View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Research Papers in Economics

Volunteering a Public Service:
An Experimental Investigation

Marc Bilodeau
Indiana University - Purdue University at Indianapolis

Jason Childs
University of New Brunswick at Saint John

Stuart Mestelman
McMaster University

Abstract: In some public goods environments it may be advantageous for heterogeneous groups to be coordinated by a single individual. This "volunteer" will bear private costs for acting as the leader while enabling each member of the group to achieve maximum potential gains. This environment is modeled as a War of Attrition game in which everyone can wait for someone else to volunteer. Since these games generally have multiple Nash equilibria but a unique subgame-perfect equilibrium, we tested experimentally the predictive power of the subgame-perfection criterion. Our data contradict that subjects saw the subgame-perfect strategy combination as the obvious way to play the game. An alternative behavioral hypothesis – that subjects were unable to predict accurately how their opponents would play and tried to maximize their expected payoff – is proposed. This hypothesis fits the observed data generally well.

Corresponding Author: Stuart Mestelman
Department of Economics
McMaster University
Hamilton, Ontario L8S 4M4
Canada

Phone: 905-525-9140 x23822
Fax: 905-521-8232
E-mail: mestelma@mcmaster.ca

Keywords: Public good; Best-Shot; Subgame-Perfection; Experiments

JEL Classification: H41, C72, C91

10 September 2003

1. Introduction

Questions regarding voluntary contributions to a public good and the free-riding problem have so far generated a sizable amount of theoretical modeling and experimental investigation.[1] The simplest way to model the free-riding problem in public good provision is through a Prisoners' Dilemma type of game.[2] A more general approach models the public good provision decision as a game akin to a Cournot duopoly game. In this game, each player selects a contribution level toward the public good.[3] In these models, the public good being provided is an incremental public good in the sense that the quantity of the public good being consumed depends on the sum of individual contributions.

One question which has received comparatively little attention in the economic literature is the problem of providing a public good that can be produced through "weakest-link" or "best-shot" technologies.[4] The provision of a public service can often be considered a best-shot decision. Examples of such situations include driving for a car pool, organizing a fund-raising event, getting up at night to quiet a crying baby, or slaying the dragon that threatens the village.[5] In this type of situation, only one individual needs to bear the cost of providing a non-incremental service that will benefit everyone. Such a situation is best modeled as a game of

---

[1] See Ledyard (1995) for a comprehensive, but now dated, survey and evaluation of laboratory research, and Chapter 6 in Plott and Smith (2003), for a summary of recent results.

[2] An early example of this can be found in Buchanan (1967).

[3] The classic reference for this approach is Bergstrom, Blume and Varian (1986).

[4] See Hirshleifer (1983).

[5] The dragon-slaying example is from Bliss and Nalebuff (1984). Bilodeau and Slivinski (1996) also mention cleaning shared toilets and chairing an academic department.

Chicken instead of a Prisoners' Dilemma.[6]

One characteristic of a game of Chicken is that it generally has multiple equilibria. For example, a two-player game of Chicken has two pure strategy equilibria in which either player gives in and the other one does not (it also has a mixed strategy equilibrium in which each player gives in with some probability). To solve this type of game, therefore, one must select an equilibrium which appears more plausible than others. Unfortunately, other than the nebulous concept of a focal point, if any can be found, there is no general method for solving static games of Chicken.

The situation is different for the dynamic version of the game of Chicken, known as a War of Attrition.[7] In this game, each player becomes progressively more injured as time wears on until one gives in and stops the game. The player giving in first gets a lower payoff than he would have received if someone else had given in. In a public service provision context, each player may choose to wait before volunteering to provide the service, thus avoiding the cost of providing it if someone else volunteers in the meantime, but no one can enjoy the benefits of the service until someone provides it. Like games of Chicken, War of Attrition games also generally have multiple equilibria in pure and mixed strategies.[8] However, one widely used equilibrium selection criterion, applicable in multistage and dynamic games, is the concept of subgame

[6] The game of Chicken is described in several introductory game theory textbooks, for example in Rasmusen (1989, p. 73). The game of Chicken is also sometimes referred to as the Hawk-Dove game, particularly in evolutionary biology.

[7] War of Attrition games can be divided into two variants according to whether they are stationary or not. See Fudenberg and Tirole (1991, pp. 119-126) for a description and discussion of both cases. The game we investigate experimentally in this paper is non-stationary.

[8] See Hendricks, Weiss and Wilson (1988) for a complete characterization of the equilibria of the War of Attrition game in continuous time.

perfection.[9] In dynamic games, some equilibria are deemed implausible because they rest on non-credible threats. Subgame perfection rules them out by insisting that the equilibrium strategy be optimal for each player not only along the equilibrium path, but at every point in the game – even those that will never be reached. When a War of Attrition game is not stationary, subgame-perfection can be applied to select a unique pure strategy equilibrium. Given the wide acceptance of subgame-perfection as an equilibrium selection criterion, it is then tempting to single out the unique subgame-perfect equilibrium as the solution of the game. This is the approach taken by Bilodeau and Slivinski (1996). The present paper is an attempt to verify experimentally whether this conclusion is warranted.

Our experimental results suggest that it is not. In 472 three-player War of Attrition games in which there was a unique subgame-perfect equilibrium (SPE), the SPE prediction was approximately realized only 133 times. Moreover, we found no statistical difference in the distribution of volunteers between games that had a unique SPE and similar-looking games with multiple SPE. Given this underwhelming support for subgame-perfection, we suggest an ex-post behavioral hypothesis that better fits the data. This ex-post behavioral hypothesis is that the subjects fail to completely account for the strategic nature of the game they are involved in and approach it essentially as if they were playing against nature.

2. Modeling the Provision of a Public Service

In this section, we present a simplified version of the public service provision game

[9] The concept of subgame perfection is explained in several game theory textbooks, for example in Fudenberg and Tirole (1991, p. 92).

modeled in Bilodeau and Slivinski (1996).[10] Consider a group of n individuals who would all stand to benefit for some time if a public service was provided by one of them. Each must decide whether and when to volunteer to provide the public service. Suppose for simplicity that time is measured in discrete increments. Each individual i is characterized by four parameters. Let v_i be the payoff he receives each period until the service is provided and let u_i be the payoff he receives each period once the public service is provided. Let C_i be the cost of providing this service. This cost must be borne entirely by the volunteer at the time he provides the service and no side payments are possible. Finally, let T_i be the (finite) time horizon during which he could benefit from the public service.

