's intentions into the thinking of the other. This makes the behavior of each player known to the other and in essence each player is in a game of certainty against nature, which is effectively

[‖] The game in Fig. 1 is referred to as Game 2 in ref. 1, where it is studied in extensive form under alternative matching protocols and compared with Game 1. The latter has the form shown in Fig. 1 except that the payoffs available at x_3 and x_5 are reversed. Consequently, in that game player 2 can either accept defection by moving left at x_5 for the outcome (60, 30) or down to punish the defection.

Table 3. Summary, extensive and normal form treatment results, branch conditional outcome frequencies

Game; form	(35, 70)	Left branch	(50, 50)	(60, 30)	(20, 20)	Right branch	(30, 60)	(40, 40)	(15, 30)	$E(\pi_2 \mid \text{Left})$
Single; extensive	0	12/26 = 0.46	6/12 = 0.50	6/12 = 0.50	0	14/26 = 0.54	0	14/14 = 1.00	0	40.0
Single; normal	0	7/24 = 0.29	1/7 = 0.14	6/7 = 0.86	0	17/24 = 0.71	3/17 = 0.18	14/17 = 0.82	0	32.9
Same; extensive	37/460 = 0.08	261/423 = 0.62	220/261 = 0.84	41/261 = 0.16	0	162/423 = 0.38	27/162 = 0.17	114/162 = 0.70	21/162 = 0.13	46.9
Same; normal	30/480 = 0.06	215/450 = 0.48	155/215 = 0.72	56/215 = 0.26	4/215 = 0.02	234/450 = 0.52	48/234 = 0.20	142/234 = 0.61	23/234 = 0.10	35.0

$E(\pi_2|\text{Left})$ reports the expected profit to player 2 of moving left at x_2, using the observed branch conditional outcome frequencies.

a robot. A similar argument explains why tournaments are so effective in inducing myopically self-regarding behavior, although this explanation is confounded with tournament incentive effects.

Reciprocity and intentionality detection alter the above analysis. If player 1 moves down at node x_1, player 2 can reasonably infer that player 1 expects to attain an outcome better than 35, which includes at least the "sure thing," 40, SP on the right. Also, player 2 knows that player 1 is not an altruist because right at x_1 costs player 1 only 5 relative to the SP outcome. Player 2, by responding with a left move at x_2, can be interpreted as saying, "Look, we can both do better at (50, 50.)" If player 2 is disposed to a cooperative stance toward people who she believes are not clearly foes (as in a tournament), but "another like person," then left at x_2 is a means of signaling an attempt to achieve (50, 50) by activating player 1's ID. Hence, player 2 is invoking her SAM and TOMM. The message to player 1 is also "I would not go left if I thought you were going to defect." Hence, once we abandon the assumptions that the player types are committed to dominance moves in single game play, and that this is common knowledge, the analysis must seek inspiration from sources beyond game theory as traditionally understood to apply.

When the game is repeated in a series of trials with an unknown end point, we expect cooperation to increase in both games. Defection on the left in trial t can always be punished at low cost in trial $t + 1$ with the right branch (40, 40) outcome. Repeat play allows any ambiguity of intentions in the play sequence to be made clearer. Also, repeat play allows more subtle outcomes, such as (35, 70) to be reached in which player 2 moves right at x_2 and down at x_6 as a trigger threat strategy to get player 1 to move right at x_1. But player 1, at a cost, can counterthreat by moving down at x_8. If reciprocity principles are significant, as we predict, such threat/counterthreat escalations are likely to be rare.

Normal Form: Simultaneous Play. We can present the game in Fig. 1 in normal form to the subjects by expressing the payoff consequences of all possible sequences of moves for each subject in rectangular matrix form. This we refer to as the matrix normal form.

Reciprocity theory implies better coordination in achieving cooperation if individuals are predisposed to cooperate, and they are matched with like-disposed persons. This is because they can interpret moves in an extensive form sequence; such moves constitute a language, and a move sequence represents a conversation, albeit not without risk and ambiguity. In matrix normal form, this conversation is broken in a single play of the constituent game, although in repeat play communication becomes possible across time.

