The Dual-Role Method and Ultimatum Game Performance

Shane J. Macfarlan

Abstract
The dual-role method (DRM) extracts subject preferences as two players in experimental economic games. Previous research indicates mixed effects of DRM on game performance. A two (role)-by-two (role order) design for the ultimatum game (UG) was implemented in a naturalistic setting across 2 years. Subjects played according to the role they were assigned (proposer or responder). Immediately after, they played as the other role. Meta-analyses show no mean effect of role order on performance across studies or between laboratory and naturalistic settings. Power analyses indicate sample sizes needed to find an effect using the DRM is larger than many small-scale populations.

Keywords
ultimatum game, dual-role method, Dominica

1 Department of Anthropology, Washington State University–Vancouver, Vancouver, WA, USA

Corresponding Author:
Shane J. Macfarlan, Department of Anthropology, Washington State University–Vancouver, VMMC 202, 14204 NE Salmon Creek Ave., Vancouver, WA 98686, USA
Email: shane.macfarlan@email.wsu.edu
Introduction
The use of experimental economic games to examine interdependent decision making has become prominent in the social sciences over the last 60 years (Colman 1995; Kagel and Roth 1995; Weber et al. 2004). Recently, anthropologists have begun to apply this framework to examine norms of fairness and economic preferences in cross-cultural settings (e.g., Henrich 2000; Kuznar 2001, 2002; Henrich et al. 2004; Marlowe 2004; Macfarlan and Quinlan 2008). Economic games place two or more individuals into cooperative or competitive settings structurally resembling a variety of real-life social situations that are of practical or theoretical importance. The rationale for economic games is that they are an efficient method for examining how people reason about social situations and they provide data that are objective and quantifiable (Colman 1982). For anthropologists, the goal of economic game research is to identify individual- and group-level variables influencing game performance within and across populations (Henrich et al. 2004). Variables influencing performance within a population shed light on locally adaptive behavior and preferences, whereas variables affecting performance across populations suggest panhuman adaptations for sociality. An underlying theme for economic game research is that it can be used to inform public policy (Roth 1995).

Experimentally minded anthropologists (e.g., Paciotti and Hadley 2003; Henrich et al. 2004a; Kuznar and Lutz 2007; Macfarlan and Quinlan 2008; Weissner 2009) and cross-cultural economists (e.g., Oosterbeek et al. 2004; Chen and Tang 2009) have given considerable attention to one game in particular, the ultimatum game (UG; Guth et al. 1982). The UG places participants into a bargaining type social setting and measures negative reciprocity—the extent to which one is willing to forgo a benefit to impose a cost on another (Camerer and Fehr 2004). The game involves two players—a proposer and a responder. The proposer is allotted a resource that she or he jointly owns with the responder and must decide how to divide it. The proposer can offer none to all of the resource. The responder can accept the offer (both participants receive their division) or reject it (neither party receives anything). When implemented within a population, responses are thought to reflect cultural models of fairness (Alvard 2004; Henrich et al. 2004; Marlowe 2004; Macfarlan and Quinlan 2008) but not necessarily real-world prosocial behavior (Gurven and Winking 2008). Two pieces of information generally are examined: (1) player 1’s offer; and (2) whether an offer was accepted or rejected.

When played using the strategy method (Oosterbeek et al. 2004), the information examined for responders is the minimal acceptable offer
(MAO)—the smallest increment of money player 2 would accept from player 1. The strategy method extracts how a subject would respond to every possible increment of the resource, rather than their reaction to a single offer. It is a preferred method because of its greater economy per data point (Brandts and Charness 2000). This method strengthens fairness considerations of responders (Guth and Tietz 1990), thus increasing the amount of money one is willing to reject (Oosterbeek et al. 2004).

Economists, psychologists, and anthropologists debate the appropriate methodological controls for experimental economics research, specifically the use of repetition and incentives (Hertwig and Ortmann 2001). The use of repetition and/or financial incentives is highly dependent on the questions one seeks to answer (Gil-White 2001). Repetition is known to affect subject behavior across a range of decision contexts (Hertwig and Ortmann 2001). Advocates of repetition suggest it allows subjects to learn the game environment, thus permitting more accurate tests of equilibrium point predictions (Roth 1995; Hertwig and Ortmann 2001). Opponents of repetition suggest that one-shot games better capture the frames of reference people bring to the game context (Henrich 2001). For the UG, repetition has mixed effects. Eckel and Grossman (2001) and Roth and colleagues (1991) show limited repetition effects on UG behavior, whereas Oosterbeek and colleagues (2004) demonstrate experience decreases proposal size but not rejection rates.