If the individual volunteers at time $t \in [\, 0, T_i \,]$, his payoff is $tv_i + (T_i - t\,)u_i - C_i$, while if someone else volunteers at time t his payoff is $tv_i + (T_i - t\,)u_i$. Assume that $u_i - v_i > 0$ (everyone benefits from the public service), $C_i > 0$ (volunteering is costly), and $T_i u_i > C_i$ (volunteering is not a dominated strategy). Also assume that every player's rationality and all the parameters are common knowledge. Then, the game is a well-defined War of Attrition in discrete time with complete information.[11]

This game has Nash equilibria in which any one of the players volunteer immediately and all the others wait. Intuitively, if player i volunteers immediately then everyone else is better off

[10] Bilodeau and Slivinski (1996) model the public service provision decision as a War of Attrition in continuous time with discounting. They allow individuals to have different discount rates and allow the cost of providing the public service to include both a one time cost and an ongoing cost. They also allow the benefit of the service to vary depending on who provides it.

[11] For completeness, we also need to define the payoffs if two or more players volunteer simultaneously and if no one ever volunteers. We could do this in many ways; but the simplest formulation, assuming that both would incur the cost C_i if they volunteer simultaneously and that all will have a payoff of $T_i v_i$ if no one ever volunteers, will do.

waiting; and, if everyone else is waiting, then player i is better off volunteering immediately. The same holds for all players.

However, only one of these equilibria is selected by subgame-perfection. To show this, we note first that since the payoff if no one ever volunteers is $T_i v_i$, it is a dominant strategy for i not to volunteer at any t such that $tv_i + (T_i - t)u_i - C_i < T_i v_i$. Since the payoffs are decreasing in t, we can calculate from this the critical value $t_i^* = T_i - C_i/(u_i - v_i)$ beyond which i would never rationally volunteer. Without loss of generality, order the individuals such that $t_1^* > t_2^* \geq t_3^* \geq \ldots \geq t_n^*$.[12] Now, consider individual 1. If the game reaches t_2^* or any time between t_2^* and t_1^*, it is optimal for him to volunteer since by then no one else would. His payoff from volunteering would still exceed $T_i v_i$ and would continue to decrease if he waited any longer. Expecting individual 1 to volunteer at t_2^*, all others would then hold off on volunteering for some interval immediately preceding t_2^*. Therefore, individual 1 could do no better than to volunteer at any time in this interval. The others would then hold off volunteering in another preceding interval and so on until individual 1 volunteers at $t = 0$ and the others wait. The game has a unique SPE in which individual 1 volunteers immediately and all others abstain.[13]

The critical value t_i^* depends on the parameters that characterize each individual and is increasing in T_i, decreasing in C_i and increasing in $(u_i - v_i)$. So *ceteris paribus*, the individual

[12] If $t_1^* = t_2^*$ the subgame perfect equilibrium would not be unique. We will use that property to test whether subjects play differently in games with and without a unique SPE.

[13] To be precise, we should say that individual 1 volunteers at every t between 0 and t_1^* and abstains after that, while all others abstain at every t. A strategy in this game is a complete statement of what a player would do at every point in the game, including points off the equilibrium path.

with the longest time horizon, the lowest cost of providing the service, or who stands to benefit the most from the public service will be the one for whom the critical value t_i* is the largest. If the game unfolds according to the subgame perfect equilibrium prediction, this individual alone will volunteer immediately.

3. Laboratory Representation of the Public Service Provision Game

In the public service provision game, players receive a flow of payoffs each period, both before and after someone volunteers, until they reach the end of their time horizon. Since they receive a smaller payoff until the public service is provided, waiting is costly: The maximum possible payoff a player could receive is $T_i u_i$ if someone else volunteers immediately, and each period spent waiting for a volunteer reduces this maximum by $(u_i - v_i)$. To replicate this type of environment in the laboratory, subjects were given an initial endowment, E_i, and were told that this would be their payoff if someone took action immediately to stop the game. Subjects were also told that their payoff would decline at the rate of s_i per second until one of them took action to stop it. The cost of stopping the game, C_i, if the subject chose to volunteer, would then be deducted from his payoff. Stopping the decline of everyone's payoffs is a public service that the volunteer provides to everyone at a private cost to himself. The game played in the laboratory is strategically equivalent to the public service provision game outlined in section 2 if and only if $E_i = T_i u_i$ and $s_i = (u_i - v_i)$ for all players.

There were 14 sessions in this experiment run over the computer network at the McMaster University Experimental Economics Laboratory in Hamilton, Ontario, Canada. Twelve subjects participated in each session of the experiment and a total of 168 subjects were recruited from the student population of McMaster University. Each session of this experiment consisted of twelve games or rounds (these terms are used interchangeably).

In each round, subjects were randomly assigned to a group with two other subjects. These groupings changed every round and no two subjects were in the same group in consecutive rounds. At no time were subjects made aware of who were the other members of their group. The three parameters (E_i, s_i and C_i) which characterize the subjects were varied across subjects systematically.

Table 1 presents the five treatments in the experimental design. The third column, $E_{i,}$ is the subjects' initial endowment. The next column, $s_{i,}$ shows how much they lost each second until the game is stopped. The third column, C_i, lists the cost of providing the service. When the entry is a triplet the first value was assigned to subject A, the second to B, and the third to C. When there is one entry, all subjects had the same value for that parameter.

The subjects' endowments, costs and payoffs were initially reported in laboratory dollars (L$). Participants were told at the start of their sessions that their laboratory dollar payoffs from each of the rounds in which they participate will be added up at the end of their sessions and will be converted to Canadian dollars (C$) at the rate L$1 = C$0.06. Subjects earned an average of C$9.50 (the standard deviation was C$0.75) and the range was from C$11.08 to C$7.50. Participants also received a "show-up" payment of C$5.00 for arriving on time. Including the reading of instructions and answering questions before the rounds began, the twelve rounds were completed within forty-five minutes and an hour.

All games lasted a maximum of 90 seconds. In some treatments, the payoffs of some players decrease to zero before 90 seconds have elapsed. In this case, the payoffs stop declining and remain at zero until the end of the game. In other treatments the payoffs of some players are still positive at $t = 90$ seconds. This would then be their payoff if the game ends without a volunteer. A player could receive a negative payoff by volunteering at a time when his remaining payoff was less than C_i.

When a round begins, each subject's screen displays three graphs: one that represents his payoff and two representing the payoffs of the two other subjects with whom he was matched. Figure 1 presents the screen that would have been seen by player C (labeled "you") in a round of Treatment 3.1. His initial endowment was L$14, the second player had an initial endowment of L$17, and the third, an initial endowment of L$20. All three see their payoffs decline by L$0.20 per second until the game is stopped. Each can stop the game any time at a personal cost of L$5. The snapshot was taken 34 seconds after the payoffs had started declining (the timer shows 56 seconds remaining). At that point, player C's payoff was down to L$7.20. Once a player's payoff reaches 0 (after 70 seconds for player C and 85 seconds for the second player in this figure) his payoff stops decreasing. If time expires (at 90 seconds) without a volunteer, the first and second players in Figure 1 would receive 0, while the third player would receive L$2 since his payoff has not reached 0 yet.

