Fairness through the Lens of Cooperative Game Theory: An Experimental Approach*

Geoffroy de Clippel† Kareen Rozen‡

November 2020

Abstract

We experimentally investigate how impartial observers allocate money to agents whose complementarity and substitutability determine the surplus each group can achieve. Analyzing the data through the lens of axioms and solutions from cooperative-game theory, a one-parameter model (mixing equal split and Shapley value) arises as a parsimonious description of the data.

*First version October 2013. We are grateful to Pedro Dal Bó, Mark Dean, Kfir Eliaz, Louis Putterman, Larry Samuelson, Roberto Serrano, Rajiv Vohra, Leeat Yariv, three anonymous referees and various seminar audiences for valuable suggestions, and also thank Roberto Serrano for arranging part of the funding for this project. The paper benefited from the excellent research assistance of Camilo Dominguez Gonzalez and Emily Nix in the initial stages, and Zeky Murra in later stages. We also thank Eli Zvuluny of Possible Worlds Ltd. for programming the experiment.

†Brown University, Department of Economics.

‡Brown University, Department of Economics.

1 Introduction

In many economic settings, including trading and joint production, the surplus to be shared is created through collaboration. Complementarity and substitutability among agents determine how much a group of agents can share when they cooperate. Consider, for instance, three musicians who can play together as a duo or a trio for an event (but not as soloists). They will collect $900 for performing as a trio. Should they instead perform as a smaller ensemble, they would be paid less: Musicians 1 and 2 could collect $800, Musicians 1 and 3 could collect $600, and Musicians 2 and 3 could collect $400. If you could decide, as a neutral outside party, how to split the $900 earnings of the trio between them, what would you do?

Intuitively, the allocated reward for collaborating may be small if an agent's role in creating the surplus is limited. By contrast, an agent judged as playing a more critical role might be rewarded more. The long economic literature on other-regarding preferences has so far studied the notion of fairness from a very different angle. It was not designed to address scenarios with such complementarity and substitutability among agents,[1] as subcoalitional worths are not taken into account. If one were forced to apply this literature to our problem, then choices for others would be independent of the worths of subcoalitions (and in many cases, would be an equal split).

As its main takeaway, this paper provides robust evidence that coalition worths do matter when choosing for others, and that principles from cooperative game theory have strong explanatory power in such situations.[2] We test axioms and compare competing solutions. At least in the context of the problems studied here, we find that choices are well understood with a one-parameter solution that finds its roots in cooperative game theory.

[1]See Fehr and Schmidt (1999), Bolton and Ockenfels (2000), Andreoni and Miller (2002), Charness and Rabin (2002), Karni and Safra (2002), and Fisman, Kariv and Markovits (2007), among others. Discussions of this literature can be found, for instance, in the book by Camerer (2003) and the survey by Sobel (2005).

[2]See Moulin (2003) for a textbook introduction to cooperative games from a normative perspective, which is the perspective we adopt here.

In our experiment, three subjects are randomly designated at the start of each session as *Recipients*, while others are designated as *Decision Makers*. In each of seven rounds, the Decision Makers are provided the set of coalition worths for the three Recipients (a *characteristic function*, in the terminology of cooperative game theory). These worths correspond to the value of different combinations of the Recipients' 'electronic baskets', whose composition is decided by the performance of each Recipient on an earlier quiz. Decision Makers play the main role in our experiment, as only they provide our choice data. For each characteristic function, we ask each of them to decide how to split the worth of the grand coalition between the three Recipients. At the end of a session, the Recipients are paid according to a randomly selected Decision Maker's allocation for a randomly selected characteristic function. Our experimental design ensures Decision Makers are 'impartial observers', in the sense that their monetary payoffs are independent of their recommendation (in contrast to dictator and ultimatum games). Moreover, the design eliminates strategic channels that might affect recommendations (in contrast to ultimatum games, or settings where reciprocity is a concern).

Understanding people's views when allocating money in such settings is important, both for its own sake, as well as to shed light on the fair reference point to use when assessing intentions and reciprocity in multiagent settings. One could, for instance, think of Decision Makers as capturing arbitrators, who are impartial observers with no immediate stakes in the decisions they make. Our design eliminates strategic considerations to pinpoint fairness views in their purest form, but our findings likely have important implications for more complex settings with strategic considerations, where fairness ideals may suggest focal equilibria or act as reference points. For instance, being offered a reward that is considered unfair may make a musician resentful, inducing her to either refuse joining the ensemble, or to exert relatively little effort if she does join. These considerations are left for future work.

Many cooperative game-theory solutions have been proposed over the years. A first question is whether or not their predictions are borne out in the data, and in particular which of them is most successful at describing surplus-sharing decisions. But testing axioms, in addition to examining the explanatory power and relative prevalence of some known solution concepts, offers a fuller picture of what people view as fair. Certain properties are satisfied by multiple solution concepts, and may thus appear, at least on a theoretical level, to be more universal and fundamental. Others are satisfied by a narrower class of solutions, and capture the essence of what distinguishes these from others.

Nearly all our Decision Makers choose equal split for a characteristic function where all standard solution concepts agree that is the solution. Yet these same Decision Makers often choose unequal splits in other, asymmetric characteristic functions. Our data analysis, which discusses average, aggregate and individual behavior, provides strong evidence in support of the axioms of *Symmetry*, *Desirability*, *Monotonicity*, and *Additivity*. However, the *Dummy Player* axiom, whereby a Recipient who adds no value to any coalition should get a zero payoff, is clearly violated. We show that satisfying Symmetry and Additivity (along with Efficiency, which must be satisfied in our experiment) means Decision Makers' choices are characterized by a linear combination of the Equal Split solution and Shapley value, with weights summing to one.[3] We use the data to estimate the resulting one-parameter, linear model–which we refer to as the *Equal-Split Shapley (ESS)* model.

In the concluding section, we discuss two alternative treatments we designed later to test further questions. In one variant, we eliminate the quiz to test whether Decision Makers ignore coalitional worths that are randomly assigned. As will be seen, the results are similar to our main treatment, which may suggest that absent detailed information about the quiz, Decision Makers considered baskets nearly randomly determined. Perhaps surprisingly, though, the similarity shows Decision Makers take basket worths seriously even when 'unearned.' In another robustness treatment, we survey Decision Makers about profit sharing among hypothetical musicians performing as a trio, similarly to

[3]This theoretical result holds for the set of 3-player characteristic functions studied here, although Casajus and Huettner (2013)'s result tells us the result essentially extends to any number of players and general characteristic functions, provided one adds a mild requirement that null players receive a nonnegative amount when the grand coalition has a positive amount to share. Alternate axiomatizations are given in van den Brink et al. (2013).

the story in the first paragraph. That treatment establishes that our qualitative analysis is portable across situations, though quantitative estimates–a single, easily identifiable parameter–may vary.