Additionally, the incentive structure of game environments can affect performance (Hertwig and Ortmann 2001; Oosterbeek et al. 2004). Incentive considerations fall into two classes: (1) Does one use financial incentives at all; and (2) if yes, how much money should be allocated to stake the game. Advocates of financial incentives suggest pecuniary rewards make participants behave more thoroughly, with less variability, and that results have greater interpretability (Hertwig and Ortmann 2001). Opponents of financial incentives suggest this option decreases intrinsic motivation, is cheaper, and is more easily implemented (Thaler 1987). Ultimately, the use of monetary incentives is dependent on whether one seeks to understand actual behavior or normative reactions. When financial incentives are used, stake size has mixed effects on UG behavior. The meta-analysis by Oosterbeek and colleagues (2004) of UG behavior shows larger stake sizes cause both proposal sizes and rejection rates to decrease. However, Hoffman et al. (1996), Munier and Zaharia (2002), and Slonim and Roth (1998) find increasing stake sizes decreases rejection rates, while proposals (as proportions of the total stake) remain static.
Dual-Role Method and Bargaining Games

The dual-role method (DRM) is rarely used in economic game research. Methodologically, the DRM is an example of stationary replication; participants make decisions repeatedly in the same game context but from two different perspectives (Hertwig and Ortmann 2001). In the UG, the DRM allows subjects to play both as a proposer and as a responder, allowing a participant’s full complement of preferences to be known. The DRM is useful for anthropologists because it (1) increases statistical power when implemented in small populations; (2) allows researchers to assess consistency of individuals’ preferences (e.g., whether a subject who gives less as a proposer also accepts less as a responder; Guth et al. 1982; Carter and Irons 1991); and (3) can be cheaper when the marginal cost of collecting data from a single individual repeatedly is lower than collecting it from additional subjects (Baron 2001).

Effects of the DRM on game performance are mixed. In the original UG experiment, Guth and colleagues (1982) found participants exposed to the DRM increase offers compared to those playing one role. They did not show how the DRM affected MAOs, nor was the effect of role order on performance evaluated. Weg and Smith (1993) show no difference in offers made by individuals subjected to the DRM but do not show how MAOs are affected. In a review of UG behavior, Guth and Tietz (1990) suggest subjects exposed to the DRM make greater proportions of 50/50 splits; however, they do not show how the DRM affects MAOs nor do they examine how role order affects performance. Burks and colleagues (2003) show subjects exposed to the DRM, but not told they would play both roles prior to game play, perform similarly to those playing only one role in trust games. When subjects were told they would be subjected to the DRM prior to game play, trust and reciprocity behaviors decreased compared to playing a single role.

Given the limited and mixed experimental evidence of the DRM on performance, the lack of systematic analysis of the effect of the DRM on MAOs and the lack of experimental data concerning the DRM on game play in non-laboratory settings, I sought to determine how the DRM affected game performance in a naturalistic setting using a two (role)-by-two (role order) group contrast research design. Results suggest role order does not affect performance and the lack of an effect is consistent across laboratory and naturalistic settings.

Study Site

The Commonwealth of Dominica is a mountainous island located in the Caribbean Sea between the French islands of Guadeloupe and Martinique.
The village of Bwa Mawego is one of the most underdeveloped in the country. A 2007 census revealed the village has a population of 408 individuals, with 500–600 full- to part-time residents (Macfarlan and Quinlan 2008). Although inheritance is patrilineal, village life is matrifocal (Quinlan 2006). Villagers rely on a mix of slash and burn horticulture and small-scale commercial activities. The primary cash opportunity is distillation of bay oil, an essential oil used in the cosmetic industry (Quinlan 2004). Anthropologists have examined the village for 20 years (Flinn and England 1997); however, a systematic analysis of economic reasoning has yet to occur. The UG was introduced as one component of a multiphase project to examine the logic of cooperation and economic reasoning.

Method

The UG was implemented following the protocol of Macfarlan and Quinlan (2008). To determine the effect of the DRM in a naturalist setting, the game was played in subjects’ homes or local shops. Although laboratories provide greater control for researchers to isolate variables, the nature and effect sizes of results from field and laboratory settings are comparable across a range of domains (Anderson et al. 1999), and I sought to determine this for the DRM.

A two (Role: Proposer, Responder)-by-two (Role Order: Propose first Respond second; Respond first, Propose second), group-contrast design was implemented to determine whether the order a role was received affected game performance. Specifically, I was interested if a mean difference existed between individuals who proposed first or second and between those responding first or second.

