Old World Monkeys are More Similar to Humans Than New World Monkeys When Playing a Coordination Game

Sarah F. Brosnan  
*Georgia State University, sbrosnan@gsu.edu*

Bart J. Wilson  
*Chapman University, bjwilson@chapman.edu*

Michael J. Beran  
*Georgia State University, mberan1@gsu.edu*

Follow this and additional works at: https://scholarworks.gsu.edu/psych_facpub

Part of the *Psychology Commons*

**Recommended Citation**

Old World monkeys are more similar to humans than New World monkeys when playing a coordination game

Sarah F. Brosnan\textsuperscript{1,2,3,*}, Bart J. Wilson\textsuperscript{4,*}, Michael J. Beran\textsuperscript{1,2,*}

\textsuperscript{1}Language Research Center, Georgia State University
\textsuperscript{2}Department of Psychology, Georgia State University
\textsuperscript{3}Neuroscience Institute, Georgia State University
\textsuperscript{4}Economic Science Institute, Chapman University
\textsuperscript{*}All three authors contributed equally to this manuscript

Corresponding Author: Sarah Brosnan, Georgia State University Department of Psychology, PO Box 5010, Atlanta, GA 30302 USA; sbrosnan@gsu.edu
Abstract

There is much debate about how humans’ decision-making compares to that of other primates. One way to explore this is to compare species’ performance using identical methodologies in games with strategic interactions. We presented a computerized Assurance Game, which was either functionally simultaneous or sequential, to investigate how humans, rhesus monkeys, and capuchin monkeys utilized information in decision-making. All species coordinated via sequential play on the payoff-dominant Nash equilibrium, indicating that information about the partner’s choice improved decisions. Furthermore, some humans and rhesus monkeys found the payoff-dominant Nash equilibrium in the simultaneous game, even when it was the first condition presented. Thus, Old World primates solved the task without any external cues to their partner’s choice. Finally, when not explicitly prohibited, humans spontaneously used language to coordinate on the payoff-dominant Nash equilibrium, indicating an alternate mechanism for converting a simultaneous move game into a sequential move game. This phylogenetic distribution implies that no single mechanism drives coordination decisions across the primates, while humans’ ability to spontaneously use language to change the structure of the game emphasizes that multiple mechanisms may be used even within the same species. These results provide insight into the evolution of decision-making strategies across the primates.
Introduction

Along with the dispositions to trust and reciprocate and the propensity to exchange, the human ability to coordinate activities is a pillar upon which the flourishing of the species is built. The ability of two individuals to coordinate, literally to mutually arrange, an activity presupposes firstly that two individuals cognize that the outcomes of their actions are interdependent. Secondly, successful coordination assumes a shared attention and agreement on the ends to be achieved by mutually arranging a pair’s activities. Within the Pleistocene tribe or the modern small group of family, friends and neighbors, these conditions are almost trivially met as personally known individuals share the habits, knowledge, and beliefs about the methods and possibilities necessary to coordinate successfully. But what happens when modern strangers face a novel task of playing a simple 2 x 2 normal form game of coordination? How well can the personally unknown extend to each other the assumptions of interdependent decisions and agreement on the ends? As van Huyck et al. (1990) suggested, not as robustly as we might hubristically expect. The question then becomes more interesting when posed of our primate relatives. Do we share the ability to cognize actions as interdependent and to share attention on the ends achieved?

There is little research addressing the question, but what there is indicates that several primates may share these abilities with humans. For instance, in one Ultimatum game study with chimpanzees (Jensen et al. 2007), the apes’ decisions were dissimilar from human-typical behavior (Camerer 2003). However a subsequent study found that the protocol designed for chimpanzees led to a similar outcome in humans (Smith & Silberberg 2010). Our previous work investigated the Assurance game, which is a well-known model of social interactions (Skyrms
2003). We found that while humans found payoff dominant outcomes more readily than did either capuchin monkeys or chimpanzees, pairs of all species were able to find these outcomes, indicating that selection has favored similar outcomes (if not similar cognitive mechanisms) across these three primates (Brosnan et al. 2011). Moreover, we found that chimpanzees with greater experience in cognitive testing found the payoff dominant outcome far more readily than did chimpanzees with little or no cognitive testing experience, indicating that, as with humans, experience may play a role in outcomes.

For the current study, we chose to investigate the Assurance Game using a computerized methodology. This provided a number of advantages for further research. First, there is a long history in comparative research suggesting that the format in which one presents the same kind of task to nonhuman animals can have radical effects on performance. For example, spatial discontiguity between response loci and stimuli was recognized as an obstacle to learning in animals (e.g., McClean & Harlow 1954; Murphy & Miller 1959; Stollnitz & Schrier 1962). However, training primates to use a joystick gets past that problem of discontiguity and produces markedly different patterns of results (see Beran et al. 2007; Rumbaugh et al. 1989), and the same may be true for performance in economic games. Thus, comparing our previous results to those from a computer task may help to highlight factors which affect decision-making.

Another advantage of a computerized task is the abundance of data with respect to other cognitive abilities which are relevant. This allows us to not only consider a priori whether species might be able to solve the task, but if our predictions are proven false to re-consider how the subjects might perceive the game. Considering cognitive mechanisms, success in the game seems to require, at minimum, an ability to respond flexibly (e.g., contingent upon one’s partner’s decisions) and, related to this, an ability to inhibit (e.g., avoid the temptation of a large
Considering first the role of phylogenetically widespread learning mechanisms, a number of primate species, including the rhesus monkeys and capuchin monkeys who participated in the current study, have shown substantial behavioral flexibility in responding to game-like tasks presented on computer screens. These include tasks that involve behavioral inhibition, tracking of relative rates of reward for different responses, and even information-seeking behavior (Beran et al. 2011; Beran et al. 2008; Evans 2007; Flemming et al. 2011). These skills, along with the monkeys’ clear interest in maximizing their food intake during these tasks, indicate that comparative assessments of cooperation using computer tasks are likely to provide the most compelling data for understanding the emergence of cooperation in humans.