Further related literature

To our knowledge, this is the first paper to experimentally investigate people's views on monetary allocations in situations involving substitutability and complementarity among recipients, and to empirically show that principles of cooperative game theory can prove helpful for this purpose. Another novelty of our approach is that, in addition to checking the relative prevalence of various solution concepts, we assess the empirical validity of various axioms. We believe that testing axioms instead of testing specific functional forms could be informative when it comes to better understanding other-regarding preferences in other contexts as well.

There is a small experimental literature testing cooperative games from a different perspective, allowing multiple subjects to bargain given a characteristic function. Kalisch, Milnor, Nash and Nering (1954), one of the earliest papers in experimental economics, informs subjects of their role in a characteristic function and lets them interact informally. Others impose a formal bargaining protocol, in addition to specifying a characteristic function, to concentrate on a particular question of interest. For instance, Murnighan and Roth (1977) consider the effect of messages during negotiation, and the announcement of payoff decisions, on the resulting allocations; while Bolton, Chatterjee, and McGinn (2003) study the impact of communication constraints in a three-person bargaining game in characteristic-function form. Yan, Friedman and Munro (2016) study the validity of extreme core predictions when using various market institutions to trade a single unit of an indivisible good. Nash, Nagel, Ockenfels and Selten (2012) are interested in whether efficient outcomes arise from a 40-times repeated bargaining game, with each stage following their 'agencies' bargaining protocol; they study who is appointed to split the pie (e.g., will it be the 'strongest' player in the characteristic function?), and how the appointee's split compares to some known solutions. On balance, a fair allocation can potentially serve as a focal or reference point to select among multiple equilibria in games. For complex strategic games, where many conflicting aspects play a role in players' decisions,[4] the 'fair' benchmark against which offers may be measured can be difficult to tease out.

One interpretation of the Shapley value is that it rewards people for their role in creating the surplus, which Shapley measures by their marginal contributions. Konow (2000) and Cappelen, Hole, Sørensen, and Tungodden (2007) also touch upon the theme of rewarding contributions, but in a two-player dictator game where the pie to split is the sum of the two subjects' 'contributions' in an earlier production phase. To understand how the dictator's choice depends on factors within versus beyond his control, a subject's contribution is the product of a chosen investment level and an exogenous rate of return. Among other questions, Konow studies whether liberal egalitarianism explains observed allocations when entitlement follows an accountability principle. Cappelen et al. studies the relative prevalence of fairness ideals beyond liberal egalitarianism, such as strict egalitarianism and libertarianism. We study scenarios that differ on multiple dimensions. First, instead of being specified as the sum of individual contributions, the amount to split arises from complementarity and substitutability across agents. A main question is then *how* Decision Makers assess individual contributions in such settings. Do they use marginal contributions, as suggested by the Shapley value? Many other measures are conceivable as well. Second, we provide no quantifiable information to express coalition worths as a precise function of effort and luck parameters. Besides keeping the analysis focused on our main point of interest

[4]Consider for instance Nash et al (2012)'s repeated game. Each stage starts with one player out of three being selected by the 'agencies' protocol to allocate the coalition's profit (including to himself). Repeating this 40 times, choices can reflect negative reciprocity (they show "the more aggressive the demand of one player is, the more aggressive are those of the others"), reputation building, strategic experimentation (how much disparity others tolerate), and end-game effects (will the last appointee take all?). They show splits vary widely as a function of the appointee, who always either favors himself or splits equally (thus departing from all cooperative solutions whenever the appointee is not the 'strongest' player). For each characteristic function, they quantify how much the average split over 40 rounds departs from different solutions using MSE.

– whether and how Decision Makers reward people for their role in creating the surplus – we see it as a realistic feature of some applications. For instance, the musicians' opportunities are quantifiable in terms of profit, but we would not expect the musicians themselves, and a fortiori impartial observers, to understand or agree on the differential impacts of talent and hard work in generating them.

Decision Makers' choices have no material consequences for themselves in our experiment. This approach is borrowed from earlier experimental papers studying fairness ideals in other contexts, where conflicting notions of fairness might coexist; these include Konow (2000), Cappelen, Konow, Sorensen and Tungodden (2013), and Cappelen, List, Samek and Tungodden (2020) among others. This approach has also been used very recently to test paternalism in the lab; see Ambuehl, Bernheim, and Ockenfels (2020).

2 Theoretical Benchmark

Let I be a set of n individuals. A *coalition* is any subset of I. Following von Neumann and Morgenstern (1944), a *characteristic function* v associates to each nonempty coalition S a worth $v(S)$. The amount $v(S)$ represents how much members of S can share should they cooperate. That is, an allocation, or payoff vector, x is *feasible* for S if $\sum_i x_i \leq v(S)$. Assuming that the grand coalition forms (that is, all players cooperate), how should $v(I)$ be split among individuals? A solution ϕ associates to each characteristic function v a set $\phi(v)$ of payoff vectors that are feasible for N. When $\phi(v)$ is a singleton, $\phi(v)$ will also denote the unique element of the set.

A significant part of cooperative game theory aims at defining normative principles that a solution might satisfy, and understanding which combinations characterize which solution concepts. We list some such principles in Section 2.1, and consider prominent solution concepts in Section 2.2. We use the characteristic functions from our experimental design to illustrate the axioms and solution concepts, thereby previewing their implications for choices in our experiment. These characteristic functions (CFs) are defined for three individuals (recipients called R1, R2 and R3) and listed in Table 1 below. We summarize the theoretical motivations for these CFs in Section 2.3.

	{R1,R2}	**{R1,R3}**	**{R2,R3}**	**{R1, R2, R3}**
CF1	60	0	0	60
CF2	40	40	0	40
CF3	40	40	20	50
CF4	80	60	40	90
CF5	30	15	15	30
CF6	40	40	0	70
CF7	40	40	40	60

TABLE 1: The seven characteristic functions (CF) studied in the experiment; the worth of singleton coalitions is zero.

2.1 Normative Principles

Individual i is a *dummy player* if $v(S) = v(S \setminus \{i\})$, for any coalition S containing i. The *Dummy Player* axiom stipulates that such individuals receive a zero payoff. Formally, ϕ satisfies the axiom if $x_i = 0$ whenever $x \in \phi(v)$ and i is a dummy player in v. The Dummy Player axiom can be tested in CF1, where R3 plays the dummy role. The worth of the grand coalition in CF1 is the same as in the fully symmetric CF7, making it a direct test of whether Decision Maker's choices vary with subcoalitional worths.

Individual i is *more desirable* than j if for any non-singleton coalition containing individual j but not i, replacing j with i strictly increases profit. If replacing j with i never makes a difference, i and j are *symmetric*. A payoff vector *respects symmetry* if it allocates the same amount to symmetric individuals. It *respects desirability* if it allocates a strictly larger amount to i than to j when i is more desirable than j.[5] Formally, ϕ satisfies the *Desirability* (resp., *Symmetry*) axiom if, for all characteristic function v, each payoff vector in $\phi(v)$ respects desirability (resp., symmetry).