The matrix normal form is shown in Table 1. The column strategies are for player 1, the row strategies are for player 2. From Fig. 1, it is seen that player 1 has 10 strategy combinations of moves: two moves at each of the five nodes x_1, x_3, x_4, x_7, and x_8. At x_1 the move right invokes the outside option, yielding (35, 70), and precludes all later move options for player 1. Where asterisks follow a move set this indicates that a later possible move or moves are precluded by the indicated move. Hence, the first column heading is "R****", meaning "right at x_1, precludes the four possible choices at x_3, x_4, x_7, and x_8." Column 2, "DLR**," means "down at x_1, left at x_3, and right at x_4, precluding moves at x_7 and x_8." Note that column 2 yields payoffs of (60, 30) or (30, 60), depending upon whether player 2 moves left or right at x_2. Player 2 has four possible strategy combinations of moves: two moves at each of the nodes x_2, x_5, and x_6 (the move at x_2 does not yield a direct outcome; only a choice between the two outcomes at x_5 or at x_6). Thus row 3, with heading "R*R" means "right at x_2, preclusion of a move at x_5, and right move at x_6."

Note that we have *not* eliminated dominated strategies in Table 1, as is common in game theoretic analysis. This is because they are not eliminated in the extensive form; i.e., they are available to be chosen, and might be chosen for defensible reasons. Hence, referring to Table 1, column 7 weakly dominates 8, and column 3 weakly dominates 4. Similarly, row 3 weakly dominates 4, etc. There are many weakly dominated strategies. The extensive form SP outcome (40, 40) is a Nash equilibrium at (row, column) = (3, 3): given that player 2 chooses row 3, player 1 can do no better than column 3, and *vice versa*. Also (4, 1) is Nash, but note that in repeat play if player 2 persists in choosing row 4, player 1 can punish with column 3 (or 7 or 8), at a personal cost. Eliminating weakly dominated columns like 4, 8, or 10, precludes a maximum punishment (and cost) response.

In a single play of the matrix normal form, the scope for intentionality detection is sharply reduced. Coordination now requires each player to conduct an elaborate thought experiment

Table 4. Binomial tests comparing extensive and normal forms

Test	Treatment (Research hypothesis)	Extensive form proposition, x_e/N_e*	Normal form proposition, x_n/N_n†	Unit normal deviate approximation of binomial‡	P
1 Left play more likely	Single (Extensive > matrix normal)	12/26	7/24	1.23	0.11
2 Left: Cooperation more likely	Single (Extensive > matrix normal)	6/12	1/7	1.61	0.05
3 Right: Noncooperation more likely	Single (Extensive > matrix normal)	14/14	14/17	1.66	0.05

*x_e is number of pairs playing left branch in the extensive form out of a total of N_e playing left or right.

†x_n is number of pairs playing left branch in the normal form out of a total of N_n playing left or right.

‡All tests are one-tailed based on *a priori* prediction of which outcomes will be greatest.

Table 5. Same branch conditional outcomes by trial block

Trials	(35, 70)	Left	(60, 30)	(50, 50)	(20, 20)	Right	(30, 60)	(40, 40)	(15, 30)	(0, 0)
Extensive form										
1–5	15/115 = (0.13)	45/100 = 0.45	16/45 = 0.36	29/45 = 0.64	0	55/100 = 0.55	7/55 = 0.13	42/55 = 0.76	6/55 = 0.11	0
6–10	5/115 = (0.04)	65/110 = 0.59	10/65 = 0.15	55/65 = 0.85	0	45/110 = 0.41	11/45 = 0.24	26/45 = 0.58	8/45 = 0.18	0
11–15	7/115 = (0.06)	73/108 = 0.68	8/73 = 0.11	65/73 = 0.89	0	35/108 = 0.32	6/35 = 0.17	23/35 = 0.66	6/35 = 0.17	0
16–20	10/115 = (0.09)	78/105 = 0.74	7/78 = 0.09	71/78 = 0.91	0	27/105 = 0.26	3/27 = 0.11	23/27 = 0.85	1/27 = 0.04	0
Normal form										
1–5	7/120 = (0.06)	44/113 = 0.39	16/44 = 0.36	28/44 = 0.64	0	69/113 = 0.61	23/69 = 0.33	38/69 = 0.55	4/69 = 0.06	4/69 = 0.06
6–10	9/120 = (0.08)	51/111 = 0.46	12/51 = 0.24	39/51 = 0.76	0	60/111 = 0.54	6/60 = 0.10	41/60 = 0.68	7/60 = 0.12	6/60 = 0.10
11–15	7/120 = (0.06)	63/113 = 0.56	15/63 = 0.24	46/63 = 0.73	2/63 = 0.03	50/113 = 0.44	8/50 = 0.16	30/50 = 0.60	7/50 = 0.14	5/50 = 0.10
16–20	7/120 = (0.06)	58/113 = 0.51	13/58 = 0.22	43/58 = 0.74	2/58 = 0.02	55/113 = 0.49	11/55 = 0.20	33/55 = 0.60	5/55 = 0.09	6/55 = 0.11