Five field assistants (four researchers per year) and I implemented the experiment within the village on 1 day, in each of two field seasons (2007 and 2008). Each researcher was assigned one of five hamlets to walk through. When a researcher encountered a villager and that individual agreed to play, the villager was introduced to the game via a verbal script. If more than one person was present, one was picked to participate first, in private, so responses were anonymous. Each subject was told he or she would play the game anonymously with another individual from the village. Roles were assigned according to how the researcher encountered subjects: The first individual encountered was assigned to propose first (respond second); the next subject encountered responded first (proposed second). This method was alternated over the day. The game was incentivized with 10 Eastern Caribbean (EC) dollars decomposed into 10 1-dollar increments.
($3.75 U.S., or one-third of a day’s wage for unskilled labor). I chose to incentivize the game to remain consistent with anthropologists who have implemented games in small-scale communities (e.g., Henrich et al. 2004) and because this community is accustomed to monetary compensation for anthropological research.

Examples of game play were conducted until the subject understood the game. Proposers chose how much they wished to allocate to a responder. Responders played the game using the strategy method (Oosterbeek et al. 2004). After a subject made their decisions, they were asked to respond as the other role. Subjects were not informed prior to game play that they would play the game as both roles. After responding as the second role, subjects were told that their payout would be determined by their performance as the first role only.

Subjects were told not to discuss the game until they received their money at the end of the day. Once all researchers finished walking through the assigned hamlet, we reconvened and randomly assigned a proposer’s offer to a responder’s MAO. When offers were accepted, individuals were paid accordingly. All subjects were given three dollars for participating; however, they did not realize this during game play. So no one could determine performance (thus reducing animosity from stingy play), all were given an extra $2, making the lowest amount of money earned $5 EC. Most subjects received payments the same day. In both years, several did not receive payments until the next morning, because of time constraints (the author had to find every subject who participated).

Results

Because multiple experimenters implemented the game and the game was administered in two field seasons, methodological effects are examined first. Both male and female researchers administered the games. Researcher sex did not affect offers (2007: \( t = -0.18; \) two-tailed \( p = 0.86; n = 47 \); 2008: \( t = -0.50; \) two-tailed \( p = 0.62; n = 68 \)) or MAOs (2007: \( t = -1.0; \) two-tailed \( p = 0.30; n = 47 \); 2008: \( t = 0.53; \) two-tailed \( p = 0.60; n = 68 \)). Sex was dummy coded as \( 1 = \text{male}; 0 = \text{female} \) for all analyses. Analysis of variance (ANOVA) with least significant difference (LSD) multiple comparisons did not show any significant interaction effects between subject sex and research sex on proposals (2007: \( F = 0.87; \) two-tailed \( p = 0.46; \) 2008: \( F = 0.44; \) two-tailed \( p = 0.73 \)) or MAOs (2007: \( F = 1.52; \) \( p = 0.22 \); 2008: \( F = 0.62; \) \( p = 0.61 \)). Four researchers implemented the game in each year. ANOVA with LSD multiple comparisons show no differences
among the four researchers for proposals (2007: $F = 0.41; p = 0.75$; 2008: $F = 0.81; p = .50$) or MAOs in 2007 ($F = 1.56$; two-tailed $p = 0.21$); however, a significant difference did exist between researchers for MAOs in 2008 ($F = 5.7; p = 0.002$). This effect may represent differences between hamlets rather than researcher bias, as such the data are included for all analyses. The UG was performed in 2 years. There was no difference between years on proposals ($t = -0.69$; two-tailed $p = 0.49$; $n = 115$) or MAOs ($t = 1.84$; two-tailed $p = 0.07$; $n = 115$; year was dummy coded as: $1 = 2007$; $0 = 2008$). Nonparametric tests returned identical results.

Results for proposals and MAOs are presented in Table 1. The order that the game was played had no effect on proposals or MAOs. Subjects who proposed first made offers similar to those proposing second (2007: $t = 1.22$; two-tailed $p = 0.23$; $n = 47$; 2008: $t = -0.64$; two-tailed $p = 0.52$; $n = 68$). MAOs were not significantly different between participants who played the game as a responder first or second (2007: $t = 0.52$; two-tailed $p = 0.61$, $n = 47$; 2008: $t = -1.42$; two-tailed $p = 0.15$; $n = 68$).