[5]Comparisons of payoffs in terms of the individuals' relative desirability were first suggested by Maschler and Peleg (1966).

	CF 1 and 5	CF 2, 3 and 6	CF4	CF7
Rankings	R1∼R2≻R3	R1≻R2∼R3	R1≻R2≻R3	R1∼R2∼R3

TABLE 2: The ranking of Recipients in each of the seven characteristic functions, where Ri≻Rj (Ri∼Rj) means that Ri is more desirable than (symmetric to) Rj.

In each of CF1-7, every pair of Recipients can be ranked in terms of either symmetry or desirability. In particular, Ri is more desirable than (symmetric to) Rj if and only if $v(\{i,k\}) > v(\{j,k\})$ (resp., $v(\{i,k\}) = v(\{j,k\})$). Table 2 shows the rankings of Recipients in CF1-7. Symmetry and Desirability have implications within each characteristic function, with the exceptions of CF4, where only Desirability applies (as it is fully asymmetric) and CF7, where only Symmetry applies (as it is fully symmetric). Notice that R1 is always more desirable than, or symmetric to, R2; and in turn, R2 is always more desirable than, or symmetric to, R3. This is just for the purpose of normalizing CF1-7. As will be discussed in Section 3, Decision Makers' screens display random permutations of the Recipients true identifiers ($i = 1, 2, 3$), so that they see a permutations of each CF; hence they cannot detect the ranking above.

The axioms discussed so far apply pointwise: i.e., for given characteristic functions. The next properties relate payoff vectors *across* characteristic functions. Suppose that one selects a payoff vector x for a characteristic function v, and a payoff vector $\hat{x}$ for a characteristic function $\hat{v}$. Suppose further that the only difference between v and $\hat{v}$ is that the worth of coalition S has increased. Then the payoff vectors x and $\hat{x}$ *respect Monotonicity* if the payoff of each member of S increases, that is, $\hat{x}_i > x_i$ for all $i \in S$. Formally, ϕ satisfies Monotonicity if $\hat{x}_i > x_i$ for all $\hat{x} \in \phi(\hat{v})$, $x \in \phi(v)$, and all $i \in S$ whenever the only difference between v and $\hat{v}$ is that $\hat{v}(S) > v(S)$.

Monotonicity has multiple implications for CF1-7. First, in going from CF2 to CF6, only the value of the grand coalition changes (it increases from \$40 to \$70), and so the axiom says that every Recipient must receive more in CF6 than in CF2. But the axiom can also be applied iteratively. For instance, if one modifies CF3 to a new CF3', by increasing $v(\{1,2,3\})$ from

\$50 to \$70, the axiom requires all payoffs to increase; and if one then modifies CF3' to CF6, by reducing $v(\{2,3\})$ from \$20 to \$0, then the payoffs of R2 and R3 should decrease. Hence the axiom calls for an unambiguous increase for Recipient 1 when moving from CF3 to CF6. Slightly more complex reasoning shows R1's payoff should increase in going from CF5 to CF2. Similarly, R3's payoff should increase going from CF1 to CF7. Lastly, the payoffs of R2 and R3 should increase when moving from CF2 to CF3, and when moving from CF2 to CF7.

The *Additivity* axiom is a cornerstone of Shapley (1953). Given two characteristic functions v and $\hat{v}$, the sum $v+\hat{v}$ is the characteristic function where the worth of each coalition is the sum of its worth in v and in $\hat{v}$. The solution ϕ *respects Additivity* if $\varphi(v+\hat{v}) = \varphi(v) + \varphi(\hat{v})$ for all characteristic functions $v, \hat{v}$. In the case of a single-valued solution, for instance, if φ selects the payoff vectors x for v and $\hat{x}$ for $\hat{v}$, then it selects $x+\hat{x}$ for $v+\hat{v}$. As is well known, Additivity is equivalent to linearity with respect to rational coefficients: $\varphi(\alpha v + \beta\hat{v}) = \alpha\varphi(v) + \beta\varphi(\hat{v})$, where $\alpha, \beta \in \mathbb{Q}_+$. The case $\alpha = \beta = 1/2$ will be useful for us, and is simple to prove; since $\varphi(2v) = 2\varphi(v)$,

$$\varphi(\frac{1}{2}v + \frac{1}{2}\hat{v}) = \varphi(\frac{1}{2}v) + \varphi(\frac{1}{2}\hat{v}) = \frac{1}{2}\varphi(v) + \frac{1}{2}\varphi(\hat{v}).$$

We have two ways of testing Additivity, even though no two of our characteristic functions immediately add up to a third. First, under the reasonable assumption that Decision Makers would choose an equal split in a hypothetical characteristic function where only the grand coalition has positive worth (equal to \$30), the Additivity axiom can be examined using Decision Makers' choices in both CF2 and CF6: for each Recipient, the amount allocated in CF6 should be \$10 larger than in CF2. Second, since CF3 is the average of CF2 and CF7, for each Recipient the amount allocated in CF3 should be the average of the amounts allocated to that Recipient in CF2 and CF7.

2.2 Solution Concepts

The *equal-split* solution (ES) simply divides $v(I)$ equally among all individuals, for any v. By contrast, cooperative game theory provides a variety of solution concepts that account for the worths of sub-coalitions, each capturing a distinct notion of fairness. Prominent solution concepts are the *Shapley value* (Shapley, 1953), the *core* (Gillies, 1959), the *nucleolus* (Schmeidler, 1969), and the weak- and strong-constrained egalitarian allocations (Dutta and Ray, 1989 and 1991). We provide a quick primer for the reader below.

The Shapley value. Consider building up the grand coalition one person at a time, giving each i his *marginal contribution* $v(S \cup \{i\}) - v(S)$ to the set S of individuals preceding him. The Shapley value (Sh), a single-valued solution, achieves a notion of fairness by averaging these payoffs over all possible ways to build up the grand coalition. That is,

$$\mathrm{Sh}_i(v) = \sum_{S \subseteq I \setminus \{i\}} p_i(S)[v(S \cup \{i\}) - v(S)],$$

where $p_i(S) = \frac{|S|!(n-|S|-1)!}{n!}$ is the fraction of possible orderings in which the set of individuals preceding i is exactly S.

The core. The core looks for payoffs $x \in R^I$ such that there is no coalition whose members would be better off by cooperating on their own; that is, $\sum_{i \in S} x_i \geq v(S)$ for each coalition S, with $\sum_{i \in I} x_i = v(I)$ for the grand coalition. While often interpreted from a positive standpoint, it also has normative appeal, as it respects property rights for individuals and groups: picking payoffs outside the core means robbing some individuals from what they deserve.