which anticipates the counterpart's thinking, unassisted by move information. At minimum this creates an added cognitive cost of decision making.

Experimental Design; Matching Protocols and Game Forms. Our 2 × 2 block experimental design is shown in Table 2. The extensive and normal forms are executed in single-play protocols in which 8–12 subjects in each session are randomized into the player 1 and 2 positions and randomly matched for one round of play. Distinct inexperienced groups participate in both the extensive and normal form representation. The payoffs in cents in Fig. 1 and Table 1 are multiplied by 20 for the single-play treatments.

Both forms are also run in a repeat-play protocol using the same partners, and referred to as "Same." Subjects in these sessions are matched once at random, assigned player roles at random, then play the constituent game for 20 periods, but the number of periods is unknown to them. All Same-pair subjects are recruited for a 2-h experiment, but the sessions last only about 1 h, so that it is credible for them to expect a much greater number of repetitions.

Subjects receive a $5 fee for arriving on time for a session, and are paid their earnings privately at the end of the session. All subjects are carefully screened, using a data bank, before they are recruited, to eliminate anyone who has participated in a previous experiment.

Hypotheses

The game theoretical equivalence of the normal and extensive forms of a game leads to two hypotheses which can be contrasted with alternatives based on reciprocity.

H1: The extensive form version of Single as compared with the normal form will favor (*i*) cooperation, conditional on left branch play; (*ii*) noncooperation conditional or right branch play. This is because the higher payoff from cooperation will induce left play for some subjects; those playing the left game will coordinate better to achieve the cooperative (50, 50) outcome, whereas those playing the right game will coordinate better on the noncooperative (40, 40) outcome.

H2: In Same, cooperation will improve over time in both game forms, but the path of increasing left play, and of cooperation, will be higher in the extensive form than in the normal form because the former better facilitates the reading of mental states.

Tests of Hypotheses H1 and H2

Table 3 summarizes the experimental results for all treatments by listing the observed outcomes and frequencies for each payoff box in left or right branch conditional form, which is the form we use for testing hypothesis H1. The entries for Same in the extensive and matrix normal forms are aggregated across all 20 (nonindependent) trials in all sessions. In this section we report only tests using the Single data, where the observations are independent.

Tests of H1 are shown in Table 4 as follows:

Row 1. Left play in the extensive form exceeds that in the normal form as predicted, but the results are not statistically significant at conventional levels for sample sizes of 24 and 26.

Row 2. Contingent on play in the left branch, the hypothesis that cooperation will be higher in the extensive than in the normal form is supported.

Row 3. Contingent on play entering the right branch, coordination on the SP prediction (40, 40) is higher in the extensive than the normal form.

Repeat Play with Same Pairs

Table 5 provides the branch conditional outcome frequencies by blocks of five trials for the Same, extensive, and matrix normal form experiments. Observe that *both* left play and the (50, 50) outcomes in the extensive form dominate those for the normal form in every trial block. Over time, left branch play and support for the cooperative outcome builds steadily, with 91% (last trial block) cooperation in the extensive form, but in the normal form cooperative support is both weaker and more erratic—even under repetition, coordination in the normal form is illusive and difficult.