On each of the three graphs is a light blue line representing the payoff this subject will receive if the service is provided by someone else. As time elapses, this line is filled in progressively with darker blue until someone stops the game. Below the blue line, each graph also contains a white line representing the subject's payoff if he volunteers. It extends below the horizontal axis to show that subjects could receive a negative payoff if they volunteer late in the game. This line is filled in with red as the round progresses until someone volunteers to stop it.

This gives the subjects a visual indication of their current payoffs at any point in the game. Located above the graphs is a button labeled "Action" on which subjects can click using their computer mouse if they wish to volunteer.

In each round, after the subjects' computers displayed the screens, subjects were given 15 seconds to make a decision or formulate a strategy before their payoffs started to decline. This period was intended to give subjects an opportunity to familiarize themselves with their own payoffs and the payoffs of the two other subjects with whom they were matched before having to make a decision. During these 15 seconds it was possible for one or more subjects to volunteer. Anyone volunteering during these 15 seconds of "frozen time" was considered to have volunteered at time 0. Subjects were not informed that someone had volunteered until the end of that 15 second period.[14] Further, subjects were never told whether there were multiple volunteers.

After this initial 15-second period, the payoffs started to decline until one of the subjects stopped the game by pushing the button labeled "Action" on his screen. In this experiment, we always referred to volunteering as "taking action" in order to prevent framing effects. Subjects were never informed of which subject had volunteered; they were only told that someone had.

The critical time t_i* is the point at which the payoff from volunteering becomes smaller than the payoff to letting time expire without a volunteer. For the first two players in Figure 1 this is the point where the white line crosses the horizontal axis, at 45 and 60 seconds respectively. They have no incentive to volunteer beyond that point, although nothing prevents

[14] Go to http://socserv2.socsci.mcmaster.ca/~econ/mceel/papers/bcmvpsin.pdf for the instructions read to and by participants.

them from pushing the "Action" button anyway and getting a negative payoff for this round. For the third player, the payoff to volunteering falls below L$2 after 65 seconds, so from that point on none of the subjects have incentives to volunteer anymore. The t_i* values for all treatments are reported in Table 1.

Looking at Table 1, we see that in treatments 1.1 to 2.4 subgame-perfection predicts that player A will volunteer. However, since all eight treatments make the same prediction, finding that player A is the most frequent volunteer wouldn't be convincing evidence in support of the subgame-perfection hypothesis because we wouldn't know whether player A volunteered because all the players saw the SPE strategies as the obvious way to play the game or for some other reason, e.g., simply because he has the lowest cost of volunteering or the highest cost of waiting. We designed the other treatments so that similar-looking games would yield different SPE predictions.

The games in treatments 3.1 to 3.4 are similar to each other. In each case, subjects' payoffs have the same cost of volunteering and the same slope (the cost of waiting) but different starting endowments. However, the games in treatments 3.1 and 3.2 have a unique SPE while those in 3.3 and 3.4 do not. This difference will be used to test whether subjects played differently in the games that have a unique SPE and those that do not.

In treatments 4.1 to 4.4, subjects' payoffs had the same endowment and the same cost of volunteering, but different slopes. However, in treatments 4.1 and 4.2 the unique SPE has subject B volunteering while in treatments 4.3 and 4.4 subject C is the predicted volunteer. This difference will be used to test whether subjects played differently in games that yielded different SPE predictions.

In treatments 5.1 to 5.4, subjects' payoffs had the same slope but different initial endowment and different costs of volunteering. All four treatments make different SPE predictions. Again, we will use these differences to test whether subjects played differently in these games.

Finally, in treatment 2.1 and 2.3 there is no Nash equilibrium in which player C volunteers because volunteering is a dominated strategy for him. This difference will be used to test whether subjects played differently in games where some had no incentive to ever volunteer.

4. Experimental Results

In the games that have a unique SPE, the subgame-perfection refinement yields a clear prediction: the subject with the largest t_i* value will volunteer *immediately* and all others will wait. The performance of the SPE prediction is summarized in Table 2.[15] In the 472 rounds of our experiments in which there was a unique SPE, this prediction was exactly realized only 97 times. This is an accuracy of just 20%. However, it could be argued that including the rounds in which the predicted subject volunteered shortly after time 0 is reasonable because in the laboratory environment it is not a large deviation for a subject to wait until the count-down actually begins before volunteering. Including the rounds in which the predicted subject volunteered after 1 or 2 seconds increases to 133 the number of rounds in which the SPE prediction was approximately realized. Even if we included the rounds in which the predicted subject waited more than two seconds before volunteering (there were 61 such rounds with an

[15] Go to http://socserv2.socsci.mcmaster.ca/~econ/mceel/papers/bcmvpsdata.txt for the data used in the analysis.

average stopping time of 13 seconds), this would still be a success rate of only about 41%. By comparison, subjects other than the one predicted by the subgame-perfection refinement volunteered first 273 times and 5 rounds had no volunteer. The identity of volunteers by treatment is presented in Table 3.

Nonetheless, the subject identified by the subgame-perfection refinement volunteered in almost 49% of the rounds[16]. In 11 of the 17 treatments that had a unique SPE, the predicted subject volunteered more often than the others, sometimes (e.g., in 1.4) up to three quarters of the time. So it is worth verifying whether it is the presence of a unique SPE or some other factor that is driving these observations.

We compared the distributions of volunteers observed in each sub-treatment of Treatments 1 through 5. Overwhelmingly, in all cases, we cannot reject the null hypothesis of no significant difference against the alternative hypothesis that there are significant differences among the distributions of volunteers within a treatment. In treatments 1 and 2, for which each sub-treatment had the same unique SPE, there was no significant difference in the pattern of play across sub-treatments ((χ^2 test, $p > 0.500$ for treatment 1 and $p > 0.300$ for treatment 2). In treatment 3, there was no significant difference in the pattern of play between games with a unique SPE and games with multiple SPE (χ^2 test, $p > 0.800$). In treatment 4, we observed the same pattern of play when subgame-perfection predicted that subject C would volunteer as when it predicted that subject B would volunteer (χ^2 test, $p > 0.100$). In treatment 5, we also observed no significant differences in the pattern of play even though each sub-treatment yielded different

[16] This includes 37 rounds in which the predicted subject volunteered simultaneously with at least one other.