However, despite equivalent performance on basic learning tasks (e.g., two choice discrimination and learning set tasks), not all primates are equally adept at performing higher-level cognitive tasks. Relevant to this task, rhesus monkeys, but not capuchin monkeys, show evidence for metacognitive monitoring during psychophysical judgment tasks (Beran et al. 2009) and information-seeking paradigms (Beran & Smith 2011). Thus suggests a species difference in monitoring ongoing performance. Thus emerges a potentially important way to determine which cognitive mechanisms are important; it is possible that rhesus and capuchins would do equally well on games of coordination where contingencies for responses are clearly presented, that is, in a situation similar to a basic learning task, while diverging in performance when immediate cues are not present. Finally, it is possible, although we think unlikely, that even higher-order cognitive mechanisms are involved, such as theory of mind, which could be activated in this case due to the social nature of the task. If this is the case, we expect humans to outperform the other species (Penn & Povinelli 2007), despite some basic perspective abilities which have been seen in rhesus monkeys (Flombaum & Santos 2005). Thus, testing these species on the Assurance
game may shed light on not only performance levels, per se, on the game, but also the nature of
the game itself from the perspective of the individuals playing it.

To explore these issues in more detail, we here investigated the role of information in
coordination decisions amongst three primate species, humans (*Homo sapiens*), rhesus monkeys
(*Macaca mulatta*), and capuchin monkeys (*Cebus apella*). We re-designed the typical, normal-
form (NFG) Assurance Game methodology (Cooper et al. 2003; Ochs 1995; Smith 1982; van
Huyck et al. 1990) specifically to work across species, holding the methodology as constant as
possible (see Methods for details). We had two hypotheses for the current work. First, based on
our prior results with the exchange version of the task, we expected all species to be capable of
successfully navigating the task, but we predicted that a higher percentage of human pairs would
find the payoff dominant outcome as compared to the monkey species. Our second hypothesis
was that outcomes would change as the task parameters varied. Specifically we predicted that if
pairs could see each other’s choices prior to making a decision, the task could be solved by a
cognitively simpler matching rule, meaning that all species would perform equally well. On the
other hand, we predicted that in the situation in which their partners’ choices were not available,
higher-order mechanisms might be required, separating the species based on their aptitude at the
tests of cognitive abilities we proposed above as relevant to this task, that is, humans performing
better than the other primates.

**Methods**

*General Methods*

*The Assurance Game*
The game we used was a common game of coordination called the Assurance Game, sometimes referred to as the Stag Hunt game. The reward structure was such that mutual Stag play was the most beneficial (4 units), mutual Hare resulted in a low payoff (1 unit each), and the uncoordinated payoff of playing Stag when one’s partner plays Hare was unrewarded, while the individual who played Hare received 1 unit. This game has two pure strategy Nash equilibria: (Stag-Stag), which is the payoff dominant equilibrium (the outcome that maximizes payoff to both individuals), and (Hare-Hare), the outcome which is payoff dominated. This well-known coordination game is interesting to economists because strategic uncertainty plays a key role in the selection of the equilibrium, yet the players’ objectives are aligned (for a summary see Ochs 1995). In the Assurance game, the objectives may be the same (Stag, Stag), but the question of strategic interest is how sure a given player is that the other player will play Stag when he or she plays Stag. Evidence from coordination game experiments with humans indicate that the payoff dominant equilibrium is not a focal point with repeated interactions as in (van Huyck et al. 1990), or with anonymous play with different individuals as in (Cooper et al. 1990).¹

We explicitly incorporated a number of features common to non-human studies but different from traditional NFG experiments with humans to facilitate cross-species comparisons. First, subjects received no verbal instruction or pre-testing so that individuals had to discover the payoff structure during the course of the game (note that there were only two options from which to choose, and thus four possible outcomes). Second, all subjects, including humans, had participated in other experiments in the laboratory prior to this study so they were aware that decisions would result in tangible rewards (e.g. food or cash). Third, subjects were paid on a

¹ Though we should note that the former experiment involves more than two players with more than two potential actions, and the latter experiment uses a 3 x 3 coordination experiment.
trial-by-trial basis, in case the immediacy of receiving rewards on each trial affected behavior.

Fourth, for most treatments (see exception, below), subjects sat directly next to one another and
were not anonymous, so subjects could potentially communicate (Brosig 2002; Fetchenhauer et
al. 2010; we saw no attempts to communicate between the primates). Fifth, neither humans nor
rhesus monkeys received any pre-tests designed to assess Assurance game understanding, so all
subjects, including humans, had to discover the payoff structure during the course of the game.
Capuchins had previously participated in an exchange-based version of the task (Brosnan et al.
2011), but had no additional training. Finally, we manipulated whether or not they could see
their partners’ decisions to investigate how information affected decision-making.

General computerized design

Decisions were made by choosing one of two icons on each side of a split computer
screen, one of which represented Stag (a red square) and one of which represented Hare (a blue
circle). Icons were presented in a vertical distribution, with the order of presentation randomized
both across trials and across individuals within the same trial. Subjects of all species made a
choice using a joystick. We chose to study their behavior as naturally as possible, and so did not
constrain the order of play or the timing of decisions. Once both subjects had made a choice,
each subject received (or not) rewards dependent upon both what they chose and what their
partner chose, following the payoff structure of the assurance experiment.

There were two conditions, Synchronous, in which subjects did not know what their
partner had chosen until both choices were complete, and Asynchronous, in which decisions
were revealed as they were made (e.g., subjects potentially had information about their partners’
responses). To block any information transfer in the Synchronous condition, the joystick itself
was occluded and the cursor did not move; when they joystick was manipulated, both options
and the cursor disappeared simultaneously, and both subjects’ choices were displayed
simultaneously once both decisions had been made. Thus it was, to our knowledge, impossible
to determine the partner’s behavior by observation in this procedure, other than knowing that
their partner had made a choice. In the Asynchronous game, the procedure was identical, except
that choices were displayed as they were made, so that their partner could see their choice and
potentially use the information when making their decision.