The nucleolus. Like the Shapley value, the nucleolus (Nuc) prescribes a unique solution in all cases. Given a payoff vector x, the *excess surplus* of a coalition S is the amount it receives net of what it could obtain on its own, that is, $\sum_{i \in S} x_i - v(S)$. The nucleolus interprets excess surplus as a welfare criterion for a coalition, and chooses among all feasible payoff vectors the one that lexicographically maximizes all coalitions' excess surpluses, starting from the coalition with the lowest excess surplus and moving up. By contrast, the core simply requires each coalition's excess surplus to be nonnegative. Hence, whenever the core is nonempty, it must contain the nucleolus.

Constrained egalitarian allocations. The constrained egalitarian allocation combines egalitarianism with protection of individual interests. The notion of egalitarianism is based on the Lorenz ordering, which is a partial ordering over allocations such that x Lorenz-dominates y if, loosely speaking, x can be derived from y through a sequence of transfers from 'rich' to 'poor.' The *Lorenz core* of the grand coalition is recursively defined. The Lorenz core of a singleton coalition $\{i\}$ is simply $\{v(i)\}$. The Lorenz core of a coalition S is then the set of feasible allocations for S such that there does not exist any $y \in T \subset S$ such that y is Lorenz-undominated within T and the members of T 'all prefer' y to x. The solution concept picks those allocations that are Lorenz-undominated within the Lorenz core of the grand coalition. The idea in this recursive definition is that objections must themselves be egalitarian. The solution concept has two versions, Strong and Weak, which differ in what 'all prefer' means: in the Strong version (sCEA), all must be strictly better off, while in the Weak version (wCEA), all must be weakly better off, with at least one strict improvement. This seemingly small difference can yield very different predictions. The sCEA may be multi-valued and is always nonempty; but the wCEA, when it exists, selects a unique allocation.

Mixtures of Equal Split and the Shapley Value Our theoretical and empirical analysis will lead us to consider the parametrized class of solutions that are convex combinations of equal split and the Shapley value: $\mathrm{ESS}_\delta(v) = \delta \mathrm{Sh}(v) + (1-\delta)\mathrm{ES}(v)$ where $\delta \geq 0$ is a fixed parameter representing how much weight is placed on the Shapley value.

Table 3 reports what a Decision Maker who perfectly follows one of the above solutions would select for CF1-CF6. For brevity, the table omits CF7, as

	CF1	CF2	CF3	CF4	CF5	CF6
ES	$(20, 20, 20)$	$(\frac{40}{3}, \frac{40}{3}, \frac{40}{3})$	$(\frac{50}{3}, \frac{50}{3}, \frac{50}{3})$	$(30, 30, 30)$	$(10, 10, 10)$	$(\frac{70}{3}, \frac{70}{3}, \frac{70}{3})$
Sh	$(30, 30, 0)$	$(\frac{80}{3}, \frac{40}{6}, \frac{40}{6})$	$(\frac{70}{3}, \frac{40}{3}, \frac{40}{3})$	$(40, 30, 20)$	$(\frac{25}{2}, \frac{25}{2}, 5)$	$(\frac{110}{3}, \frac{50}{3}, \frac{50}{3})$
Core	P_1	$(40, 0, 0)$	$(30, 10, 10)$	$(50, 30, 10)$	$(15, 15, 0)$	P_2
Nuc	$(30, 30, 0)$	$(40, 0, 0)$	$(30, 10, 10)$	$(50, 30, 10)$	$(15, 15, 0)$	$(40, 15, 15)$
wCEA	$(30, 30, 0)$	-	-	$(40, 40, 10)$	$(15, 15, 0)$	$(\frac{70}{3}, \frac{70}{3}, \frac{70}{3})$
sCEA	P_3	$(20, 10, 10)$	$(20, 15, 15)$	$(40, 25, 25)$	$(15, 7.5, 7.5)$	$(\frac{70}{3}, \frac{70}{3}, \frac{70}{3})$

Table 3: What the solution concepts prescribe for CF1-CF6, where $P_1 = \{(x, 60 - x, 0) \mid x \in [0, 60]\}$, $P_2 = \{(70 - x - y, x, y) \mid x, y \in [0, 30]\}$, and $P_3 = \{(30, 15, 15), (15, 30, 15)\}$. wCEA does not exist in CF2-CF3.

all solutions agree on splitting the \$60 equally. It also omits the parametrized solution ESS_δ, which is easily derived by combining the first two rows.

Solution concepts and axioms are complementary: some solutions arose from axioms (e.g., the Shapley value), while other solutions were motivated differently (e.g., the constrained egalitarian solutions) and studied from an axiomatic perspective only later. For each solution concept and each axiom from Section 2.1, Table 4 summarize whether the solution concept always satisfies that axiom (✓) or whether it can violate it (✗) for 3-person characteristic functions.[6] If such a violation cannot be observed over CF1-7, we notate that as ✗* (e.g., wCEA is empty-valued for CF2-3, precluding a violation of additivity here). See Online Appendix A for the underlying arguments.

[6]Table 4 considers only axioms studied in Section 2.1, and thus does not necessarily describe necessary and sufficient axioms for each solution concept. However, the classic result of Shapley (1953) tells us that the Shapley value is the only single-valued solution satisfying the Dummy Player, Symmetry and Additivity axioms (under our running assumption that solutions pick efficient splits of the grand-coalition worth). Other axiomatizations of the Shapley value have been proposed over the years (see Winter (2002) for a survey). In Observation 1 below, we note that dropping the Dummy Player axiom leads to a characterization of ESS_δ for three-player characteristic functions as tested in this paper (see Casajus and Huettner (2013) and van den Brink et al. (2013) for axiomatic results beyond three players). Typical axiomatizations of the core and the nucleolus involve reduced-game properties, restricting how the solution varies as the set of individuals changes; see e.g. Peleg (1986) and Potters (1991). Dutta (1990) also used such properties to characterize wCEA on the class of convex games. The number of recipients is fixed in our experiment, and hence testing reduced-game properties remains an open question.

Axioms	ES	Sh	Core	Nu	wCEA	sCEA	ESS_δ
Dummy	✗	✓	✓	✓	✓	✗	✗
Desirability	✗	✓	✗	✗*	✗	✗	✓
Symmetry	✓	✓	✗	✓	✓	✗	✓
Monotonicity	✗	✓	✗	✗	✗*	✗*	✓
Additivity	✓	✓	✗	✗	✗*	✗	✓

Table 4: Solutions and Axioms for 3-Player CFs ($\delta \in (0, 1)$)

2.3 Motivations for CF1-7

CF1-7 allow us to assess the relative prevalence of solution concepts and directly test well-known axioms. Section 2.1 describes how chosen allocations can be used to test the Dummy, Desirability, Symmetry, Monotonicity and Additivity axioms. Table 3 shows the varied set of predictions for CF1-6 offered by prominent solution concepts other than the equal split solution.