SPE predictions (χ^2 test, $p > 0.300$). Therefore, even in the rounds where the SPE prediction was accurate, it appears that this was just a coincidence and that something else was guiding the subjects' decisions. Incidentally, even the presence of a dominated strategy didn't seem to affect the pattern of play: we observed the same patterns in treatments 2.1 and 2.3 where volunteering was a dominated strategy for player C as in treatments 2.2 and 2.4 in which volunteering was not dominated.

Table 4 reports the frequency distributions of the time elapsed before the first subject took "action" and Figure 2 shows the cumulative incidence of volunteering in the first 20 seconds of each game. Of the 544 rounds played, the median time was 0 (54% of the rounds were stopped at $t = 0$) and the median time for "action" in the remaining rounds occurred in the fourth second. Even if volunteering *immediately* can be extended through the first couple of seconds of a round, this leaves approximately 28% of the rounds without a volunteer after the first two seconds. Unless these subjects were playing a mixed strategy and randomizing their stopping time, this observation suggests that many subjects were either expecting someone else to stop the game and needed a few seconds to revise their strategy once they saw that the payoffs were continuing to decline, or, had deliberately chosen to wait some time in the hope that someone else would volunteer first. We will explore this possibility below.

5. Discussion

5.1 Subgame-Perfection

Three main reasons lead us to doubt the appropriateness of subgame-perfection as an equilibrium selection criterion in this experimental context.

First, subgame-perfection is an equilibrium selection criterion whose purpose is to eliminate implausible equilibria, namely those that rest on non-credible threats. For example, in a simple two-stage entry-deterrence game, the implausibility of the non-subgame-perfect equilibrium outcome is apparent to any casual observer.[17] In this case, predicting that this or some other similar game would unfold according to the SPE prediction seems warranted. However, in a War of Attrition game such as the one induced in our experiments, it is not clear that any of the Nash equilibria are implausible. Individual A taking action and the other two abstaining sounds just as plausible as individual B or C taking action and the other two abstaining. Instead, it is the application of subgame-perfection which seems rather implausible, as the mere statement of the subgame-perfect equilibrium strategy (player A volunteers at all t between 0 and t_A*) seems to defy common sense (a natural reaction to this statement is: how can player A volunteer at all t if he is stopping the game at t = 0?). Many subjects may have simply conceived their strategy as a particular stopping time, e.g., "wait 4 seconds," rather than as a conditional statement outlining what they would do at every point of the game if this point was reached, e.g., "wait 4 seconds then push the action button, but if for some reason that doesn't stop the game then continue pushing the action button repeatedly every second until 65 seconds have elapsed and stop pushing the action button after that." Unless subjects defined their strategies in this fashion in their mind (and we doubt that they did), they would be unable to even notice that a particular strategy combination may not be best responses to each other, say,

---

[17] A description of the two-stage entry-deterrence game can be found in several introductory game theory textbooks, for example in Rasmusen (1989, pp.85-87). This game has two Nash equilibria in pure strategy (Stay Out; Fight) and (Enter; Collude); however, only the second is subgame-perfect.

between the 60th and 65th seconds.

Second, even if they had defined their strategy sets correctly in their minds, identifying the subgame-perfect equilibrium strategies may be too difficult for most subjects. Contrary to a simple Prisoner's Dilemma, subjects do not have a dominant strategy in a War of Attrition game. Selecting a *best response* to every strategy combination by the other players requires recalculating an optimization problem for each of their strategies. Figuring out an *equilibrium strategy combination* is even more complex than merely selecting a best response. It requires that each player solve the game not only from his point of view but also from the point of view of all the other players; and, they must verify that each is playing a best response to everyone else. Verifying *subgame-perfection* adds an even thicker layer of complexity by requiring that the players figure out what everyone would rationally do at every point in the game, even those they are convinced will never be reached. This may be too complex a task for many subjects, especially in a laboratory environment in which they only have 15 seconds to decide on a strategy.[18]

Third, even though we provide subjects with complete information about everyone's payoffs, we can never truly elicit a game of *complete information* between them because it is impossible to insure that everyone's rationality is common knowledge (i.e., all members of a group believe that all other members of their group are rational, and their beliefs about others' beliefs are equally certain and known up to an infinite regress). If subjects have any doubt about

[18] Past experimental research has so far failed to find convincing evidence that subjects in sequential decision games systematically play subgame-perfect equilibrium strategies; this is particularly true for complex games or games involving long decision chains. See for example Roth (1995) and Davis and Holt (1993, pp.102-109).

another's rationality or about whether anyone doubts anyone else's rationality, they will be unable to predict each other's best responses accurately. But the ability to accurately predict everyone's rational best responses to actions off the equilibrium path is crucial to the application of subgame-perfection, and the subjects who participated in these experiments almost certainly lacked this ability. In fact, before beginning these experiments none of us knew with certainty how the subjects would behave. So, it seems likely that the subjects who participated in the sessions were at least as ignorant as we were about how their opponents would behave. But if a player is unsure of how the other players will behave, how should he rationally play this game? The answer is definitely not to just play blindly the strategy identified by the subgame-perfection refinement. Instead, he would use whatever beliefs he has about the likelihood that other players will behave in certain ways to determine an optimal strategy.

Since our experimental results confirm these doubts concerning the accuracy of the prediction that the game will unfold according to the subgame-perfect equilibrium, we suggest an alternative hypothesis about how subjects may behave in this environment. It rests on the hypothesis that instead of looking backward from the end of the game to unravel the equilibria that do not meet the subgame perfection criterion, subjects will simply compare the expected benefits and costs of waiting for some time at the beginning of the game.[19]

---

[19] Thaler (2000) predicts an increase in the use of descriptive theories such as this in economics. Descriptive theories are based on the observance of empirical regularities in the behavior of human subjects and the context of the economic problem being considered.

5.2. Play Against Nature [20]

If subjects are unable to quickly and accurately forecast each other's strategies, they will approach this game as if they were playing against an unpredictable opponent instead of against two other *purposeful* players and will simply choose a waiting time that equates the expected cost and benefit of waiting an additional second.[21] The benefit of waiting is that it increases the probability that someone else will stop the game first and therefore that the player will avoid the cost of volunteering. The cost of waiting is the loss of payoff if no one else has stopped the game by then.

Suppose that Nature takes action randomly at some time t according to a (discrete) probability distribution $g(t)$ over $[0, T_i]$ known by i.[22] We could also interpret $g(t)$ as the subjective belief that someone else will volunteer at time t. There may be a positive probability that nature takes action at $t = 0$ and so $g(0) \geq 0$. Furthermore, $g(T_i) \geq 0$ allows for a positive probability that nature will not take action during individual i's time horizon. Normalizing

---

[20] Another alternative hypothesis is that subjects may have been attempting to coordinate their actions around a number of seemingly relevant focal points. We evaluate this hypothesis in an appendix that we are not including in this paper due to space constraints, but which is available at http://socserv2.socsci.mcmaster.ca/~econ/mceel/papers/bcmvpsfocal.pdf . The "focal point" hypothesis does not explain the data as well as the "play against nature" hypothesis.