Non-human primates

All non-human primates were socially housed at the Language Research Center of
Georgia State University. Rhesus monkeys were all adult males who were moved to a specially
designed paired testing area. Capuchin monkeys were socially housed in multi-male, multi-
female social arrangements and voluntarily separated in an adjacent cage for testing, to limit
distractions. Only adults were tested with members of their social group, and in multiple
pairings from within the same social group whenever possible. No individual was ever food or
water deprived for testing. During test sessions, pairs did not always finish a trial block, and
pairs completed different numbers of blocks during each testing session. Thus the number of
trials varied across both pairs and sessions.

The capuchins and the rhesus monkeys were used to somewhat different testing
schedules, so to avoid changing their schedules and causing unnecessary stress, we initially
proceeded using their typical schedules. Rhesus monkeys were given 6 hour testing sessions
consisting of 60-trial blocks with a 30 minute interval between blocks. Pairs could complete as
many trial blocks as they chose. Capuchins were initially given a single 40-trial session per day,
as per their norm, but of four pairs, only one achieved the Stag-Stag outcome in the
Asynchronous version, and even this pair did not maintain it. We then implemented a more
The non-human primates did require some training to learn to use the split screen (a novel experimental feature) and to make choices within the same time frame. This occurred through a two-stage training process. First, two monkeys worked together to learn that they had to both hit a single, solitary target on their side of the screen before both would receive a food pellet. Then, they had to progress to a point where they would make those same responses within a 5 second window from the initial presentation of a trial. Note that these training stimuli were not those used in the Assurance game, and there was no choice behavior on the part of the primates. There was only one icon on the computer screen which could be contacted with the cursor. This training was used to teach them which half of the screen presented their choices (and outcomes), and to teach them that they needed to respond relatively quickly when a trial was presented. This training assured that, at minimum, subjects knew that 1) rewards were not given without both individuals making a response and 2) they could control only their cursor.

Although of course we cannot know how the monkeys actually interpreted the task, we do know they at least understood how to generate responses that might bring rewards, and what limitations had been imposed.

*Capuchin monkeys* Based on our previous research (Brosnan et al. 2011), we knew that the capuchins would have more difficulty with the task than did the humans. Thus, we started all capuchins with the *Asynchronous* version of the task, and once they had learned it, gave them the *Synchronous* version. The subjects had far more difficulty with this, so to verify that they were
still able to do the basic task, we repeated the *Asynchronous* version. Finally, we repeated the *Synchronous* version to see if the extensive experience had increased their skill level.

*Rhesus monkeys*  Half of the rhesus monkeys were started on the *Synchronous* and half with the *Asynchronous* version. All of the subjects on the *Synchronous* version succeeded, and so we did not return them to the *Asynchronous* version (see Results and Table 2). Subjects that started on the *Asynchronous* version were subsequently run on the *Synchronous* version.

*Humans*

Undergraduate subjects were recruited from the general student body at Chapman University, Orange, CA, USA. Subjects were randomly recruited via an electronic email system and paid $7 for showing up on time, plus what they earned in the experiment. Each subject had participated in at least one economic experiment sometime prior to this session (participating in a previous study involving Normal Form game experiments or the Assurance Game disqualified individuals as a participant) so that they had experience with receiving actual payment for their decisions in this laboratory. No subject participated in more than one pairing or more than one version of the task.

The humans’ only instruction the *Synchronous* treatment was limited to the following six points:

1) Have you participated in an economic experiment before?

   (Both must reply with a “yes” to participate.)

2) In this experiment you will be making decisions using a joystick attached to a computer.

   Use the left thumb pad to make a decision.

3) As the experiment progresses you may be paid in quarters by the machines next to your computer.
4) Please collect the coins in the yellow cups provided so as to not clog up the machines.

5) These are the only instructions you will receive in the experiment. Once the experiment begins, the experimenter will not be allowed to answer *any* questions until the experiment is over.

6) Do you have any questions before the experiment begins?

Subjects initially began with the *Synchronous* task, as described above. Pairs of participants, who were the only two individuals in the room, sat next to one another at a single computer and used a joystick. The lack of anonymity enhances the likelihood of achieving the Pareto dominant outcome. Pairs received payment in quarters (from a coin dispenser, an analogue to the primates’ pellet dispenser) and payoffs were in the same ratios as those of the monkeys (accumulated coins were converted into large bills at the conclusion of the experiment). However, the results of this game indicated that language was an important characteristic in determining the pairs’ outcomes (see Results for details), thus a true comparison between the conditions could not be done. While we could have simply asked participants not to talk with each other during the game, this differed from the other primates, who could communicate to the fullest extent of their abilities, and may have led to an awkward social environment. Thus, we instead investigated the *Synchronous/Asynchronous* comparison using an alternate procedure. This procedure was based on a typical Normal Form Game (NFG) procedure with a 2 x 2 matrix of payoffs and strategies. For the NFG treatments, the game and payoffs were the same, however participants were isolated at individual study carrels playing on their own computer against an anonymous opponent drawn from amongst the other participants in the room. To hide who was partnered with whom, rewards accumulated and were paid out at the end of the session, rather than using a coin dispenser after each decision (the noise from coin dispensers would have
served as a cue). As with the monkeys, in the NFG Synchronous version, both partners’ choices were displayed simultaneously to both players after both decisions had been made, while in the NFG Asynchronous version choices were displayed to both players as they were made. As an additional benefit, having results in a traditional NFG procedure allowed us to see how our Synchronous results with humans compared to typical NFG experimental procedures (e.g. involving instruction).

Fifty-two undergraduate subjects participated in the study in pairs (i.e., in 26 separate sessions).