The above reported data for Same are too aggregated across individual observational pairs to give a sense of the dynamics of individual play outcomes. In particular, it does not convey information on the number of pairs, if any, that achieve the (50, 50) cooperative outcome in all 20 trials, or in no trials. This is remedied in Fig. 2, which plots the (cumulative) number of pairs that achieve

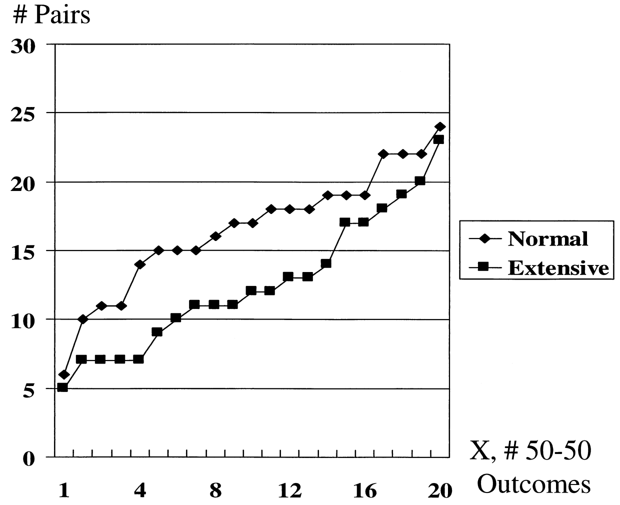

Fig. 2. Number of pairs achieving *X* or fewer (50, 50) outcomes in 20 trials. Subjects reach (50, 50) more often in the repeated 20-period play of the extensive form game, shown in Fig. 1, than in the repeated 20-period play of the comparable normal form game. These data support the hypothesis that move information is used in the extensive form game to inform subjects' theory-of-mind reasoning about the intentions of their counterpart.

Table 6. Regression analysis

Variable name	Estimated coefficient	Standard error	*T* ratio (36 df)	*P*
Logit 1 [proportion left branch play | down at x_1] Raw moment $r^2 = 0.6858$				
Constant	−0.25887	0.1360	−1.904	0.065
Trials, *t*	0.03673	0.01136	3.232	0.003
Game Form	−0.08076	0.2061	−0.3919	0.697
(Game Form) *t*	0.04291	0.0180	2.384	0.023
Logit 2 [proportion (50, 50) outcome | left at x_2] Raw moment $r^2 = 0.7482$				
Constant	−0.027058	0.2257	−0.1199	0.905
Trials, *t*	0.055026	0.01726	3.188	0.003
Game Form	0.92697	0.2049	4.524	<0.001
Logit 3 [proportion (40, 40) outcome | right at x_2] Raw moment $r^2 = 0.4610$				
Constant	0.32615	0.2292	1.423	0.163
Trials, *t*	0.02805	0.01850	1.125	0.268
Game Form	0.1438	0.2171	0.6596	0.514

Observations have been adjusted for heteroskedasticity.

any given number of (50, 50) outcomes or less. Thus, in normal form 11 pairs cooperate only 3 or fewer times in the 20 trial sequence, whereas in extensive form 11 pairs have cooperated 9 times. Both treatments yield some pairs who never cooperate and some who cooperate for the entire 20 trial sequence.

We turn now to a report of the results of a logit analysis of the trend pattern of change over successive trials in Same, and provide formal tests of the hypothesis H2 set forth above.

We report three regressions of the form

$$\ln(p/(1-p)) = \beta_0 + \beta_1 t + \beta_2 \text{ (Game form)},$$

where t refers to trial $t = 1, 2, \ldots, 20$ of an experiment and Game Form is an indicator (dummy) variable that has value 1 if Game Form is extensive and 0 if Game Form is normal. In one of the regressions below we add the interaction variable (Game Form) t to capture suspected interaction effects of Game Form on the time trend in $p/(1-p)$).[††] All regressions are weighted to correct for heteroskedasticity in the observations on the dependent variable.

In logit 1, p is the proportion of pairs (players 2) that choose to play in the left branch of the decision tree conditional on player 1 moving down at x_1; in logit 2, p is the proportion of pairs that achieve the (50, 50) outcome, conditional on player 2 moving left at x_2; and finally, logit 3 is the proportion of pairs that achieve the (40, 40) outcome, conditional on player 2 moving right at x_2.

