Humans live in two worlds – one of personal, the other of impersonal exchange; chimps, capuchins, and so on, live only in the former (de Waal 2005). Humans miraculously leveraged the former, the world of personal exchange, into cultures that enabled our ancestors to settle almost all the globe before they became herder-farmers.

Henrich et al.'s monumental and valuable cross-cultural study is *not* about human sociality – equity (equal opportunity), reciprocity, and sharing across time in personal exchange. It is about rejecting selfishness defined as choosing dominant outcomes independent of context, history, and the future. It is about attacking this narrow and misguided view of the self interest in social interactions, and shades of standard externality theory. It is also about adding arguments to utility functions – equity, reciprocity, and "fairness" as preferences, not as norms allowing self-betterment through gains from exchange across time.

I have argued (Smith 1998; 2003) that "property;" that is, human rights to act, arose endogenously by consent in social exchange; that norms like reciprocity – you have to give in order to receive – long preceded market exchange; and that these ancient emergent "laws" continue alive and well today, as revealed in the universal human expression, "I owe you one," in which A voluntarily acknowledges a debt to B for favors. Negative reciprocity is the endogenous policeman that reminds of debts not paid. It dictates punishment when the expectations of favors are not fulfilled.

These rules survive in the laboratory when subjects move in single-play trust games (TG) with or without punishment options, and cooperate with monotonous increasing average frequency in repeat play, as we vary the probability of being matched with the same person from zero to unity (McCabe et al. 1996). Also, when B sees what player A forgoes when risking defection by B, it doubles the cooperative response compared with B seeing that A has no alternative but to pass to B. This cannot be due to preferences regarding one's own and others' payoffs (McCabe et al. 2003). I see behavior in more austere Ultimatum Game (UG), Dictator Game (DG), and Public Good Games (PGG) as flatland projections of richer TG environments; the former are too easily over-interpreted in terms of utility rather than sociality. The evidence that children and young adults lean toward dominate outcome choices, and are vulnerable to drug abuse, suggests the likelihood that the superior benefits of delayed gratification from reciprocity have not yet become fully integrated into experience and behavior.

Our TG studies primed me for the central claim here that impersonal market exchange enhances, rather than crowds out, personal exchange as a viable hypothesis, credible if not predictable from extant theory. The cross-cultural data presented by Henrich et al. are consistent with this hypothesis, and the authors have done an excellent job of answering my earlier criticism that I believed their study suffered from a confounding of context and culture. As Henrich et al. note, the ad hoc need to adapt procedures to the field realities led to several variations, and these turn out to not materially affect the behavior. For future studies, I think they are now in a position to make such variations part of the design.

I believe such exercises in *deliberate* variation in the defined context are especially important in cross-cultural studies, in which there can be no presumption, given the wide variation in the techniques needed to convey comprehension, that different experimenters using different language, formats, descriptors, protocols, currencies, group sizes, and settings are all implementing the same "abstract game." Given all this variation, I don't know what it means to say the "researchers stuck to entirely abstract explanations of the game" (target article, sect. 4.3, para. 2). There is no such thing as a context-free experiment. All memory is autobiographical and context dependent, and our brains draw on that memory when deciding in an unfamiliar situation. What the experimenter thinks is "neutral," may not be so for the subjects. Hence the need to vary protocols and to empirically determine what is or is not a contextual "treatment." Recall that the canonical model predicts that none of these issues matter, yet they are obviously a consequence of variation in human sociality. When theory fails, it ceases to provide any guidelines as to how to understand "abstract" play, and you are back to the starting point.

These considerations are underlined by the wide variation found in our (Hoffman et al 1994) studies of UG – mean proposals varied from 27.8% to 44.1% with undergraduates. Our variations shifted distributions, without altering their basic shapes, which I think is an interesting and unique feature of the cross-cultural data. But how is this affected by deliberately varying protocols? I assume that field anthropologists have some conjectures as to what in a particular culture might account for these "outlying" results, and through deliberate design variations they can increase our understanding of the culture. For our North American culture, we explored "earned rights" and the buy-sell context, and we were surprised that we could move the data across such a range with these treatments and their conjunction. Also, we conjectured that the social connection – we called it "distance" – may not have gone far enough in the DG (Hoffman et al 1996b). We were surprised, not just that the "double blind" made a large economic difference – giving nothing rose from 20% to 64% – but that, by relaxing the double blind conditions incrementally, the results stepped down incrementally to the baseline.