NFG Synchronous and NFG Asynchronous versions

One hundred eighteen undergraduate subjects were recruited by the same protocols above, except that 12-24 people participated at the same time. Fifty-eight people in three sessions of 22, 24, and 12 participated in the NFG Synchronous treatment and 60 people in three sessions of 24, 24, and 12 participated in the NFG Asynchronous treatment. The subjects were simultaneously seated in a computer laboratory at visually-isolated carrels and instructed not to talk to one another. They then read self-paced instructions on how to participate in the experiment. These subjects were privately paid their total accumulated earnings at the conclusion of the experiment; they did not receive payment as they made each decision to avoid possible cuing to one’s partner’s identity.

Statistics

Statistics are non-parametric due to small sample sizes. Primate results are based on individual analyses, while for humans we include both individual analyses and inferential statistics that allow generalizations about the population. This difference in approach is because
fewer monkeys were available for the study, as few are sufficiently well trained for computerized testing. All statistics are two-tailed.

Note that in many cases both chi-square and Fisher’s exact tests were impossible due to the large number of cells with zero values. Thus, to determine whether a pair showed a pattern in their decision-making, we considered it meaningful if the subject showed an 80% or greater preference for one of the four options (for the pair; chance was actually 25% in this case) or one of the two options (for the individual; chance is 50%). This percentage is significant for a binomial test for 20 trials and, as all of our subjects had at least 40 trials (and for pairs, chance was 25%, not 50%), this represented a conservative estimate for what constituted a significant pattern to their decision-making.

Results

Non-human primate results

Both monkey species did very well in the Asynchronous task. All capuchins began with the Asynchronous version due to previous results indicating that they would have trouble finding the payoff dominant outcome (Brosnan et al. 2011). One pair reached the payoff-dominant outcome in our initial 40-trial sessions and the other three did so when switched to 60-trial sessions (see Methods, above). Three of the rhesus monkey pairs (composed of four unique individuals) first played the Asynchronous game. Two reached the 80% Stag-Stag criterion within a single session (see Table 2) and the third pair did so in the second session.

Despite this similarity, the monkeys differed in their outcomes in the Synchronous task. None of the capuchin pairings showed any preference for playing Stag-Stag (or any other outcome) when tested in a novel pairing on the Synchronous task. To see if this was due to a
lack of understanding of the task, we re-ran the *Asynchronous* version with the capuchins. Eight monkeys were paired in multiple pairings (range: 1-3; see Table 1) for a single session consisting of as many trials as they chose to complete in 2 hours (mean=253 trials). All but one pair chose *Stag-Stag* at least 80% of the time (range: 80.6%-96.1%; see Table 1) and the exceptional pair chose *Stag-Stag* 71% of the time. Nonetheless, when retested on the *Synchronous* task, outcomes remained poor. *Stag-Stag* was maintained in only one of the five pairs which had been tested together previously (see Table 1). Intriguingly, the exceptional pair showed the lowest frequency of *Stag-Stag* outcomes in the *Asynchronous* test, and the pair which had not previously been tested together also most often played the *Stag-Stag* outcome (see Table 1). Note that capuchins’ poor performance occurred despite both previous experience with the Assurance game and introductory experience in the presumably easier *Asynchronous* version of the game.

On the other hand, the two pairs of rhesus monkeys that were tested first on the *Synchronous* version quickly found the *Stag-Stag* outcome. Although it took them slightly longer to reach the 80% criterion than did the pairs that played the *Asynchronous* version first, both pairs did so within 4 trial blocks, and at the same frequency as those pairs which first played the *Asynchronous* version (see Table 2). Note that they accomplished this despite never having had the opportunity to match a partner’s play after having seen it, ruling out this simple associative mechanism.

Finally, the two rhesus pairs which showed the highest frequency of *Stag-Stag* choices in the *Asynchronous* game were subsequently given the opportunity to play the *Synchronous* version. One pair maintained a preference for playing *Stag-Stag*, while the other started at a lower level of *Stag-Stag* choices and subsequently declined further. This seemed to be primarily due to one individual who began preferentially choosing *Hare*. Thus, unlike with the capuchins,
at least one pair was able to maintain the Stag-Stag outcome when switched to the Synchronous task.

Human Results

Among 27 human pairs in the Synchronous condition, 22 ultimately settled on Stag-Stag and 5 settled on Hare-Hare. What was notable was the perfect correlation between outcomes and pairs’ discussions. Although all pairs spoke to each other, textual analysis of video recordings revealed that not every pair spoke about the game. Amongst the 5 pairs who settled on Hare-Hare (range: 34-36 Hare-Hare choices in 40 trials; see Table 3), not a single pair spoke about the game (henceforth, non-communicators), while among the 22 pairs who did so (henceforth, communicators), every pair ultimately settled on Stag-Stag, choosing it in at least seven of the last 10 choices (13 pairs did so on every one of these choices). One-third of communicators chose Stag-Stag a minimum of 80% of the time overall (range: 32-39 Stag-Stag choices) and more than half (55%) did so at least 70% of the time. No non-communicators ever played Stag-Stag, indicating that they did not explore the decision space as thoroughly as did the other pairs (or the monkeys).

Given the larger human sample size, we can consider variation using inferential statistics. The overall payoff between communicators and non-communicators differed by a factor of 3 (Independent samples Mann-Whitney U test: p<0.001, mean±SE payoff per trial for pairs of communicators: $1.48±.08; non-communicators: $0.47±.002). Moreover, among communicators, there was an increase in payoff between the first quartile and last quartile (Wilcoxon Signed-ranks test, p <0.001, quartile 1 mean±SE: $0.98±.11; quartile 4: $1.89±.03). On the other hand, among non-communicators, the payoffs increased across these quartiles only because the decrease in number of Stag choices stabilized payoffs at one quarter for each
individual by the fourth quartile (Wilcoxon, p=0.039; quartile 1 mean±SE: $0.44±.01; quartile 2: $0.50±.00). Thus, sociality is not synonymous with coordination; humans had to actually
discuss the game in order to benefit from language. Moreover, humans were not intrinsically
better than the other primates at the *Synchronous* condition. They appeared to use
communication to turn the *Synchronous* game in to an asynchronous game, giving humans an
additional mechanism for coordination.