[21] This hypothesis is similar to what Nagel (1995) calls "first order beliefs." Players who hold first order beliefs assume that others behave randomly and choose their best response to this random behavior. Nagel (1995) contrasts this to zero-order beliefs, in which players simply pick a strategy at random without forming beliefs about what the others will do, and *n*th-order beliefs in which the players reason deeply enough to form beliefs about the (n-1)th order beliefs of the others.

[22] If the individual did not know $g(t)$, we could reinterpret it as his belief about the probability that nature would take action at time t. Since we would then have to consider the possibility that the player could update these beliefs every second of the game, this would complicate the analysis without shedding additional light on the question at hand.

payoffs so that $v_i = 0$ for simplicity, individual i's payoff from taking action at time t against an opponent who plays according to $g(t)$ is:

$$P_i(t) = \{1 - \sum_{j=0}^{t-1} g(j)\}[\, s_i(T_i - t) - C_i] + \sum_{j=0}^{t-1} g(j)\, [\, s_i (T_i - j)] \qquad (9)$$

The first term on the right-hand side is the expected payoff if nature has not taken action by time t and individual i takes action himself. In this case, he pays the cost C_i and receives s_i from time t through T_i. The second term is the expected payoff if nature takes action before time t. In this case the individual receives s_i from that time on through T_i. It is noteworthy that $P_i(t)$ is not necessarily decreasing monotonically in t, so volunteering at $t = 0$ is not necessarily optimal. Ex-ante, given their incomplete information, it may even be optimal for everyone to wait before taking action or even for everyone to volunteer at $t = 0$. The specific form of the function depends upon $g(t)$.

Given this expected payoff function, the individual will choose to take action at the time t which maximizes $P_i(t)$. The expected gain from waiting one more second is:

$$P_i(t+1) - P_i(t) = g(t)C_i - \{1 - \sum_{j=0}^{t} g(j)\}s_i \qquad (10)$$

The first term to the right of the equal sign contains the probability of avoiding the cost C_i by waiting one more second and the second term contains the probability of losing the return s_i if nature does not act by time t. The trade-off between taking action and waiting is clear: Each second the individual waits costs him s_i if nature does not take action; however, this increases the probability that nature will take action in the next second sparing him the cost C_i. *Ceteris paribus*, the larger C_i, the longer an individual will choose to wait before taking action, and the larger s_i, the shorter the time the individual will choose to wait before taking action. It is noteworthy that the time horizon, T_i, and the initial endowment, E_i, do not appear in the equation. When individuals have incomplete information, or are unable to accurately forecast others' strategies, the decision of how long to wait before taking action depends only on the expected gain and cost of waiting and T_i and E_i are irrelevant to this decision.

If individuals have the same beliefs, this hypothesis leads to the following predictions: In the treatments in which individuals differ with respect to the cost of volunteering, C_i, (Treatments 1 and 5) the one with the smallest C_i will wait the shortest time before taking action. In Treatments 2 and 4 where individuals differ with respect to the cost of waiting, the individual with the highest cost of waiting will wait the least. If the individuals do not differ with respect to C_i or s_i (Treatment 3), we cannot predict who will wait the least, thus a default prediction of random behavior is maintained.

In the 224 rounds in which subjects differed according to C_i (Treatments 1 and 5), the subject with the lowest C_i volunteered 145 times, the subject with the middle C_i volunteered 67 times, and the subject with the highest C_i volunteered only 46 times. In the 224 rounds in which subjects differed according to s_i (Treatments 2 and 4), the subject with the highest cost of waiting volunteered 123 times, the subject with the middle cost of waiting volunteered 69 times, and the subject with the lowest cost of waiting volunteered only 49 times. In both cases, the distribution of volunteers is significantly different from a random draw. By contrast, in the 96 rounds in which subjects did not differ by either C_i or s_i (Treatment 3), the three subjects volunteered 43, 29 and 36 times each. These numbers are not significantly different from a random draw.[23]

It is worth noting that the play-against-nature hypothesis does not predict that the individual with the lowest cost of volunteering (or the highest cost of waiting) will *always* volunteer first. The decision of how long to wait also depends on the players' unobserved beliefs, $g(t)$. What the play-against-nature hypothesis predicts is that the subjects who have a lower cost of volunteering or a higher cost of waiting should be observed to volunteer *more often* because they tend not to wait as long. This is exactly what we observed. Over Treatments 1, 2, 4, and 5, for which the PAN hypothesis predicts $n_A > n_B > n_C$ where n_i is the number of rounds in which subject i volunteers, we observed $n_A = 268$, $n_B = 136$ and $n_C = 95$. We can reject the null hypothesis that the true distribution is a random distribution of volunteers across the three subject types ($\chi^2 = 98.29$, p = 0.000). In addition to this overall observation, this order is observed in all four treatments ([88, 27, 26], [53, 24, 23], [70, 45, 26], and [57, 40, 20]). Because of the smaller sample size the numbers are not as unambiguous when we break down the data by sub-treatment, but nonetheless in 13 of 16 sub-treatments $n_A > n_B$, in 9 of 13 $n_B > n_C$, and in 15 of 16 $n_A > n_C$.

Looking at the correlation between the stopping time and the size of the parameters C_i and s_i, we ought to observe fewer multiple volunteers and longer stopping times in treatments 1.2, 1.4, 3.3 and 3.4 where the cost of waiting is small ($s_i = 0.1$) than in treatments 1.1, 1.3, 3.1, 3.2 and 5.1-5.4 where it is large ($s_i = 0.2$). The data only weakly confirm that this is what happened: when s_i is smaller, we observed fewer multiple volunteers at $t = 0$ (11% of the rounds

[23] The χ^2 statistic for the test of the hypothesis that all three subjects volunteer as often as each other is 63.28 for treatments 1 and 5, 36.48 for treatments 2 and 4 and 2.72 for treatment 3. The critical value of the χ^2 statistic with 2 degrees of freedom at a 5% significance level is 5.99.

when $s_i = 0.1$ and 16% when $s_i = 0.2$), fewer games ending at $t = 0$ (54% when $s_i = 0.1$ and 60% when $s_i = 0.2$), but nearly identical average stopping times (3.72 seconds when $s_i = 0.1$ and 3.78 seconds when $s_i = 0.2$).[24] It may be that the differences in the costs of waiting in these experiments (10 cents versus 20 cents per second) are not large enough to prompt significantly different waiting times. It would have been interesting to see whether subjects would have waited significantly longer if waiting had cost them only 1 or 2 cents per second instead.