Next, I calculated the weighted mean effect of the DRM on performance across years using meta-analysis. Meta-analyses combine results and pool effect size estimates across studies, thus providing synthesized effect sizes with considerably more statistical power and accurate assessments of relationships between variables than is possible with single studies alone (Lipsey and Wilson 2001; Rosenthal and DiMatteo 2001). Effect sizes and confidence intervals were calculated using MIX Meta-Analysis 1.7 (Bax et al. 2006). Conventional standards in meta-analytic research suggest effect sizes smaller than 0.2 should be considered weak (Lipsey and Wilson 2001).¹ When confidence intervals include zero, the effect size is statistically

### Table 1. Descriptive Statistics for UG Performance in Bwa Mawego

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>Mean</th>
<th>Median</th>
<th>Standard Deviation</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Propose 1st 2007</td>
<td>25</td>
<td>$4.16$</td>
<td>$4$</td>
<td>$1.91$</td>
<td>$1$</td>
<td>$10$</td>
</tr>
<tr>
<td>Propose 2nd 2007</td>
<td>22</td>
<td>$4.86$</td>
<td>$5$</td>
<td>$2.03$</td>
<td>$1$</td>
<td>$10$</td>
</tr>
<tr>
<td>Propose 1st 2008</td>
<td>38</td>
<td>$4.37$</td>
<td>$5$</td>
<td>$1.22$</td>
<td>$0$</td>
<td>$7$</td>
</tr>
<tr>
<td>Propose 2nd 2008</td>
<td>30</td>
<td>$4.17$</td>
<td>$5$</td>
<td>$1.37$</td>
<td>$1$</td>
<td>$7$</td>
</tr>
<tr>
<td>Respond 1st 2007</td>
<td>22</td>
<td>$1.36$ (0.29)</td>
<td>$1$ (0.30)</td>
<td>$1.60$ (0.33)</td>
<td>$0$ (0)</td>
<td>$6$ (0.85)</td>
</tr>
<tr>
<td>Respond 2nd 2007</td>
<td>25</td>
<td>$1.68$ (0.30)</td>
<td>$1$ (0.30)</td>
<td>$2.61$ (0.24)</td>
<td>$0$ (0)</td>
<td>$10$ (1.04)</td>
</tr>
<tr>
<td>Respond 1st 2008</td>
<td>30</td>
<td>$2.70$</td>
<td>$2.50$</td>
<td>$2.22$</td>
<td>$0$</td>
<td>$6$</td>
</tr>
<tr>
<td>Respond 2nd 2008</td>
<td>38</td>
<td>$1.92$</td>
<td>$1$</td>
<td>$2.21$</td>
<td>$0$</td>
<td>$7$</td>
</tr>
</tbody>
</table>

Note: ¹log10 transformed: A constant $= 1$ was added to 2007 acceptances before logging, to deal with acceptances of zero.
The weighted mean effect of the DRM on performance in Bwa Mawego is weakly positive for proposals ($d^1 = 0.05; 95\% \text{ CI} = -0.32$ to $0.42; p = 0.77$), weakly negative for responses ($d^1 = -0.15; 95\% \text{ CI} = -0.52$ to $0.22; p = 0.44$), and statistically insignificant for both.

Next, I compared the insignificant effect found in the naturalistic setting to that found in laboratory conditions. The only study I am aware using the DRM, manipulating role order, and providing means, standard deviations, and sample sizes for the UG is Weg and Smith (1993); however, sufficient information exists to reconstruct proposals only (Table 2). The direction of the effect for the DRM on proposals in laboratory settings is in the opposite direction found in Bwa Mawego; however, the overall effect is nonsignificant ($d^1 = -0.16; 95\% \text{ CI} = -0.49$ to $0.17; p = 0.33$). The weighted mean effect of the DRM on proposals using both the naturalistic and the laboratory settings is weakly negative and nonsignificant ($d^1 = -0.06; 95\% \text{ CI} = -0.31$ to $0.18; p = 0.63$).

Although the DRM did not affect proposals (in naturalistic and laboratory studies) or MAOs (in naturalist settings alone), it is possible that the studies were sufficiently underpowered to detect an effect (Cohen et al. 2003). Using the weighted mean effect sizes generated from the meta-analysis, a power analysis (G*Power 3.0; see Faul et al. 2007) was performed to determine the sample sizes required to find an effect of the DRM on performance. To find a statistically significant effect using the DRM ($\alpha = 0.05$ and $\beta = 0.95$), a sample of 13,972 is required for proposals and 1,780 for MAOs.