Based on these results, humans played the *NFG Synchronous* and *NFG Asynchronous*
versions using a more traditional normal-form game set-up that prohibited discussion (see
Methods). One third (10 of 30) of the pairs in the *NFG Asynchronous* game chose *Stag-Stag* at
least 80% of the time. Intriguingly, an additional four of these pairs chose *Hare-Hare* 80% or
more (and one other did so 78% of the time), an outcome that we never saw in either of the
monkey species. This could indicate the non-human primates are more likely to explore the
decision space than are humans, or that humans are more likely to persevere on responses that
are rewarding. In the *NFG Synchronous* game, only four (14%) pairs chose *Stag-Stag* this often,
while ten (35%) pairs chose *Hare-Hare*. Thus the frequency of *Stag-Stag* and *Hare-Hare*
choices flipped between the two conditions.

Considering the data quantitatively, the overall payoffs do not differ between the games
(Mann-Whitney, z=1.47, p=0.14), although there is a non-significant trend for the players in the
*NFG Asynchronous* game to earn more by the fourth period than do players in the *NFG
Synchronous* game (Mann-Whitney, z=1.85, p=0.0643). Moreover, the number of *Stag-Stag*
choices was greater in the *NFG Asynchronous* treatment than *NFG Synchronous* treatment
(Mann-Whitney, z=1.9, p=0.0574), and twice as many partnerships failed to ever play *Stag-Stag*
in the *NFG Synchronous* as compared to the *NFG Asynchronous* treatment (10/29 vs 5/30).
Considering the games individually, participants showed a trend towards playing *Stag-Stag* more often than *Hare-Hare* in the *NFG Asynchronous* game (Wilcoxon, $z=1.86$, $p=0.0624$), but not the *NFG Synchronous* version (Wilcoxon, $z=0.46$, $p=0.6818$). Subjects improved over the course of both treatments, earning more money in the fourth quartile than in the first (*NFG Asynchronous*: Wilcoxon, $z=4.26$, $p<0.001$, *NFG Synchronous*: $z=3.42$, $p=0.0006$). Nonetheless, as with the other two primate species, humans were better at finding the payoff dominant outcome when an opportunity to coordinate was presented (e.g., sequential play).

Finally, comparing humans’ results from the nonhuman primate format (which was the *Synchronous* treatment) to the *NFG Synchronous* format, the pairs’ payoffs were higher when they could talk (mean payoffs for the *Synchronous* treatment: $1.63$; *NFG Synchronous* treatment: $1.07$; Mann-Whitney, $z=2.62$, $p=0.0087$). On the other hand, by the end of the game, subjects did equally well when they could see each other’s decision as it was made as when they could talk (comparing fourth quartile payoffs in the *Synchronous* and *NFG Asynchronous* treatments; Mann-Whitney, $z=1.00$, $p=0.3179$; see Table 3). Thus, subjects were equally able to use language or sequential moves without language to coordinate on the *Stag-Stag* outcome and achieve the same level of payoffs across procedures.

*Does finding *Stag-Stag* mark a change in behavior?*

The above results assess individuals’ tendency to find the cooperative, *Stag-Stag*, solution. That is, to what degree do they explore the problem space and encounter *Stag-Stag* as an option? A second way to consider these data is to determine the frequency with which they played stag-stag after having found it for the first time. Considering first the monkeys, there is no evidence that finding *Stag-Stag* was sufficient to alter their behavior. Amongst capuchins, no pair ever had a single trial block (60 trials) in which they did not play Stag-Stag at least once (see
Supplemental Results for details), yet despite this not a single pair was able to successfully solve the Synchronous version of the task. Rhesus similarly played Stag-Stag in every trial block, yet it was rare for a pair to settle on Stag-Stag before the third trial block (see Supplemental Results for details).

Humans showed a similar pattern. In the Asynchronous task, most players who found Stag-Stag did so quickly (within the first 4 trials), yet only three (12%) of these pairs played nothing but Stag-Stag after finding it. In the Synchronous task, no pair ever exclusively played Stag-Stag after finding it (for details of both conditions, see Supplemental Methods). Thus we find clear indication that in neither of the two computerized conditions do pairs’ payoffs change meaningfully after finding the first Stag-Stag outcome, indicating that these results cannot be explained by a lack of knowledge about the payoff matrix. Similarly, sitting next to one’s partner did not affect the frequency with which individuals stayed on the Stag-Stag payoff. Of the 22 pairs that played Stag-Stag at some point, only one played nothing else after finding that outcome. Four additional pairs played Stag-Stag all but one or two times, possibly indicating that they were either jointly exploring the parameter space, or individually exploring potentially increased outcomes (see Supplemental Results).

**Discussion**

Our results indicate that generalizing the decision-making outcomes of primates into one in which species are ‘better’ or ‘worse’ than others is not sufficiently nuanced. Instead, we find that the results vary in interesting ways depending upon context. Considering our second hypothesis first, all species were able to more easily solve the task in the Asynchronous than in the Synchronous conditions. Although these studies were explicitly designed to investigate
outcomes, not mechanism, we are able to use these results to posit necessary mechanisms. Thus it seems likely that this boost in performance was due to the availability of simple rules, such as matching-to-sample (e.g., Cumming & Berryman 1961) which could not be used in the Synchronous task since the partner’s behavior was hidden. This indicates that it is possible to solve this task with a suite of fairly simple behavioral mechanisms. We also note that, without other data, this would indicate phylogenetic continuity amongst the primates, and thus, when considered with the Synchronous results, emphasizes the utility of using multiple methods to assess behavior and cognition.