The same analysis can be made regarding the cost of volunteering. We ought to observe fewer multiple volunteers and longer stopping times in treatments 2.1 and 2.3 ($C_i = 10$) than in treatments 2.2, 2.4, 3.1, 3.3, 4.1 and 4.3 ($C_i = 5$), and the shortest stopping time in treatments 3.2, 3.4, 4.2 and 4.4 ($C_i = 1$). The data strongly confirm that this is what happened: When C_i is larger, we observed fewer multiple volunteers at $t = 0$ (6% of the rounds when $C_i = 10$, 7.5% when $C_i = 5$, and 14% when $C_i = 1$), fewer games ending at $t = 0$ (29% when $C_i = 10$, 44% when $C_i = 5$, and 71% when $C_i = 1$), and longer average stopping time (17 seconds when $C_i = 10$, 5 seconds when $C_i = 5$, and 1 second when $C_i = 1$).[25] These data confirm that subjects tended to wait longer when the cost of volunteering was higher, as predicted by the play-against-nature

[24] Using a χ^2 test we cannot reject the null hypotheses that s_i has no effect on the number of multiple volunteers when $t = 0$ and that s_i has no effect on the number of rounds that end at $t = 0$. A t-test on the difference between the mean stopping times when $s_i = 0.1$ and $s_i = 0.2$ does not permit us to reject the null hypothesis that the mean stopping times are the same.

[25] Using a χ^2 test we cannot reject the null hypothesis that the proportion of multiple volunteers is the same across the three values of C_i , but we can reject the null hypothesis that the proportion of games ending at t = 0 is the same across the three values of C_i . Pairwise one-sided Fisher exact tests support the conclusion that the negative relationship between C_i and the proportion of games ending at t = 0 is significant. Finally, t-tests on the differences between mean stopping times support the significance differences in all pairwise comparisons of average stopping times.

hypothesis.

The predictive performance of the PAN hypothesis can be compared to that of subgame-perfection if we look at the treatments in which the subject predicted to volunteer most by the PAN hypothesis is not the player predicted to volunteer by the SPE hypothesis. In treatments 4.1, 4.2, 4.3, 4.4, 5.3 and 5.4 the subject predicted to volunteer by SPE did so in 51 of 176 rounds, while the subject predicted to volunteer the most often by the PAN hypothesis did so in 99 rounds. The PAN prediction was correct twice as often as the SPE prediction. Not only does the PAN hypothesis fit the data well, but it's predictive power is clearly superior to subgame-perfection.

6. Conclusions

Who will volunteer to do a job that everyone thinks should be done but that everyone would rather let someone else do? The evidence from the experiments we conducted shows that the answer is not as clear as the theory predicted. Subjects did tend to behave in a systematic manner: in most rounds someone volunteered immediately or fairly quickly (73% of the rounds ended within the first 2 seconds), and very few rounds (1%) ended without a volunteer. However, our data contradict the hypothesis that subjects saw their subgame-perfect equilibrium strategies as the obvious way to play the game. The SPE hypothesis had very poor predictive power, being approximately correct in only 28% of the rounds. Moreover, even in the cases where the SPE prediction was realized, the subjects' play was likely governed by other considerations since they appeared to play the same way in games with and without a unique SPE. We proposed a simple behavioral hypothesis to explain the observed data: the participants in our sessions were unable to predict how the others would play, and therefore tried to maximize their expected payoffs by choosing a waiting time that equates the expected cost and benefit of waiting an additional second – as if they were playing against an unpredictable opponent. We called this hypothesis "play-against-nature"(PAN). The data on the timing and identity of volunteers is generally consistent with this hypothesis, and its predictive power is much superior to the SPE prediction.

We can draw two important lessons from this exercise. First, we are reminded yet again that even though useful insights about incentives and strategic behavior may be obtained from complete information models, one should be cautious to extend the predictions obtained from these models to actual strategic interaction situations. Since it is impossible to insure that the rationality of all the subjects involved in a game is common knowledge, we cannot generate a game of complete information in the laboratory setting, much less expect to observe one in the field. This is significant because even if all players are fully rational, backward induction reasoning can break down completely when the slightest bit of incomplete information is introduced.[26] Second, these experiments also suggest that given *homo sapiens'* limited cognitive ability, decision-making time is a relevant factor in experimental design. Figuring out how the other subjects would play may have been too complex a task in the time we allowed them to make a decision. We may wonder, for example, whether subjects would have played the same way if they had been given 24 hours to research and think about a strategy instead of 15 seconds. This is an issue for further research.

[26] See Kreps (1990) p. 536-543 for a discussion and other examples.

Acknowledgments

Marc Bilodeau and Stuart Mestelman are grateful to the Social Sciences and Humanities Research Council of Canada for research funding. Helpful comments were received from Robert Sandy and from participants at a session of the meetings of the Public Choice Society and Economic Science Association in New Orleans and from anonymous referees.

References

Bergstrom, T., L. Blume and H. Varian. 1986. "On the Private Provision of Public Goods." *Journal Of Public Economics*, 29, 25-59

Bilodeau, M. and A. Slivinski. 1996. Toilet Cleaning and Department Chairing: Volunteering a Public Service. *Journal of Public Economics* 59, 299-308.

Bliss, C. and B. Nalebuff. 1984. Dragon Slaying and Ballroom Dancing: The Private Supply of a Public Good. *Journal of Public Economics* 25, 1-12.

Buchanan, J. M. 1967. Cooperation and Conflict in Public Good Interaction, *Western Economic Journal*, March 1967, 109-121

Davis, D. D. and C. A. Holt. 1993. *Experimental Economics*. Princeton University Press, Princeton, NJ.

Friedman, D. and S. Sunder. 1994. *Experimental Methods: A Primer for Economists*. Cambridge University Press: New York, NY.

Fudenberg, D. and J. Tirole. 1991. *Game Theory*. MIT Press: Cambridge, MA.

Hendricks, K., A. Weiss and C. A. Wilson. 1988. The War of Attrition in Continuous Time with Complete Information. *International Economic Review* 29, 663-680.

Hirshleifer, J. 1983. "From Weakest Link to Best Shot: The Voluntary Provision Of Public Goods," *Public Choice* 41, 317-386

Kreps, D.A. (1990). *A Course in Microeconomic Theory*, Princeton University Press: Princeton, NJ.

Ledyard, J. O. 1995. "Public Goods: A Survey of Experimental Research", in *Handbook of Experimental Economics*, J. Kagel and A. Roth, eds. Princeton University Press:

Princeton, NJ, 111-194.