To ensure that subjects are not overwhelmed by numbers, we test only characteristic functions for which the payoff of singleton coalitions is zero. We introduce some variation in the worth of the grand coalition, but construct CF1 and CF7 so that the grand coalition is worth \$60 in both cases. CF1 thus provides not only a test of the Dummy axiom, but a direct test of whether the worths of subcoalitions matter through the contrast with CF7.

All prominent solution concepts from cooperative game theory prescribe equal split for CF7. We thus use the selection of equal split in CF7 as a screening device, focusing our analysis on what these subjects will do in other characteristic functions where solutions can depart from equal split.

To make the contrast between the Shapley value and the core most meaningful, we include some characteristic functions whose core is single-valued (CF2-CF5). With three individuals and singleton coalitions that generate zero profit, the core is single-valued if and only if $v(\{1, 2\}) + v(\{1, 3\}) + v(\{2, 3\}) = 2v(\{1, 2, 3\})$. Under this condition, the Shapley value is exactly halfway between the equal-split solution and the single payoff vector in the core (when the core is single-valued, it also coincides with the nucleolus). We also include two characteristic functions with multi-valued cores (CF1, CF6).

3 Experimental Design and Procedure

We study how individuals (Decision Makers) allocate money to three other participants (Recipients), in view of how much different coalitions of Recipients would be worth. That is, Decision Makers' information is in the form of a *characteristic function*.

At the start of each session, three subjects are chosen through uniform randomization and designated Recipients 1, 2 and 3, respectively. All other subjects are designated Decision Makers. Subjects stay in the same role for the entire session. Each session has seven rounds.

At the start of each round, each Recipient has an empty 'electronic basket.' By answering trivia questions correctly, a Recipient earns some fictitious objects (e.g., two left shoes, a bicycle frame, one bicycle wheel) for his or her basket. Combinations of objects that form a "match" have monetary value. For instance, in a given round a complete pair of shoes – left and right – may be worth \$15, while a bicycle frame with two wheels may be worth \$40. The objects available to each Recipient in a round have been selected so that only combinations of two or three Recipients' baskets may have positive worth. The worth of each basket combination corresponds to the maximum possible sum of values that the objects inside generate. To continue the example above, if combining two particular baskets leads only to a complete pair of shoes and a complete bicycle, then that basket combination would be worth \$55.

Before discussing our control over the values of possible basket combinations, we discuss what the Decision Makers do with those values. Once baskets are determined in each of the seven rounds, Decision Makers are told the values of the different basket combinations. As noted in Section 2, each Decision Maker knows the Recipients are identified to him only through randomly generated aliases in each round, with the characteristic function shown to the Decision Maker permuted accordingly.[7]

[7]That is, Recipients have a true fixed identity as R1, R2 or R3 but their names and

The Decision Maker is permitted to allocate, as he or she deems fit, the monetary proceeds of the three-basket combination among the Recipients; we only require chosen monetary allocations to be nonnegative and efficient (no money is left on the table). The Decision Maker receives \$1 for each round where he opts to make a decision, on top of the \$5 show-up fee. Decision Makers determine Recipients payoffs as follows. At the end of the session, one round and one Decision Maker (who participated in that round) are randomly chosen. Recipients receive the monetary payoffs determined by the chosen Decision Maker in the chosen round, in addition to the \$5 show-up fee. Subjects are informed only of their own payoff, and do not learn which roles other subjects played during the experiment.

Given our interest in testing specific axioms and solution concepts, we opted to maintain some control over the set of characteristic functions faced by Decision Makers. Subjects were told that Recipients would be earning objects in each round by answering quiz questions correctly, but were not told how those objects and their values were selected. For each round, we chose the available objects and values of object combinations with the following goal in mind: if Recipients were to earn all the objects available to them in a round, then one of the seven characteristic functions in Table 1 would be generated.[8] Precisely to reduce the probability that some other characteristic functions would be generated, Recipients were afforded multiple opportunities to earn available objects. In all our sessions, Recipients did indeed earn all available objects, so that CF1-7 are the relevant characteristic functions to study.[9]

We ran six different sessions, using a Latin square design for CF1-6. This allows us to test for potential effects from the order in which the characteristic

---

the characteristic function are permuted on the Decision Makers' screens, with the aliases independently redrawn in each round. This rules out the possibility a Decision Maker's payoff allocation for a Recipient is influenced by earlier choices for that Recipient.

[8] We explain below the session-dependent map from rounds to characteristic functions. For testing other characteristic functions, the authors can provide an algorithm showing how to generate any desired superadditive characteristic function (if all objects are earned), by selecting object values and which objects are available for each Recipient to earn.

[9] Since the number of questions was fixed in advance, it was possible for Receivers to earn fewer objects. In that case, in line with the above description, some other characteristic functions would have been generated, based on which objects were earned and their values.

functions are presented to Decision Makers, and if needed, help wash these out in the aggregate.[10] Online Appendix E details the session-dependent mapping between rounds and characteristic functions. All standard solution concepts agree on an equal split for CF7, and we leave it as a consistency check in the final round of each session, where it cannot affect subsequent behavior.

3.1 Comments on the design

To keep our setting as close as possible to standard split-the-pie problems, our design attempts to mitigate the possibility that information extraneous to the monetary values of basket combinations affects Decision Makers' choices. For this reason, subjects remain in separate roles throughout the experiment, so that Decision Makers cannot differentially consider their personal experience as a Recipient when determining payoff allocations. Decision Makers cannot communicate, and never learn others' allocation decisions. Moreover, a Decision Maker's chosen payoff allocation does not reflect strategic concerns, both because it cannot influence his or her own payoff, and because Recipients have no strategic role. Finally, Decision Makers are not given information about Recipients' performance in the quiz, or the mapping between performance and basket values. The motivation for this choice, which bears some realism,[11] is twofold. First, attention remains focused on the characteristic function itself. Second, even in a controlled laboratory setting, providing more information could lead to less control: how difficult the Decision Maker finds the task, and whether they find the skill it tests valuable, can confound their interpretation of the results in idiosyncratic ways. The above features have the added benefit of simplifying the Decision Maker's problem from a computational standpoint.

The design also aims to provide a simple and economically relevant context where substitutability and complementarity of individuals arise intuitively. We subsequently tested alternative designs in robustness treatments, to vary how

[10] As seen in Online Appendix H, we do not find such order effects.

[11] It echoes aspects of real-life problems where endowments are taken as given but their origin is unknown. For instance, it might not be known whether a valuable skill is innate or acquired through hard work, whether a patented discovery was obtained after years of research or through sheer luck, whether a piece of land was bought or inherited, etc.

characteristic functions arise. Section 5 discusses these, showing that our qualitative analysis is portable, even if quantitative results may vary. The inclusion of a quiz in the design aimed to create a sense of earned worths, because we initially conjectured that subjects would ignore randomly-determined coalition worths. As discussed later, one of the robustness treatments suggests that we had limited success in creating a real sense of earning, and demonstrates that coalitional worths are important even when unearned.