On the other hand, not all species were able to solve the Synchronous task. No pair of capuchin monkeys ever solved this task, even after experience (successfully) solving the Asynchronous task. Thus, the monkeys’ inability in the Synchronous pairing was not due to a failure to understand the task outcomes. Instead, we think that the best explanation for the decline in performance in the Synchronous task is likely related to the fact that they could no longer see their partners’ decisions and use a simple strategy such as matching. These results also imply that the one pair’s success in the previous, exchange version (Brosnan et al. 2011) was likely due to the fact that they could see their partners’ choices. On the other hand, some rhesus monkeys and humans found the payoff dominant outcome in the Asynchronous task, and required very few trials in order to do so. Thus, Old World primates outperformed New World primates, rather than humans outperforming nonhumans.

This has several very important implications when considering the mechanisms required to solve the Assurance game. First, of course, the task can be solved using fairly simple learning-based mechanisms when appropriate cues are available (e.g., the Asynchronous task). Second, the high performance of both rhesus and humans implies that the presence of theory of
mind is not required. Although one can see how the ability to predict one’s partners’ behavior would be useful, it seems the task can be solved without it. It will be interesting to see whether this remains true in more complex games which lack a mutually beneficially payoff dominant outcome. Finally, in the absence of cues, it seems likely that other more complex cognitive abilities are required to solve the task. We note that these results are particularly strong given to our choice of a New World monkey species. Capuchins have an unusually large (for a monkey) brain-to-body ratio (Rilling & Insel 1999), a general high ability in cooperative tasks (e.g., de Waal & Berger 2000; Mendres & de Waal 2000), and equal rhesus in many learning tasks (see Introduction). Any of these might have led to the prediction that capuchins would cluster with the Old World monkeys, even if uniquely amongst New World primates. On the other hand, despite the capuchins’ apparent advantages (for solving this task) over other New World monkeys, cognitive differences favoring rhesus monkeys over capuchins have been seen in other recent tasks involving higher-order cognitive abilities, such as metacognition (Beran et al. 2009; Hampton 2001).

Our final intriguing result is that outcomes in these games, and in particular the inefficiencies in decision-making, were based on an inability to coordinate on the payoff-dominant outcome, not an inability to locate it due to a lack of exploration of the problem space. None of our subjects, including humans, immediately reverted to the payoff dominant choice after having first experienced it. This was true whether they could or could not see their partners’ choice when it was made or, in the case of humans, whether or not they could talk to their partner. This observation is clearly contrary to what would be expected if finding the payoff dominant solution is the only thing required to solve this game. Moreover, this result indicates that there are really two aspects involved in successful performance. First is exploring
the parameter space, and locating the payoff dominant outcome. In this, we saw great variability within each species, as indicated in our analysis of their overall results. Second is the ability of pairs to coordinate on that outcome once they have found it. This is explored in our analysis of the choices following the first Stag-Stag play. The data indicate that both of these aspects of the Assurance game present challenges to all three species; obviously individuals who do not fully explore the parameter space may never find the payoff dominant outcome, but even after this has been found, coordination is not assured.

While we are strongly in favor of testing additional primates (and non-primates) of all taxa, given these data we predict that evidence will continue to favor a New World/Old World split in decision-making outcomes. We are also enthusiastic to see how the inclusion of different game features which may require other abilities, such as theory of mind, will affect species’ play. We predict that such studies will serve to illuminate not only the phylogeny of decision-making, but also the cognitive requirements of different decision-making situations, as illustrated by various economic games. In particular, some rhesus monkeys were notable for their success, a pattern of results which clearly deserve additional investigation. One possibility for investigating these mechanisms is to see whether humans and rhesus monkeys differ in the flexibility with which they can switch between strategies. This question is more than academic; if humans and rhesus monkeys are using similar mechanisms (e.g., a cognitive homology), then we know that humans’ abilities are built on shared foundation that extends back at least as far as the split with Old World monkeys. On the other hand, if humans and rhesus monkeys have converged on similar outcomes despite using different cognitive mechanisms to reach those outcomes, then we can begin to investigate what shared social or ecological factors in humans’ and rhesus’ environments selected for these outcomes.
Finally, uniquely among the primates, language is an important mechanism for solving coordination tasks in humans, much as additional information was used by all species in the *Asynchronous* task. In other words, we hypothesize that humans may use communication to transform a *Synchronous* task into an *Asynchronous* one by providing information about future moves. This may indicate that some underlying mechanisms, such as the ability to increase performance when cues are present, are conserved between humans and other primates, albeit with differing specific mechanisms. We also think the most likely explanation for the humans who did not communicate about, and thus did not solve, the task is not that they were incapable of doing so. Even subjects who did not solve the task spoke to one another. Instead we propose that they thought that they *had* solved the task, did not see the necessity of exploring other options, and so did not converse about it. Future research aimed at determining which factors cause these different reactions may help to clarify both individual (e.g., personality) and contextual (e.g., social) factors which affect humans’ ability to coordinate in such situations.

The picture that is emerging makes it clear that humans are not alone in our ability to find efficient, pay-off maximizing outcomes in a coordination game. Moreover, other species’ behaviors can be measured using the same experimental mechanisms common in humans, that is, economic games, and results can be made comparable when procedures are equalized across species, including humans. Future work should be done to investigate how cognitive mechanisms interact with game structure, and whether there is homology in the underlying cognitive mechanisms or instead whether similar outcomes are reached in analogous ways. Finally, it will be interesting to see whether this similarity in outcome remains in other more challenging decision-making situations.
Acknowledgements

This research was funded by an NSF HSD grant SES 0729244 to SFB, BJW and MJB. SFB was additionally funded by NSF CAREER award SES 0847351; MJB was additionally funded by NSF grants BCS-0924811 and BCS 0956993 and NICHD grants HD-38051, HD-061455, and HD-060563; and BJW was additionally funded by NSF SES 0833310. We thank Betty Chan, Theodore Evans, Audrey Parrish, and Kate Talbot for assistance with data collection at the LRC, Jennifer Cunningham for subject recruiting and data collection at Chapman University, and Jeffrey Kirchner for software programming at Chapman University. We additionally thank the animal care and enrichment staff for maintaining the health and wellbeing of the primates and making this research possible. All authors contributed equally in experimental design, data collection, analysis, and writing. The authors declare no competing interests.