For DG, the authors should consult Cherry et al. (2002), who compare endowments given by the experimenter with money earned by the subjects: the percent of dictators giving nothing jumps from 19% to 79%. And the latter rises to 97% when double blinded. Strengthening the sense that the stakes belong to the dictator, combined with "no one can know" his or her decision, all but eliminates dictator giving. We don't know how playing with the money you earned impacts UG, PGG, and cross-cultural results; it would be interesting to find out.

Methods do matter: Variation in experimental methodologies and results

Richard Sosis

Department of Sociology and Anthropology, The Hebrew University of Jerusalem, Mt. Scopus, Jerusalem, 91905, Israel. **richard.sosis@uconn.edu**
www.anth.uconn.edu/faculty/sosis/

Abstract: Henrich et al.'s findings are generally consistent with the findings of other experimental field studies. However, the methodological variations deployed across field sites produced several systematic biases in their results. Hence, although the project is destined to become a watershed study, their results should be interpreted prudently.

Henrich and colleagues are to be commended for initiating what is surely a landmark interdisciplinary study that promises to have a profound impact on numerous fields in the social and biological sciences. Indeed, there is already a burgeoning anthropological and economic literature employing economic games among non-student populations (e.g., Carpenter et al. 2005; Lesorogol 2005b; Paciotti & Hadley 2003, Paciotti et al. 2005; Sosis & Ruffle 2003; 2004), some of which has been stimulated by Henrich's (2000) pioneering work among the Machiguenga. The findings of the target article are largely consistent with these studies: namely, non-student populations behave differently than students, the selfishness axiom is not supported, daily experiences influence game responses, and demographic factors have little impact on experimental decisions. The current study, as well as their companion co-edited volume, Henrich et al. (2004), will set the standard for the field. Nonetheless, I am concerned about the variance in methods employed across field sites and about how these variations influenced the experimental results.

Henrich et al. argue that the variation in their methodologies had little or no impact on the decision-making behavior of their subjects. Their data, however, cannot fully address this issue because there is no baseline data for any populations other than the Michigan and UCLA samples. We do not know whether method-

https://doi.org/10.1017/S0140525X05420142 Published online by Cambridge University Press

ological variation affects different groups differently (which seems likely and is a point Henrich et al. acknowledge), and what aspects of decision-making the methodological variations are likely to affect. We can assume that group differences emerging from identical protocols (Mapuche, Machiguenga, and UCLA; Quichua and Achuar) are real, but the lack of differences between (1) Kazaks and Mongols, (2) Au and Gnau, (3) Shona resettled and unsettled, and (4) Sangu herders and farmers could be either real or a consequence of methodological biases. Additional data will be necessary for this to be determined.

My methodological concerns stem from my own experience (or more precisely, inexperience) in conducting similar economic experiments. I have repeatedly found slight variations in methods to have generally unwanted but significant effects. I offer two illustrative examples. During research aimed at examining cooperation among Israeli kibbutzim, Bradley Ruffle and I conducted various pilot studies to refine our methods. To test several variations of the common-pool resource game, we conducted experiments with Israeli students and kibbutz members from several kibbutzim not in our research sample. Unexpectedly, we found that students had higher levels of cooperation than did kibbutz members. We realized that kibbutz members and students had different expectations, which presumably influenced their decisions. Kibbutz members had anticipated our experiments for about a week, since we had written and phoned them to arrange meeting times in their homes. In contrast, during the pilot studies at Ben Gurion University we simply approached students in the hallway and asked them to participate. These students had few expectations and consequently kept very little of their endowment (i.e., they exhibited high levels of cooperation). Post-experiment interviews among kibbutz members, however, revealed that they often retained part of their endowment (i.e., they behaved uncooperatively) because they did not want to leave the experiment without payment; since they had already invested time in scheduling the appointment with us, they expected to make some money and wanted to be assured of this outcome (they were not paid fees for showing up). Hence, they appeared less cooperative than Israeli university students. Once we controlled for this methodological disparity, the behavioral differences disappeared. Additional work has further shown that, as one would expect, when paired with fellow anonymous kibbutz members, kibbutzniks are more cooperative than city residents paired with (anonymous) city residents (Ruffle & Sosis, in press; Sosis & Ruffle 2004). It is not clear whether an "expectation" bias occurred in the Henrich et al. project, but, given that researchers obviously had different means of recruiting subjects and that some researchers conducted their experiments in a day while others carried them out over weeks, it is a possibility that should be examined.