Nagel, R. 1995. "Unraveling in Guessing Games: An Experimental Study", in *The American Economic Review* 85(5), 1313-1326.

Plott, C. R. and V. L. Smith. 2003. *Handbook of Experimental Economics Results*. Elsevier Science: New York, NY.

Rasmusen, E. 1989. *Games and Information: An Introduction to Game Theory*. Basil Blackwell: New York, NY.

Roth, A. E., 1995. "Bargaining Experiments" in *Handbook of Experimental Economics*. J. Kagel and A. Roth, eds. Princeton University Press: Princeton, NJ, 254-349.

Thaler, R. H., 2000. "From Homo Economicus to Homo Sapiens", in *Journal of Economic Perspectives* 14(1), p. 133-141.

Table 1 Experimental Design

Treatment Parameter Set	Observations	E_i	s_i	C_i	t_i*	SPE Prediction
1.1	32	20	0.2	1, 3, 5	85, 75, 65	A
1.2	32	20	0.1	1, 3, 5	80, 60, 40	A
1.3	32	10	0.2	1, 3, 5	44, 34, 24	A
1.4	32	10	0.1	1, 3, 5	80, 60, 40	A
2.1	24	18, 12, 6	0.3, 0.2, 0.1	10	26, 9, -	A
2.2	24	18, 12, 6	0.3, 0.2, 0.1	5	42, 34, 9	A
2.3	24	20, 15, 10	0.4, 0.3, 0.2	10	24, 16, -	A
2.4	24	20, 15, 10	0.4, 0.3, 0.2	5	36, 32, 24	A
3.1	24	20, 17, 14	0.2	5	65, 59, 44	A
3.2	24	20, 17, 14	0.2	1	85, 79, 64	A
3.3	24	20, 17, 16	0.1	5	40, 40, 40	A,B,C
3.4	24	20, 17, 16	0.1	1	80, 80, 80	A,B,C
4.1	32	20	0.4, 0.2, 0.1	5	36, 65, 40	B
4.2	32	20	0.4, 0.2, 0.1	1	46, 85, 80	B
4.3	32	10	0.4, 0.2, 0.1	5	11, 24, 40	C
4.4	32	10	0.4, 0.2, 0.1	1	21, 44, 80	C
5.1	24	16, 18, 20	0.2	1, 5, 10	74, 64, 40	A
5.2	24	16, 18, 20	0.2	1, 3, 5	74, 74, 65	A,B
5.3	24	10, 15, 20	0.2	1, 5, 10	44, 49, 40	B
5.4	24	10, 15, 20	0.2	1, 3, 5	44, 59, 65	C

Note: The first 8 sessions included treatments 1, 2 and 4. The data from treatment 2 were inaccurately recorded because of a programming error. These data are not reported. The next 6 sessions included treatments 2, 3, and 5. These treatment 2 data are reported.

For some parameter combinations, a player's payoff would still have been positive at t=90. If the game ended without a volunteer, the payoff he received was then E_i - 90 s_i. For other parameter combinations, a player's payoff would have decreased to zero before t=90. In this case, the payoff function had a kink and followed the horizontal axis from that point on until the end of the game. Due to a programming glitch subjects received one second's payoff, s_i, instead of 0. This shortens the t_i* values by one second. The values of t_i* are therefore calculated as $t_i^* = \min\{90 - C_i/s_i, (E_i - C_i)/s_i - 1\}$.

Table 2 Summary of Subgame Perfect Equilibrium Prediction Performance

Description	Number of Rounds	Percentage
Predicted subject volunteers immediately, all others wait	97	20%
Predicted subject waits 1 or 2 seconds	36	8%
Predicted subject waits more than 2 seconds (average stopping time: 13 seconds)	61	13%
Multiple volunteers at time 0 (including predicted subject)	37	8%
Another subject volunteers but not the predicted subject	236	50%
No one volunteers	5	1%
Total	472	100%

Note: Totals exclude treatments 3.3, 3.4 and 5.2 in which there was not a unique SPE

Table 3 Identity of Volunteers by Treatment (Predicted Subgame Perfect Equilibrium Volunteers are Identified with **Bold** Font)

	Participant		
Treatment	A	B	C
1.1	**21**	11	6
1.2	**22**	7	7
1.3	**21**	5	7
1.4	**24**	4	6
2.1	**16**	4	5
2.2	**15**	7	3
2.3	**8**	8	8
2.4	**14**	5	7
3.1	**12**	8	6
3.2	**11**	7	11
3.3	8	8	10
3.4	12	6	9
4.1	21	**8**	6
4.2	14	**16**	6
4.3	20	6	**8**
4.4	15	15	**6**
5.1	**16**	7	3
5.2	12	11	8
5.3	13	**11**	7
5.4	16	11	**2**

Note: Rows do not always total 32 or 24 because there were multiple volunteers in some rounds. In some other rounds time ran out and no one volunteered.

Table 4 Frequency Distributions and Medians of Times at which Action was taken and Number of Multiple Volunteers by Treatment

Treatment	Time Elapsed Before a Volunteer Takes Action (in seconds)								Multiple Volunteers
	0	1 to 2	3 to 5	6 to 10	11 to 20	21 to 89	No Volunteer	Median	
1.1	17	8	5	1	0	1	0	0	6
1.2	16	7	2	5	0	2	0	0.5	4
1.3	16	4	5	3	1	3	0	0.5	1
1.4	16	6	6	2	2	0	0	0.5	2
Sub-total	*65*	*25*	*18*	*11*	*3*	*6*	*0*	*1*	*13*
2.1	7	4	4	1	3	3	2	4	1
2.2	10	4	2	6	1	1	0	1	1
2.3	7	4	2	3	0	5	3	4.5	2
2.4	9	5	5	2	3	0	0	1	2
Sub-total	*33*	*17*	*13*	*12*	*7*	*9*	*5*	*2*	*6*
3.1	7	4	2	6	3	2	0	4	2
3.2	19	4	1	0	0	0	0	0	5
3.3	10	8	0	1	1	3	1	1	3
3.4	19	3	1	1	0	0	0	0	3
Sub-total	*55*	*19*	*4*	*8*	*4*	*5*	*1*	*0*	*13*
4.1	18	5	1	3	3	2	0	0	3
4.2	21	10	0	0	0	1	0	0	4
4.3	17	5	3	5	1	1	0	0	1
4.4	20	9	3	0	0	0	0	0	4
Sub-total	*76*	*29*	*7*	*8*	*4*	*4*	*0*	*0*	*12*
5.1	19	0	1	2	1	1	0	0	2
5.2	16	3	2	1	0	2	0	0	7
5.3	14	5	2	0	2	1	0	0	6
5.4	16	1	2	2	1	2	0	0	4
Sub-total	*65*	*9*	*7*	*5*	*4*	*6*	*0*	*0*	*19*
Total	294	99	49	44	22	30	6	0	63

