

COMMENT AND REPLY

Comment on Mark, ASR, June 2002

Explication of the Cultural Transmission Model

Elisa Jayne Bienenstock
University of California, Irvine

Michael McBride
University of California, Irvine

In “Cultural Transmission, Disproportionate Prior Exposure, and the Evolution of Cooperation,” Mark (2002) claims to have “conducted a formal theoretical analysis that isolated the logic of cultural transmission [to] reveal an evolutionary force toward cooperation under conditions previously thought to make the evolution of cooperation impossible” (p. 324). The most striking aspect of Mark’s work is not that his simple model converges to cooperative behavior, but that it works under initial conditions severely biased against the evolution of cooperation: randomly matched interaction in a large population with only 1% initial cooperators (99% initial defectors). Although this result is compelling at first glance, we show that it is not robust to minor changes in the model: cooperative behavior evolves only when the particular model specifications chosen by Mark are selected and does not evolve when the model is generalized. In particular, we show that the result relies on a decoupling between actors’ fitness and behaviors. According to Mark’s theoretical model, this decoupling between fitness and behaviors should be associated, depends on simplifications of his own proposed mechanism, and is not robust in small populations. We conclude that the role of Mark’s formulation of disproportionate prior exposure (DPE) for the evolution of cooperation is unclear, though it

may be a relevant factor in understanding the transmission of cultural traits like language.

SUMMARY OF MARK’S THEORY AND MODEL

Mark’s thesis is that “disproportionate prior exposure creates a cultural evolutionary force toward cooperation” (p. 324). Language is the best illustration of how DPE is required for cultural transmission because all speakers of a language have been repeatedly exposed to the language. Mark’s purpose is to extend DPE to cooperative behavior; he claims that “cultural transmission of almost any characteristic produces disproportionate prior exposure to that characteristic” (p. 330). This reasoning leads him to conclude that DPE must be present if a trait is present (footnote 12, p. 330).

To investigate this relation, Mark presents a computational model that randomly pairs members of a large population in one-shot prisoner’s dilemma (PD) game. Figure 1 shows the payoffs. Each individual has two traits: a strategy (i.e., propensity to either cooperate or defect) and a fitness level with values 0, 1, 2 or 3. Departing from most work in evolutionary game theory (Leimar and Hammerstein 2001; Nowak and Sigmund 1998a; Nowak and Sigmund 1998b; Wilson 1989), Mark defines fitness as the value acquired during the previous pairing

Direct correspondence to Elisa Jayne Bienenstock, Booz Allen Hamilton, Fall Church, VA 22042 (ejb@bah.com). We would like to thank Michael Macy, Alison Bianchi and two anonymous reviewers for valuable comments on earlier drafts.

		Actor 2	
		Cooperate	Defect
Actor 1	Cooperate	2,2	0,3
	Defect	3,0	1,1

Figure 1. Mark’s Prisoner’s Dilemma Game.

rather than the cumulative payoff to that individual. Cultural transmission only occurs when the paired actors have different fitness levels: after a pair engages and payoffs are awarded, the partner who entered the encounter with a lower fitness level will copy the behavior of the partner who entered the encounter with the higher fitness level. The change in strategy occurs after both actors have played their strategies (i.e., a low fitness partner cannot know and, therefore, cannot copy the strategy of a high fitness partner until strategies are revealed). After playing their strategies, actors retain the fitness they acquired into the next round.

Figure 2 illustrates how pairings lead to cultural adoption. The rows and columns list all possible combinations of strategies and fitness levels available to the actors. Cell values represent the strategy and fitness level of the row player after an interaction with the column player. For example, if a D1 (row player) is paired with a C2 (column player), the D1 will become a C3: i.e., D1 earned a fitness level of 3 by defecting on a cooperator, while then adopting the C2's behavior because $2 > 1$. This combination of payoff structure and behavioral adoption dynamics converges to cooperative populations even when the initial population contains 99% defectors and 1% cooperators. Mark claims, "This model illustrates how the disproportionate prior exposure inherent to cultural transmission logically implies evolutionary force toward cooperation" (p. 332).

	C0	C2	C3	D0	D1	D3
C0	C2	C2	C2**	C0	D0	D0
C2	C2	C2	C2**	C0	C0	D0
C3	C2	C2	C2	C0	C0	C0
D0	D3	C3	C3*	D1	D1	D1
D1	D3	C3	C3*	D1	D1	D1
D3	D3	D3	D3	D1	D1	D1

Figure 2. Player strategy and fitness updating Matrix in Mark's Model.

* The presence of an attribution error that biases the adoption of cooperation upward.

** The presence of an attribution error that biases the adoption of defection downward.

Adoption is enclosed in dark borders.