3.2 Procedure

The six sessions were conducted in April and May 2013, and held at a computer lab at Brown University, with subjects participating anonymously through their computer terminal. The interface for the experiment was programmed by Possible Worlds Ltd. to run through a web browser. Subjects were recruited via the BUSSEL (Brown University Social Science Experimental Laboratory) website,[12] and were allowed to participate in only one of the six sessions.

Sessions lasted about thirty to forty minutes. At the start of each session, the supervisor read aloud the experimental instructions, which were also available on each subject's computer screen. The onscreen instructions contained a practice screen for inputting Recipients' payoffs, to get accustomed to the interface. The session supervisor then summarized the instructions using a presentation projected onto a screen. The instructions and presentation are available in Online Appendices C-D. Subjects learned their role as Recipient or Decision Maker only after going through all of the instructions.

A total of 107 subjects participated in the experiment, for an average of nearly eighteen subjects per session. With three subjects selected to be Recipients in each session, a total of 89 subjects acted as Decision Makers.

[12]This site, available at bussel.brown.edu, offers an interface to register in the system and sign up for economic experiments. To do so, the information requested from subjects is their name and email address and, if applicable, their school and student ID number. The vast majority of subjects registered through the site are Brown University and RISD graduate and undergraduate students, but participation is open to all interested individuals of at least 18 years of age without discrimination regarding gender, race, religious beliefs, sexual orientation or any other personal characteristics.

Nearly all Decision Makers chose to actively participate in each round.[13] All subjects received payment in cash at the end of the session.

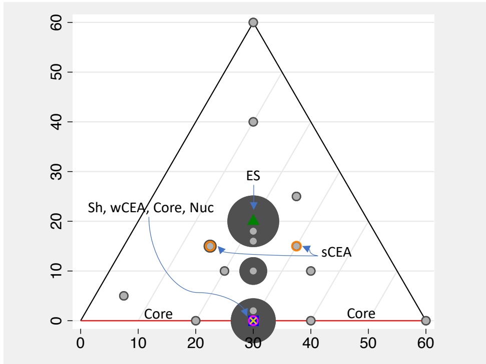
After completing all seven rounds but before learning their payoff, subjects in each session were presented with an optional exit survey via the computer interface. This survey collected basic demographic information (major, gender, age and number of siblings) and allowed subjects to describe how they made their choices as Decision Makers, if applicable.

4 Data analysis

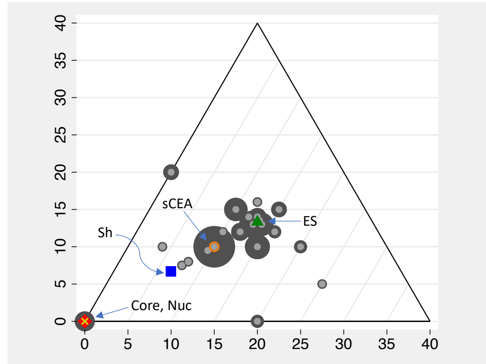
Figure 1 visualizes Decision Makers' choices for CF1-7 with frequency-weighted scatterplots in imputation triangles. Imputation triangles are commonly used in the cooperative games literature, and are read as follows: R3's payoff is given by the vertical axis, R2's payoff is read from the diagonal indifference lines emanating from the horizontal axis, and R1's payoff is what remains from the total. Hence the top (bottom right, bottom left) corner of the simplex corresponds to giving everything to R3 (R2, R1). For CF7, the only choice consistent with standard solution concepts is to split proceeds equally among the Recipients. Figure 1G shows that only 5 subjects who participated in CF7 chose an unequal allocation for that characteristic function. We drop these subjects from our ensuing analysis.[14]

[13] In the first couple of sessions, after everyone except one or two Decision Makers had completed all seven rounds, a connectivity issue with the server prevented the remaining Decision Makers from entering their choice in the final one or two rounds. Of course, the last round was always CF7. Since it was through no fault of their own, those few subjects were paid \$1 for each of those missing decisions. This did not affect any of the remaining payment process. The connectivity problem was ultimately identified and corrected. Aside from this, two Decision Makers voluntarily opted out of one round, and one opted out of three rounds. Letting n_i be the number of responses for CFi, we have $n_1 = 88$, $n_2 = 89$, $n_3 = 88$, $n_4 = 88$, $n_5 = 86$, $n_6 = 87$, $n_7 = 84$.

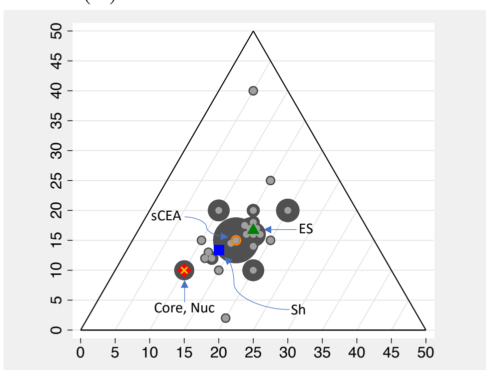
[14] These 5 subjects were also outliers in other characteristic functions. Some of their survey responses suggest a lack of understanding of basket worths or of the setting, or that they were intentionally allocating payoffs in an arbitrary manner; e.g., in describing how they made their choices in the exit survey, one of these five outliers wrote "Pretty arbitrary", and another explained that "i gave one person all of the money because i thought it would increase the recipients average earnings" (sic).