Beran, M. J. & Smith, J. D. 2011 Information seeking by rhesus monkeys (Macaca mulatta) and capuchin monkeys (Cebus apella). Cognition 120, 90-105.


Evans, T. A. 2007 Performance in a computerized self-control task by rhesus macaques (Macaca mulatta): The combined influence of effort and delay. Learning and Motivation 38, 342-357.

Fetchenhauer, D., Groothuis, T. & Pradel, J. 2010 Not only states but traits - Humans can identify permanent altruistic dispositions in 20 s. Evolution & Human Behavior 31, 80-86.


Hampton, R. R. 2001 Rhesus monkeys know when they remember. PNAS 98, 5359-5362.


Table 1: Comparing the capuchin pairs’ performance on the second asynchronous and synchronous games (note some pairs had previous experience on both the asynchronous and synchronous games; see Results for details).

<table>
<thead>
<tr>
<th>Pairings</th>
<th>Asynchronous Game (second)</th>
<th>Synchronous Game (second)</th>
<th>Overall</th>
<th>First Session</th>
<th>Second Session</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% Stag-Stag</td>
<td># trials (1 session)</td>
<td>% Stag-Stag</td>
<td># sessions/trials</td>
<td>% Stag-Stag</td>
</tr>
<tr>
<td>Drella Wren</td>
<td>80.6</td>
<td>201</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Griffin Drella</td>
<td>93.0</td>
<td>196</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Griffin Wren</td>
<td>96.1</td>
<td>360</td>
<td>33.1</td>
<td>8/3042</td>
<td>34</td>
</tr>
<tr>
<td>Lily Wren</td>
<td>85.2</td>
<td>240</td>
<td>38.9</td>
<td>7/2176</td>
<td>81.1</td>
</tr>
<tr>
<td>Griffin Lily</td>
<td>81.7</td>
<td>68</td>
<td>28.4</td>
<td>6/2864</td>
<td>36.1</td>
</tr>
<tr>
<td>Liam Logan</td>
<td>71.3</td>
<td>240</td>
<td>78.1</td>
<td>6/1467</td>
<td>88.2</td>
</tr>
<tr>
<td>Logan Gabe</td>
<td>89.2</td>
<td>240</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nala Logan</td>
<td>93.5</td>
<td>420</td>
<td>34.6</td>
<td>5/2260</td>
<td>74.2</td>
</tr>
<tr>
<td>Nala Liam</td>
<td>60.1</td>
<td>4/1437</td>
<td>78.3</td>
<td>300</td>
<td>64.0</td>
</tr>
</tbody>
</table>
Table 2: Comparing rhesus performance by trial.

<table>
<thead>
<tr>
<th>Game</th>
<th>Game Type</th>
<th>Overall %</th>
<th># sessions/trials</th>
<th>% Stag-Stag</th>
<th>Session 1</th>
<th>Session 2</th>
<th>Session 3</th>
<th>Session 4</th>
<th>Session 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obi Han</td>
<td>Asynchronous</td>
<td>74.6</td>
<td>5/1483</td>
<td>53.7</td>
<td>77.9</td>
<td>84.6</td>
<td>76.7</td>
<td>81.0</td>
<td></td>
</tr>
<tr>
<td>Luke Obi</td>
<td>Asynchronous</td>
<td>94.5</td>
<td>5/1876</td>
<td>90.9</td>
<td>95.3</td>
<td>96.1</td>
<td>93.9</td>
<td>96.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Synchronous</td>
<td>65.7</td>
<td>2/824</td>
<td>76.7</td>
<td>54.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chewey Han*</td>
<td>Asynchronous</td>
<td>87.8</td>
<td>5/1440</td>
<td>85</td>
<td>94.2</td>
<td>85.3</td>
<td>87.2</td>
<td>87.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Synchronous</td>
<td>72.2</td>
<td>3/360</td>
<td>66.7</td>
<td>68.8</td>
<td>81.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hank Gale</td>
<td>Synchronous</td>
<td>84.0</td>
<td>2/720</td>
<td>74.5</td>
<td>93.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Willie Murph</td>
<td>Synchronous</td>
<td>88.1</td>
<td>3/472</td>
<td>31.6^</td>
<td>81.3</td>
<td>95.7</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Both pairs which played both games played Asynchronous first, followed by synchronous.

^This represented only a single trial block; in actuality Willie and Murph reached 90% Stag-Stag preference in their third trial block, which was in their second session. Hank and Gale did not reach 90% until their fifth trial block, which was in their first session.
Table 3: Human preferences across the three games. Earnings measure the mean earnings per trial for the pair, and so range from 0.25 (one individual earned a quarter from playing Hare and their partner earned nothing from playing Stag) to 2.0 (both earned a dollar from playing Stag). An average of 0.5 means that both played Hare on every trial. A mean of 1.825 indicates a single Hare play and a mean of 1.85 indicates a single Hare-Hare trial.