Mark further claims that his analysis "reveals the substantive reason cultural transmission favors the evolution of cooperation" (pp. 336–7). Defectors become cooperators only after interacting and benefiting from a cooperator's behavior. When a defector adopts cooperation, he maintains his high fitness level. As seen in Figure 2, this occurs when a D0 or D1 meets a C2 or C3 and becomes a C3. Conversely, when a C0 meets a D1 or D3, or when a C0 or C2 meets a D3, it becomes a D0 and the average fitness level of defectors decreases. Over time, this process causes the relative proportion of cooperators to increase because more actors benefit from meeting cooperators, who are more likely to have high fitness levels. Mark's explanation is correct on the surface, but it misses the more fundamental and problematic question: *Why* does this model produce cooperation? A closer look reveals that more work is required to explain "the prevalence of cooperation in human populations" (p. 325).

MECHANISM FOR STRATEGY SWITCHING

Mark's model converges to cooperation primarily because his mechanism for strategy switching has a systemic asymmetric bias that is at odds with the theoretical motivation of his own proposed mechanism. To justify using the high fitness level as motivation for cultural adoption, Mark cites empirical work that shows that the behavior of high status individuals is more likely to be imitated than that of low status individuals [Eisenberg-Berg and Geisheker 1979; Boyd and Richerson (1985); Katz and Lazarsfeld (1955)]. Research not cited in Mark's study suggests that those with higher status are thought to be more competent (Webster and Foschi 1988), and when individuals of high status are shown to be less competent, then the level of their influence decreases (Wagner *et al.* 1996). Mark's implied assumption is that the observed fitness level and behavior are associated, so the behavior of higher status actors is imitated. However, in Mark's model, fitness level and strategy are decoupled. When a D0 or D1 meets a C3 with a high fitness level (that could have resulted only from a prior defection), the D0 or D1 copies the current behavior (cooperation) of this higher fitness partner, even though this fitness is the outcome of a prior defection. It is as if the D0 or D1 incorrectly

attributes the high fitness to the observed behavior. Asterisks in the bottom left quadrant of Figure 2 mark instances of this type of “attribution error.” If switching did not occur in these instances, there would be fewer conversions to cooperation. Double asterisks in the top left quadrant of Figure 2 mark a different but related situation: cooperators (C0 and C2) remain cooperators after meeting more fit cooperators (C3) who achieved this high fitness level by defecting, thus slowing the defection rate.

Our concern is not that strategy and observed fitness are decoupled *per se*, but that it is this decoupling that actually drives the convergence to cooperation. The success of Mark’s model necessitates these biases, which only occur when an individual is matched with a fitter cooperator who benefited by defecting on another’s cooperation. Limiting the certainty of decoupling prevents the system from achieving cooperation. Table 1 lists the minimum probability of a switch, when a low C or D and a C3 are paired, required to generate a cooperative equilibrium, at different starting population distributions. For example, for an initial population of 50% cooperators and 50% defectors to converge to cooperation, a decoupling of “status” and strategy must be present 49% of the time. When the initial proportion of population has 10% cooperators, the decoupling is required 98% of the time. The justification for using “fitness” as a trigger for adopting a new strategy is based on literature that links adoption of a behavior to status through an attribution that status is related to the behavior, but for convergence to cooperation, Mark’s model requires that fitness and strategy not be connected. We find this contradiction problematic.

POPULATION SIZES

As stated earlier, one compelling feature of Mark’s model is that, unlike many models, it converges to cooperation for large, randomly matched populations. Ironically, convergence

to cooperation is not guaranteed in smaller populations unless the initial population has a sufficient proportion of cooperators. Mark models interaction using difference equations, which are good approximations for the dynamics of random interaction of strangers without exit options in very large (infinite) populations, but not so for modeling random matching in smaller (finite) populations. Suppose the initial population consists of n individuals of which k are C0s and $(n-k)$ are D0s. If the proportion of Cs is small and, by the randomness of the matching, all C0s happen to be paired with D0s in the first round, then each C0 will remain a C0, and each D0 will either become a D1 or a D3. In the second round, unless the C0 is paired with another C0, each C0 will become a D0, each D matched with a C0 will be a D3, while all other D3s become D1s. After only two rounds, the system eliminates all cooperators. This elimination is more likely the smaller the population. The exact probability of the immediate elimination is:

$$\left[\prod_{j=0}^{k-1} \left(\frac{n-k-j}{n-1-j} \right) \right]^2.$$

Table 2.1a–c presents results from simulations with direct random matching consistent with the probabilities calculated above for $n = 500, 5000, \text{ or } 10,000$, with k fixed at 1% of n (i.e., $k = 0.01 * n$) as in Mark’s own benchmark. This shows that the probability of elimination grows as population size decreases, though an increase in the proportion of initial cooperators offsets the population size effect (as seen in Table 2.2a–c). So, although Mark’s model does converge to cooperation when populations are large even when they are dominated by defectors, this convergence is unlikely to occur in small populations unless there is a sufficiently large proportion of cooperators.