Another methodological blunder on my part further illustrates the importance of consistent methodologies. In trust experiments I recently conducted at the University of Connecticut, I had difficulty recruiting students because in the standard trust experiment (Berg et al. 1995) only half the participants receive show-up fees. I decided to offer students pizza following the experiment, to encourage participation. During one session, the pizza arrived early and I (very foolishly) allowed all the subjects to eat before commencing the research. Consistent with numerous experimental results demonstrating that "cheap talk" increases prosociality (e.g., Orbell et al. 1988), this session, to my dismay, had significantly higher levels of trust than all the other sessions with similar student samples.

Henrich et al. claim that "there is no reliable correspondence between methodological variations across groups in the UG and their game behavior (compare Tables 2 and 4)" (sect. 4.3, para. 9). A comparison of these tables, however, suggests that there may be some systematic biases in their results. I will discuss each of the eight types of variation listed in Table 4 by column (which I denote as C1–C8). It is not clear how post-game interviews (C8) could influence game decisions (as they happen after the fact), nor is it likely that sham offers (C6) affect results significantly; otherwise we would be forced to dismiss the entire field of social psychology. Only Alvard used a non-monetary medium (C5), and although the Lamalera are the only group to offer more than 50% of their initial endowment, there are obviously not enough data to assess the influence of a non-monetary medium. Henrich et al. acknowledge that contextualizing (C2) experiments has an effect on game decisions, and they cite several studies in which this has been demonstrated, although it is not clear how contextualization influenced their results, if at all. There appears to be no consistent relationship between how the money was allocated (C1) and mean UG offers. However, in four remaining categories there appear to be systematic biases in the results. In Table 1, I rank the 15 studied groups by mean UG offer, as in Figure 2, and list the presence or absence of a show-up fee (C7), corralling with talking (C4), no corralling (C4), and individual instructions (C3; question marks indicate the category denoted as "both" in Table 4 of the target article). Similar to my pizza gaffe described above, we would expect higher prosociality when participants are in a communal setting and/or interact with each other prior to the experiment, as is the case when subjects are given the instructions as a group, when they are corralled, and, especially, when they are corralled with talking. We might also expect individuals who receive a show-up fee to be more generous, as they have already been guaranteed some payment. Henrich finds that show-up fees do not influence game decisions among U.S. populations; possibly U.S. students are more accustomed to participating in diverse experimental conditions and thus are less responsive to the presence or absence of pre-experimental payments. Regardless, the summary data in Table 1 here are consistent with all of these expectations and show systematic trends in the results that may be a consequence of methodological biases. I am not arguing that these biases, if real, diminish the significance of Henrich et al.'s main findings; I present them because I believe the authors' current results must be interpreted cautiously and so that future researchers employ consistent methods (as Henrich et al. have done with their second round of experiments).

Despite my concerns, I find Henrich et al.'s interpretation of their findings compelling and consistent with other research which has similarly shown that social institutions and daily life activities, ranging from economic pursuits to ritual practices, do impact game decisions (Paciotti & Hadley 2003; Sosis & Ruffle 2003; 2004). Many of the concerns raised here will likely be resolved in their second round of experiments, the results of which I eagerly await.

Table 1 (Sosis). *Group Rankings by UG Mean Offers and Methodological Variation*

Group Ranked by UG mean	Show-up fee	Corralling with talking	No corralling	Instructions to individuals only
Lamalera		X		
Ache	X	X		
Shona				
Orma	X			
Au	X	X		X
Achuar	X			
Sangu			X?	
Gnau	X	X		X
Tsimane	X			
Kazakh			X	X
Torguud			X	X
Mapuche			X	X
Hadza			X	X
Machiguenga			X?	X?
Quichua	X			

https://doi.org/10.1017/S0140525X05420142 Published online by Cambridge University Press