<table>
<thead>
<tr>
<th></th>
<th>Talking?</th>
<th>Spoke about symbols?</th>
<th>Number of pairs</th>
<th>Number of trials</th>
<th>Ever played Stag-Stag</th>
<th>Mean # Stag-Stag</th>
<th>Mean # Hare-Hare</th>
<th>Mean Q1 earnings</th>
<th>Mean Q4 earnings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paired</td>
<td>Yes</td>
<td>Yes</td>
<td>22</td>
<td>40</td>
<td>ALL</td>
<td>27.7 (69%)</td>
<td>6.2</td>
<td>0.98</td>
<td>1.89</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
<td>5</td>
<td>40</td>
<td>NONE</td>
<td>0 (0%)</td>
<td>35.2</td>
<td>0.44</td>
<td>0.5</td>
</tr>
<tr>
<td>Synchronous</td>
<td>No</td>
<td></td>
<td>30</td>
<td>40</td>
<td>25 (83%)</td>
<td>20.27 (51%)</td>
<td>10.83</td>
<td>0.93</td>
<td>1.43</td>
</tr>
<tr>
<td>Asynchronous</td>
<td>No</td>
<td></td>
<td>29</td>
<td>40</td>
<td>19 (66%)</td>
<td>13.48 (34%)</td>
<td>16.28</td>
<td>0.77</td>
<td>1.07</td>
</tr>
</tbody>
</table>
Supplemental Methods

Non-human primate housing details

All non-human primates were socially housed at the Language Research Center of Georgia State University. Rhesus monkeys were all adult males, individually housed in rooms with 1 to 2 other individuals for testing and each spent time outside in various social arrangements. Rhesus monkeys were moved to a specially designed paired testing area where they could share a single monitor between the two subjects. Capuchin monkeys were socially housed in multi-male, multi-female social arrangements with extensive indoor and outdoor living space complete with environmental enrichment. All capuchins were adults who lived in the same social group and, when possible, multiple pairings within the same social group were tested. For testing, capuchin pairs were separated from the remainder of the social group into a testing cage located in the indoor colony room to limit distractions. All separation was voluntary (i.e., subjects were only tested if they entered the testing area freely). Both species received a combination of primate chow, fruits, and vegetables as well as additional enrichment foods.

Rhesus testing schedule

Rhesus monkeys were given 6 hour testing sessions. During these sessions, pairs could complete as many trial blocks as they chose. Sessions consisted of 60-trial blocks with 30 minute long intervals between each trial block during which the program was unavailable.
**Capuchin testing schedule**

Capuchins were initially given a single 40-trial session per day, which reflected their typical testing pattern and the general trend that they complete fewer daily trials than do rhesus monkeys. However, only one of four unique pairs repeated achieved the Stag-Stag outcome in the *Asynchronous* version, and even this preference peaked at 67.5% and dropped off over the next 6 sessions to only 45% (not different from *Hare-Stag* at 40%). At this point, we implemented a more rhesus-like schedule, except that their test sessions were only two hours long (they became agitated if left in their testing cages for longer than this amount of time). Using this longer format, all three showed a 90% preference *Stag-Stag* by their first (2 pairs) or third (1 pair) session. Given these results, 60 trial sessions were used throughout the remainder of the testing.

**Supplemental Results**

*Does finding Stag-Stag mark a change in behavior?*

Of the four conditions for the capuchin monkeys, not a single pair ever had a single trial block (60 trials) in which they did not play Stag-Stag at least once (First trial block mean±SEM percent Stag-Stag choices, Asynchronous first time: 30.2±4.5, Asynchronous second time: 78.8±4.6, Synchronous first time: 14.1±7.0, Synchronous second time: 58.3±10). Thus for all test blocks (and sessions) following their first test block in the first Asynchronous condition, subjects had previously experience Stag-Stag, making it unlikely that the challenges the capuchins faced were due to not recognizing Stag-Stag as an option. In fact, in their second Synchronous condition (that is, the fourth condition chronologically), the capuchins did significantly better on the first trial block than they did overall (overall mean: 45.58±4.0;
Wilcoxon, $T=10$, $n=12$, $p=0.008$), indicating that previous exposure to Stag-Stag was not sufficient for them to solve the task.

All pairs of rhesus monkeys played Stag-Stag during the first trial block (First trial block mean±SEM percent Stag-Stag choices. Asynchronous: 59.6±15.2, Synchronous: 41.7±8.4). Nonetheless, it was rare for pairs to reach 80% Stag-Stag outcomes in the next trial block (Asynchronous: mean=10, range = 1-28 trial blocks; Synchronous: mean=3, range = 2-5 trial blocks; note that the 80% criterion was reached more rapidly in the Synchronous condition, but no pair found it on the first trial block). Thus simply having experience with the Stag-Stag outcome was insufficient to settle upon that strategy.

In the *Asynchronous* computerized task, 43% (13 pairs) of humans played Stag-Stag on their first move and 17% (5 pair) never played it (the remaining players averaged 4.5 trials, range 2-17). Of the pairs who found the Stag-Stag payoff, only three (12%) pairs subsequently played nothing but Stag-Stag after finding it. In the *Synchronous* condition, 11 of 29 human pairs never played Stag-Stag, and only one pair (3%) played it on the first trial. The remaining pairs found the Stag-Stag outcome in a mean of 3.8 trials (range 2-12). No pair played nothing but Stag-Stag after finding it. In fact, while the overall mean payoffs increased by definition when any trials prior to a Stag-Stag choice are excluded, the mean payoff increased from 1.14 to 1.19, a very modest increase (non-parametric statistics based on rank are, again by definition, significant, but we think meaningless in this case). Thus we find clear indication that in neither of the two computerized conditions do pairs’ payoffs change meaningfully after finding the first Stag-Stag outcome, indicating that these results cannot be explained by a lack of knowledge about the payoff matrix.
The primate version of the human game produced the same general results. Of the 22 pairs that played Stag-Stag at some point, they found the Stag-Stag outcome in a mean of 6.7 trials (range 1-32). Only one pair played nothing but Stag-Stag after experiencing the outcome; the rest took a mean of 21.6 trials (range 3-39) to settle on all Stag-Stag choices. Several pairs had only one or two non-Stag-Stag choices after finding the outcome, indicating that either one individual tried cheating (n=4) or they tried the Hare-Hare outcome, presumably after discussion (n=4; two pairs in both measures). If these pairs are counted from the time they went mostly to Stag-Stag, then 5 pairs (all but one of them) immediately preferred Stag-Stag.