### Table 2. Descriptive Statistics for UG Play by Weg and Smith (1993)

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>Mean</th>
<th>Median</th>
<th>Standard Deviation</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Propose 1st (Standard UG)</td>
<td>24</td>
<td>$3.95$</td>
<td>$4$</td>
<td>$1.12$</td>
<td>$1$</td>
<td>$5$</td>
</tr>
<tr>
<td>Propose 2nd (Standard UG)</td>
<td>24</td>
<td>$3.58$</td>
<td>$4$</td>
<td>$1.21$</td>
<td>$1$</td>
<td>$5$</td>
</tr>
<tr>
<td>Propose 1st (Modified UG: One-Two Game)</td>
<td>24</td>
<td>$4.29$</td>
<td>$5$</td>
<td>$1.04$</td>
<td>$1$</td>
<td>$5$</td>
</tr>
<tr>
<td>Propose 2nd (Modified UG: One-Two Game)</td>
<td>24</td>
<td>$4.21$</td>
<td>$5$</td>
<td>$1.06$</td>
<td>$2$</td>
<td>$5$</td>
</tr>
<tr>
<td>Propose 1st (Modified UG: One-One Game)</td>
<td>24</td>
<td>$3.67$</td>
<td>$4$</td>
<td>$0.96$</td>
<td>$1$</td>
<td>$5$</td>
</tr>
<tr>
<td>Propose 2nd (Modified UG: One-One Game)</td>
<td>24</td>
<td>$3.58$</td>
<td>$4$</td>
<td>$0.93$</td>
<td>$1$</td>
<td>$5$</td>
</tr>
</tbody>
</table>
Discussion and Conclusion

I conducted this study to determine (1) how the DRM affected UG performance by varying the order a role was received; and (2) whether this effect was consistent between laboratory and naturalistic settings. A two-by-two group contrast research design was used to determine the first effect. No statistically significant effects were uncovered between role and role order on UG performance in 2007 or 2008 in Bwa Mawego. Meta-analyses indicate the direction of the weighted mean effect for proposals in naturalistic and laboratory settings are in opposite directions; however, neither effect is statistically significant. The combined meta-analytic results suggest the weighted mean effect of the DRM is weakly negative and statistically insignificant for proposals and MAOs. This is the first study I know to report the direction and effect size of the DRM (varying role order) on proposals in a naturalistic setting and the first to report any effect size estimate for MAOs. The DRM appears to have no effect on proposals or MAOs in the UG in laboratory or naturalistic settings.

Null results are not evidence the effect does not exist. Power analyses indicate an effect could exist between role and role order on game performance; however, the sample sizes needed to find an effect is larger than the village of Bwa Mawego. I suggest the DRM is a practical way to extract greater information from the UG in small populations (<1,700 individuals). Researchers working with samples larger than 1,700 individuals should be cautious about potential DRM effects on game performance.

Research suggests repetition can affect game performance (Hertwig and Ortmann 2001). Because subjects play both roles in sequential order, it is possible repetition could cause performance to change from one condition to the next. However, results indicate the DRM has no effect on performance. Additionally, research suggests financial incentives (or the lack thereof) can affect subject performance (Hertwig and Ortmann 2001; Oosterbeek et al. 2004). It is possible the DRM affects performance because it requires subjects to play a portion of the game for no money. However, Burks and colleagues (2003) show participants who do not know they will be subjected to the DRM, perform statistically similar to those playing a single role. I find similar results. Incentivized, first-role UG behavior is statistically similar to non-incentivized, second-role behavior. The lack of repetition and second-role incentive effects may be due to subjects eliciting a single “cooperative” frame with which to make decisions across UG contexts that is robust to changes in game parameterization.

The DRM is an appropriate method for anthropologists working in small populations to extract economic preferences as long as subjects do not know
they will play the UG using this method. The method has three benefits; it (1) increases statistical power by extracting subjects’ preferences for both roles; (2) allows researchers to determine the consistency of those preferences; and (3) is more economical when the marginal cost of collecting data from a single individual repeatedly is lower than that of additional subjects.

**Acknowledgment**

The author thanks Rob Quinlan, Celeste Henrickson, Heather Bonander, Mark Remiker, and Davide Ponzi for administering the game and the people of Bwa Mawego for participating. Additional thanks go to Rob Quinlan for staking a portion of both games and to Larry Kuznar and three additional reviewers for their valuable comments on the manuscript.

**Declaration of Conflicting Interests**

The author(s) declared no potential conflicts of interest with respect to the authorship and/or publication of this article.

**Funding**

The author(s) received no financial support for the research and/or authorship of this article.

**Note**

1. Fixed Effects Model.

**References**