Other evolutionary game theory models have only been able to achieve cooperation when

Table 1. Minimum Attribution Error Required to Converge to All-C Given an Initial Starting Population

Initial Population%	1%	5%	10%	25%	50%	75%	99%
%C0	1%	5%	10%	25%	50%	75%	99%
%D0	99%	95%	90%	75%	50%	25%	1%
Minimum Attribution Error	0.9999	0.9954	0.9818	0.8776	0.4879	0.0480	0

Note: Calculated by modifying Mark’s difference equations to account for attribution errors. Contact the authors for details.

Table 2. Convergence Results from Simulations with Actual Randomness

	Mark's Model by Population Size			One Permutation from Mark (2002)		
	(a)	(b)	(c)	(d)	(e)	(f)
	n = 500	n = 5000	n = 10,000	Twice-in-row	Cumulative	
				2 Period	All Prior	
2.1 Defector start (1% C0, 99% D3)						
Converged to All-C	8%	61%	85%		41%	
Converged to All-D	92%	39%	15%		59%	
Average rounds to convergence (SD)	10.1 (27.7)	85.3 (68.0)	121.6 (53.3)		63.6 (72.6)	
Min. Iteration/Max. Iteration	3/177	3/264	3/252		3/187	
2.2 Random start						
Converged to All-C	100%	100%	100%			
Converged to All-D	—			100%		100%
Average rounds to convergence (SD)	27.9 (5.2)	23.1 (2.8)	27.6 (1.2)	23.1 (2.8)		40.5 (5.7)
Min. Iteration/Max. Iteration	14/49	17/38	23/32	17/38		31/81
2.3 Cooperator start (99% C0, 1% D3)						
Converged to All-C				—		
Converged to All-D				100%		100%
Average rounds to convergence (SD)				23.1 (2.8)		170.1 (14.5)
Min. Iteration/Max. Iteration				17/38		135/245

Note: Convergence results obtained from simulation data generated by authors using 1000 runs per game configuration. The Java simulation code is available from the authors upon request. All-C denotes “all donators” under altruism payoffs in column (d).

populations are small. Therefore, researchers have argued that, because humans evolved in relatively small groups (Leimar and Hammerstein 2001), models that allow for the evolution of cooperation in small groups are useful and may mimic how cooperation in human populations evolved. Finding a mechanism that fosters cooperation in large populations is an achievement. However, its relevance for explaining the evolution of cooperation is ambiguous if it cannot also explain the evolution of cooperation in small populations, where cooperation is thought to have evolved.

THE "DISPROPORTIONATE" IN DPE

The concerns above focus on Mark's model and should not be thought of as an evaluation of DPE, in general. In fact, Mark's model could better be described as a "higher status exposure" model, because only one exposure to a higher status individual is sufficient for strategy switching. In this sense, Mark's model does not capture DPE (p. 332), because in his model, the decision horizon is based only one interaction.

One way to capture the longer decision horizon implied by "disproportionate" prior exposure is by keeping the PD payoffs but requiring that an actor be matched with a fitter partner of a different strategy twice in a row, rather than once, to induce a conversion. As seen in Table 2(d), convergence to cooperation is now unattainable. A defector has to meet a fitter cooperator twice in a row to switch, but after meeting a cooperator the first time, the defector becomes a D3. There is no possibility to meet a cooperator with fitness level higher than 3, therefore, this defector cannot be converted in the next round.

A second way to broaden the horizon of actors is to define the fitness level as the cumulative score obtained through more than one payoff. By encompassing an actor's history rather than the snap shot of one event, cumulative scores better represent the impact of strategies. Earlier work on the evolution of cooperation (cited above) focuses on cumulative scores for this reason. As shown in Figure 3a, if we begin with a random distribution of strategies all with a fitness level of zero, matches produce four strategy payoff combinations. When these strategies are paired and their payoffs add to (Figure 3b), rather than replace (Figure 2), their prior fitness, there is no advantage to cooperators (cooperators comprise fewer than half of

the population). Figure 3c illustrates the fitness behavior pairings that are possible after the third iteration. Although the actual proportion of defectors depends on the random pairing process, this figure demonstrates the potential for defectors to gain high fitness levels relative to cooperators. Although not obvious from Figure 3, Table 2e-f illustrates that this process converges to defection. For Mark's benchmark of initially 99% defectors and 1% cooperators, extending the payoff window to two periods, results in a system that converges to cooperation only 41% of the time. Further extending the payoff window to include all payoffs dooms cooperation, even with initially 99% cooperators and 1% defectors. Defectors can achieve high fitness levels rapidly, maintain their high fitness levels, and never meet fitter cooperators. More generally, a player's fitness could consist of the sum of payoffs from the last x rounds. Under this formulation, the chances of converging to cooperation diminish as x increases. A slightly different formulation would have a player's fitness be a discounted sum of past payoffs, where payoffs from the distant past are discounted more heavily. Again, increasing the discount factor decreases the chances of converging to cooperation.