Table 6 reports the regression results. In logit 1 the proportion of left branch play increases significantly with trials across both Game Forms (β_1 is significantly above zero), but the extensive form yields a further significant interactive increase with trials relative to the normal form (β_3 is significantly greater than zero). This can be seen by comparing the left branch frequencies in Table 5. In both the extensive and normal forms, the left play proportions increase with trial block, but the rate of increase is faster in the extensive form. Game Form alone, however, is not significant after accounting for its interaction with trials. In logit 2 both trials and Game Form are significant in determining the proportion of pairs yielding the (50, 50) outcomes, as proposed in H2, but in logit 3 neither is significant in explaining the (40, 40) outcomes.

Conclusions

We close with the following summary and discussion of our results:

(*i*) Directly comparing the extensive and normal forms in single play, the proportion of left branch play is higher in the extensive than the normal form, but the difference is not statistically significant at conventional levels.

(*ii*) Conditional on left game play, the proportion of (50, 50) cooperative outcomes in the extensive form is significantly greater than in the normal form ($P = 0.05$).

(*iii*) Conditional on right game play, the proportion of (40, 40) noncooperative outcomes is significantly higher in the extensive than the normal forms. Hence, the ability of the extensive form to facilitate the mutual reading of intentions allows noncooperators to better coordinate in achieving the noncooperative equilibrium than when they interact under the normal form.

(*iv*) When both game forms are repeated with the same matched pairs we observe a significant trend in successive trials in both the proportion of offers to cooperate, and the reciprocating achievement of cooperation. Game Form matters in determining left play offers to cooperate, but in the achievement of cooperation the extensive form interacts with trials to accelerate the increase in cooperation relative to that in the normal form.

[‡] From the Bellsely test for multicollinearity, none of the logit regressions containing the interaction term, (Game Form) *t*, exhibit significant multicollinearity.

1. McCabe, K., Rassenti, S. & Smith, V. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 13421–13428.
2. Rapoport, A. (1997) *Int. J. Game Theory* **26,** 113–136.
3. Schotter, A., Wiegelt, K. & Wilson, C. (1994) *Games Econ. Behav.* **6,** 445–468.
4. Kohlberg, E. & Mertens, J. (1986) *Econometrica* **54,** 1003–1038.
5. Schelling, T. (1960) *The Strategy of Conflict* (Harvard Univ. Press, Cambridge, MA).
6. Pinker, S. (1997) *How the Mind Works* (Norton, New York).
7. Cosmides, L. & Tooby, J. (1992) in *The Adapted Mind*, eds. Barkow, J., Cosmides, L. & Tooby, J. (Oxford Univ. Press, Oxford), pp. 19, 136.
8. Forsythe, R., Horowitz, J., Savin, N. & Sefton, M. (1994) *Games Econ. Behav.* **6,** 347–369.
9. Hoffman, E., McCabe, K., Shachat, K. & Smith, V. (1994) *Games Econ. Behav.* **7,** 346–380.
10. Baldwin, B. A. & Meese, G. B. (1979) *Anim. Behav.* **27,** 947–957.
11. McCabe, K., Rassenti, S. & Smith, V. (1998) *Games Econ. Behav.* **24,** 10–24.
12. Brown, D. E. (1991) *Human Universals* (McGraw–Hill, New York).
13. Hoffman, E., McCabe, K. & Smith, V. (1998) *Econ. Inquiry* **36,** 335–352.
14. Hoffman, E. & Spitzer, M. (1985) *J. Legal Stud.* **15,** 254–297.
15. Shogren, J. (1997) *J. Econ. Behav. Organ.* **32,** 383–394.
16. Kalai, E. & Lehrer, E. (1993) *Econometrica* **61,** 1019–1045.
17. Ball, S. & Eckel, C. (1996) *Psychol. Market.* **13,** 381–405.
18. Orbell, J., Goldman, M., Mulford, M. & Dawes, Robyn (1992) *Ration. Soc.* **4,** 291–307.
19. Gigerenzer, G. (2000) in *The Evolution of Mind,* eds. Cummings, D. & Allen, C. (Oxford Univ. Press, Oxford), in press.
20. Baron-Cohen, S. (1995) *Mindblindness: An Essay on Autism and Theory of Mind* (MIT Press, Cambridge, MA).