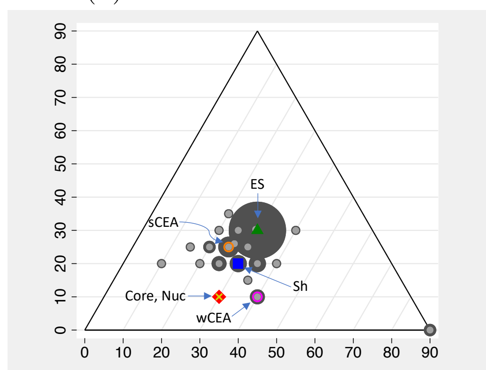
(A) Characteristic function 1



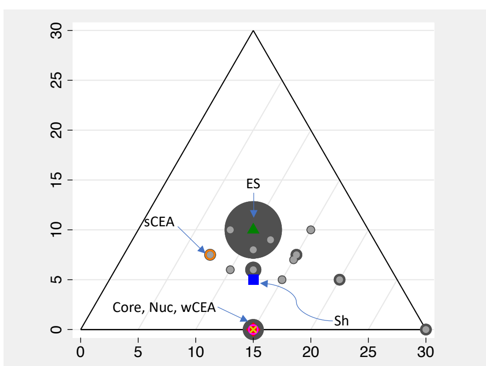
(B) Characteristic function 2



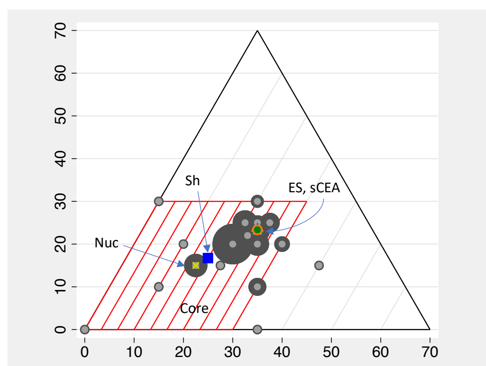
(C) Characteristic function 3



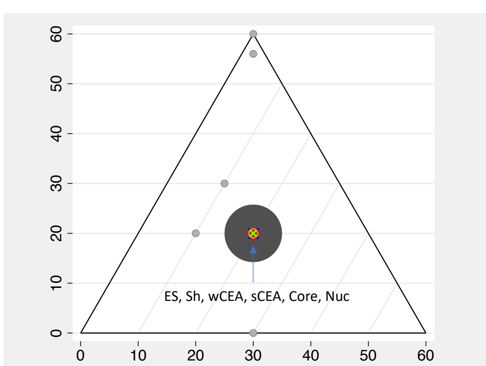
(D) Characteristic function 4



(E) Characteristic function 5



(F) Characteristic function 6



(G) Characteristic function 7

FIGURE 1: Frequency-weighted scatterplots of Decision Makers' allocations (with outliers). Vertical ticks give R3's payoff; R2's payoff is read through the diagonal indifference curves; R1's payoff is what remains. Each ball's radius is proportional to the fraction of Decision Makers who picked its center.

	CF1	CF2	CF3	CF4	CF5	CF6	CF7
Recipient 1	\$24.30 (0.73)	\$17.70 (0.73)	\$19.07 (0.52)	\$34.02 (0.70)	\$10.51 (0.23)	\$27.71 (0.70)	\$20 (0)
Recipient 2	\$24.36 (0.75)	\$11.42 (0.51)	\$15.22 (0.33)	\$29.04 (0.52)	\$11.09 (0.26)	\$21.57 (0.57)	\$20 (0)
Recipient 3	\$11.34 (1.05)	\$10.88 (0.45)	\$15.70 (0.47)	\$26.94 (0.54)	\$8.41 (0.33)	\$20.72 (0.53)	\$20 (0)
% Equal splits	41%	23.8%	18.1%	57.8%	65.4%	20.7%	100%
Observations	83	84	83	83	81	82	79

TABLE 5: Summary data. Average amounts allocated (standard errors in parentheses), along with percent of payoff allocations that are "equal splits" as defined by choosing payoffs for Recipients that differ by at most one dollar.

Since the imputation triangles are all the same size (only tick marks differ), they are comparable in terms of percentages of the total allocated to each recipient. The movement of the clouds of points across characteristic functions suggests that splits do vary with sub-coalition worths. CF1 and CF7 provide a particularly salient contrast, as they share the same total amount available but differ in the sub-coalition worths.

Table 5 summarizes the data more succinctly. The first three rows give the mean payoffs chosen by Decision Makers for each characteristic function. We see a clear departure from the equal split norm. For instance, the mean payoffs in CF1 and CF7 differ widely, although the available sum is \$60 in both cases.

Table 5 also reveals a substantial fraction of subjects depart from equal split in CF1-6. In CF2, CF3 and CF6, the worth of the grand coalition is not divisible by three. Decision Makers can input numbers with decimal places, but may find payments in whole dollars simpler. Throughout, we count an allocation as an *equal split* if the payoffs of Ri and Rj differ by at most one dollar, for all i, j. When the total worth is divisible by three (CF1, CF4, CF5 and CF7), everyone who satisfies our equal-split criterion splits *exactly* equally. There are 28 Decision Makers who split the money exactly equally in *all* four CFs where the total is divisible by three. Even allowing for differences of a dollar, the proportion of equal splits is lower in CF2, CF3 and CF6, where the total is not perfectly divisible. Imperfect divisibility might motivate Decision

Makers to think further about the problem, and take a closer look at sub-coalition worths. Alternatively, there may be a fraction of Decision Makers who desire an "equal split," but round in multiples of \$5 instead and don't discriminate regarding who get more. If one includes payoff allocations that differ by (at most) \$5, the percentages for CF2, CF3, and CF6 would be closer to the others in Table 5. But this may count too many people: in CF3, for instance, among Decision Makers satisfying the \$5-criterion but not the \$1-criterion, 73.7% choose the allocation $(\$20, \$15, \$15)$, which is compatible with rewarding the most desirable recipient R1, and treating the symmetric recipients R2 and R3 equally. That choice happens to be the sCEA in CF3, and is also exactly midway between equal split and the Shapley value.

The imputation triangles in Figure 1 mark the prediction(s) of the solution concepts from Section 2.2, and can be used to gauge how close choices are to those predictions within a single CF. To get a better sense of how close choices are to a solution concept when taking into account all CFs, one could attempt to classify each subjects' array of choices, based on the smallest mean-squared error, into either the nucleolus, the sCEA, equal split, the Shapley value, or a simple average of the Shapley value and equal split solution (to be parsimonious, as a continuum of weights are possible). For the 79 subjects who provided answers to all of CF1-6, we find that 3.8% of subjects are closest to the nucleolus, 3.8% are closest to sCEA, 43.0% are closest to equal split, 16.5% are closest to the Shapley value, and 32.9% are closest to a simple average of the Shapley value and equal split solution. The data, can, however, teach us more. Our ensuing analysis provides evidence that subjects' choices are not arbitrary, but are guided by basic normative principles.

4.1 Axioms

In this subsection, we provide the empirical support for the following result.

Result 1. *Overall, there is strong evidence for Additivity, Desirability, Monotonicity, and Symmetry. On the other hand, Dummy Player is rejected.*

Dummy player. According to the Dummy Player axiom, R3 should receive zero payoff in CF1. By contrast, Table 5 shows that R3 receives an average payoff of \$11.34, which is significantly different from zero at all conventional significance levels ($p = 0.0000$). When excluding equal splits, the average payoff of R3 is \$5.33, which is still significantly different from zero at all conventional significance levels ($p = 0.0000$). At the individual level, 34.9% of subjects satisfy the axiom, while nearly two-thirds of subjects violate it, such as by choosing equal split (41%) or a different convex combination of the equal split solution and the Shapley value (15.7%). Most in the latter category give \$10 to R3, and \$25 to each of the other recipients. There are several reasons why one may see few norms in CF1; for instance, the Shapley value is an element of the core, and coincides with the nucleolus.

Desirability. Remember that R1 $\succ$ R3 and R2 $\succ$ R3 in CF1 and CF5; R1 $\succ$ R2 and R1 $\succ$ R3 in CF2, CF3 and CF6; and R1 $\succ$ R2 and R2 $\succ$ R3 in CF4. Desirability requires the payoff difference between a more desirable and less desirable Recipient to be strictly positive. Assessing desirability is nontrivial, and Decision Makers could not rely on any patterns due to our use of randomly generated aliases for recipients. Equal splits automatically violate desirability, and we stack the deck against the axiom in our statistical analysis by including these in the sample.