While our analysis implies that Mark's model cannot account for the evolution of cooperation except when there is a large initial population

(a)		D0	C0				
	D0	D1	D3				
	C0	C0	C2				
(b)		D1	D3	C0	C2		
	D1	D2	D2	D4	C4		
	D3	D4	D4	D6	D6		
	C0	D0	D0	C2	C2		
	C2	C2	D2	C4	C4		
(c)		D0	D2	D4	D6	C2	C4
	D0	D1	D1	D1	D1	C3	C3
	D2	D3	D3	D3	D3	D5	C5
	D4	D5	D5	D5	D5	D7	D7
	D6	D7	D7	D7	D7	D9	D9
	C2	C2	C2	D2	D2	C4	C4
	C4	C4	C4	D4	D4	C6	C6

Figure 3. Player-strategy and fitness updating matrix when fitness is cumulative for three iterations.

of cooperators, the status switching mechanism may still be relevant in cultural transmission, where there is no decoupling of fitness and strategy. From a game theory perspective, language transmission is a coordination game and not a mixed motive (PD) game: both parties benefit only when speaking the same language, and adopting a new language is advantageous only if a partner speaks it. Additionally, language adoption is a learning process that requires repeated exposure so the extended time horizons implied by DPE are reasonable. However, in a PD game, copying the behavior of those from whom an actor has benefited is less likely to generate positive outcomes unless the population is sufficiently cooperative. Hence Mark's proposed mechanism may be useful for modeling the transmission of some cultural traits, but is not be as universal as Mark suggests.

CONCLUSION

Despite this model's sensitivity to randomness in small populations and dependence on short time horizons, Mark's work raises a number of interesting issues. The DPE/status model provides valuable insights into a large class of interesting phenomena: cultural transmission arising in coordination games where the processes do not exhibit asymmetric attribution biases. A systematic look at these biases may provide interesting insights into the dynamics of cultural transmission. Similar models, which focus more on disproportion than status as the criterion for transmission, might provide a contrasting result that sheds more light on the role of culture in the evolution of cooperation. Even if DPE is not an appropriate mechanism to model the evolution of cooperation, it is likely that culture will be an important component in the story of cooperation. For example, explicitly using networks to study the effects of reputation on cooperation is another promising approach for studying cultural transmission.

Elisa Jayne Bienenstock is an Associate at the consulting firm of Booz Allen Hamilton where she applies mathematical sociology to national security issues. Currently her work focuses on designing methods to eliminate bias in task groups and developing predictive models based on a combination of network analysis, decision analysis, and game theory.

Michael McBride is Assistant Professor of Economics at the University of California, Irvine. His current research uses game theory to study the formation of social networks and the causes of political instability in sub-Saharan Africa. He is currently conducting laboratory experiments to test theoretical predictions of the provision of public goods when individuals do not know how many contributions are needed to provide the good. His article on statistically testing the relationship between income and subjective well-being appeared in the *Journal of Economic Behavior and Organization* (2001).

REFERENCES

- Leimar, O. and P. Hammerstein. 2001. "Evolution of Cooperation Through Indirect Reciprocity." *Proceedings of the Royal Society* V268:745–753.
- Mark, N. P. 2002. "Cultural Transmission, Disproportionate Prior Exposure, and the Evolution of Cooperation." *American Sociological Review*. V67:323–344.
- Nowak, M. A. and K. Sigmund. 1998a. "The dynamics of indirect reciprocity." *Journal of Theoretical Biology*. V194:561–574.
- . 1998b. "Evolution of indirect reciprocity by image scoring." *Nature* V393:573–578.
- Wagner, David G., Rebecca S. Ford, and Thomas W. Ford. 1996. "Can Gender Inequalities Be Reduced?" *American Sociological Review* 51:47–61.
- Webster, Murray Jr. and Martha Foschi. 1988. "Overview of Status Generalization." Pp. 1–20 in *Status Generalization: New Theory and Research*, edited by M. Webster, Jr. and M. Foschi. Stanford, California: Stanford University Press.
- Wilson, D. S. 1989. "Levels of Selection: An Alternative to Individualism in Biology and the Human Sciences." *Social Networks* 11:257–272.