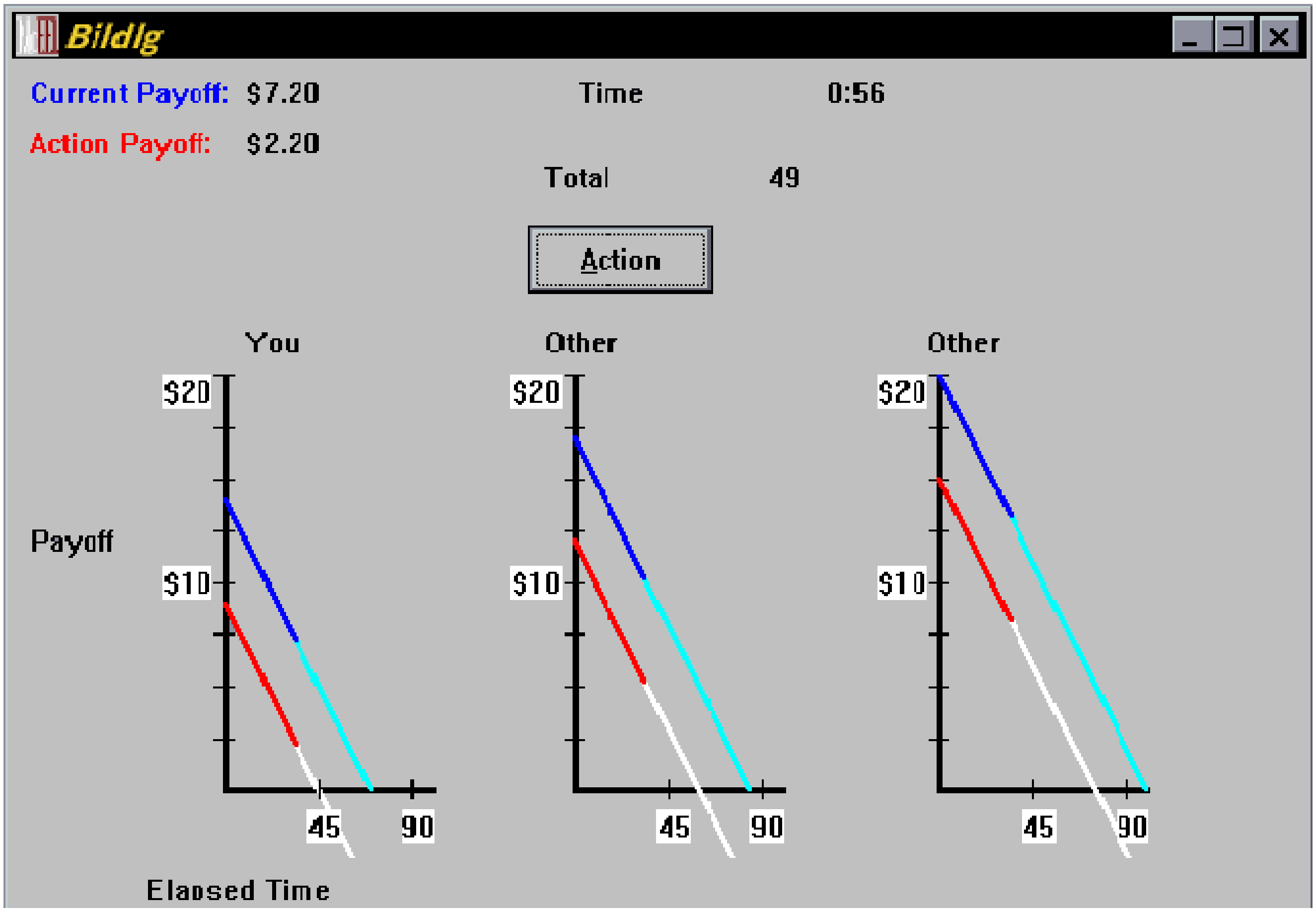

Figure 1 Subject's Screen Providing Information for the Three Participants in a Group

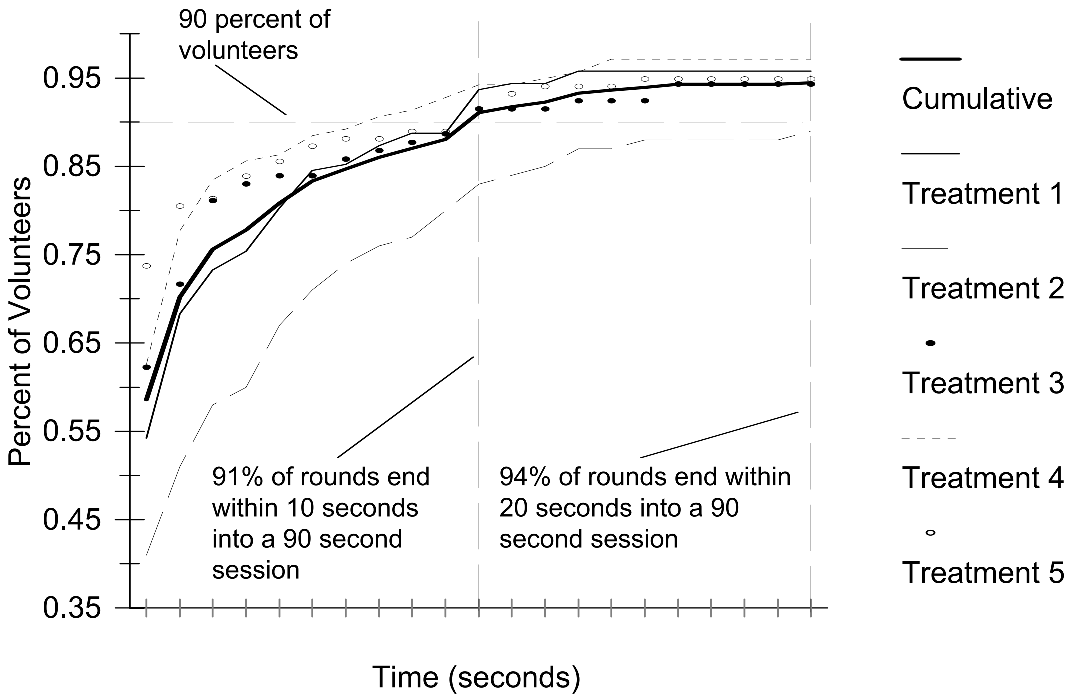

Figure 2 Cumulative Distributions of Volunteers over Time by Treatment