Nonetheless, Table 5 shows that average payoff allocations respect all desirability comparisons, with more desirable Recipients allocated strictly higher average payoffs. As a first check, a paired-sample Hotelling's T-square test (a multivariate generalization of the paired t-test) rejects the joint null hypothesis that the twelve payoff differences are all zero ($p = 0.0000$), showing that Decision Makers do not treat all Recipients equally. We then examine each desirability ranking to see which are respected. For each CF and desirability comparison R$i \succ$ Rj, a paired t-test rejects the null hypothesis that the average payoff difference is zero, with $p \leq 0.0006$ in all cases except for equality of R2 and R3 in CF4 ($p = 0.0109$). Effect sizes as measured by Cohen's d

(Cohen, 1998) range from moderate to large, with only two exceptions.[15]

We then delve into potential heterogeneity in choices. The Wilcoxon signed-ranks test for distributions (which applies to dependent samples, as we have here) rejects each null hypothesis that the payoff distributions for desirability-ranked Recipients are the same ($p = 0.0000$ in all cases except for equality of R2 and R3 in CF4, where $p = 0.0371$). Figure 6 in the Online Appendix plots the empirical CDFs of monetary differences chosen by each Decision Maker for different Recipient pairs and CFs. These show the support of the distribution is almost entirely positive when desirability applies. Of course, individual noise on a given desirability ranking may be canceled in the aggregate, and Figure 7 in the Online Appendix indeed shows that the CDF of money to Ri generally first-order stochastically dominates that for Rj, when R$i \succ$ Rj. See Figure 2 in the text for an example of each plot in the case of CF6, where R1$\succ$R2,R3.

The imputation triangles in Figure 1 provide a view of Desirability at the individual level. In our setting, the axiom implies that chosen allocations lie strictly below the vertical axis passing through equal split (R1 and R2 get at least as much as R3, with at least one strict comparison) and lie to the left of the diagonal line passing through equal split (strictly to the left if R1 is more desirable than R2). Figure 1 provides overall support for this placement of allocations. Among nonequal splits, the percentage of individuals satisfying *all* desirability comparisons is 85.7% for CF1, 56.3% for CF2, 63.2% for CF3, 31.4% for CF4, 67.9% for CF5, and 55.4% for CF6. CF4 is more complex because no two players are symmetric. Still, 94.3% of subjects in CF4 respect at least two of the rankings R1 $\succ$ R2, R2 $\succ$ R3, and R1 $\succ$ R3 (those respecting the first two inequalities satisfy the third), and Figure 7 of the Online Appendix shows that violations are canceled out in the aggregate.

[15]This is the absolute value of the mean difference in payoffs divided by the standard deviation of this difference. A value of 0.2 is considered small, 0.5 is moderate, and 0.8 is large (Cohen, 1998). Except for an effect size of 0.2860 for R2 versus R3 in CF4, and 0.3930 for R1 versus R3 in CF3, all other effect sizes are in the interval $[0.4637, 0.8637]$.

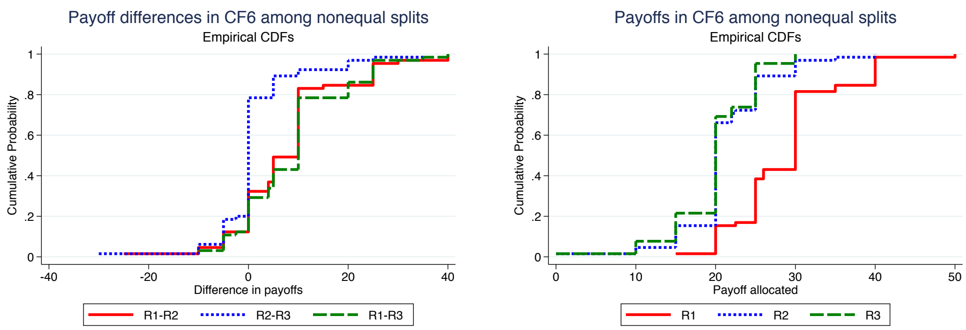


FIGURE 2: A view of desirability and symmetry in CF6, where R1≻R2∼R3.

Symmetry. Remember that R1 $\sim$ R2 in CF1 and CF5, and R2 $\sim$ R3 in CF2, CF3 and CF6. Symmetry requires symmetric Recipients to receive equal payoffs. Average payoff allocations in Table 5 appear to respect the axiom, with 'similar' average payoffs for symmetric Recipients. However, equal splits trivially satisfy symmetry and we exclude these when performing tests below.

For each CF, a paired t-test cannot reject the null hypothesis of zero payoff difference between symmetric recipients ($p = 0.1167$ for CF5, with p ranging from 0.3005 to 0.9544 otherwise).[16] All but one of the observed effect sizes are very small, with a small-to-moderate effect size for CF5: $d = 0.0082$ for CF1, $d = 0.1068$ for CF2, $d = 0.0964$ for CF3, $d = 0.3063$ for CF5, and $d = 0.1295$ for CF6. Taking into account the observed standard deviations and the number of observations left after dropping equal splits,[17] our minimum detectable effect sizes with 80% power are $d^s = 0.4085$ for CF1, $d^s = 0.3557$ for CF2, $d^s = 0.3447$ for CF3, $d^s = 0.5492$ for CF5, and $d^s = 0.3528$ for CF6.[18]

[16] One can also use the paired-sample Hotelling T-square test to check the joint null hypothesis that all 5 payoff differences between symmetric recipients are zero, conditional on nonequal splits in *all* these CFs, rather than each CF separately; again we cannot reject the null ($p = 0.3466$).

[17] After dropping equal splits in a CF, we remain with $n = 49$ for CF1, $n = 64$ for CF2, $n = 68$ for CF3, $n = 28$ for CF5, and $n = 65$ for CF6.

[18] When equivalence is the hypothesis of interest, an alternative to effect sizes is Schuirmann (1987)'s two one-sided tests. This approach is commonly used in the bio-statistics literature (e.g., to test whether a cheap new drug is as effective as an existing one) and also pointed out by List, Sadoff and Wagner (2011). After choosing some acceptable upper and lower bounds Δ_U and $-\Delta_L$ (often symmetric), two composite null hypotheses about the difference in means, Δ, are tested: $\Delta \leq -\Delta_L$ and $\Delta \geq \Delta_U$. Equivalence is

Taking into account heterogeneity in choices, we cannot reject equality of the distribution of payoffs to symmetric recipients in any CF, using the Wilcoxon signed-ranks test ($p = 0.1956$ for CF5, with p-values ranging from 0.4894 to 0.7673 otherwise). To get a sense of monetary differences in a Decision Maker's choices for symmetric Recipients among nonequal splits, we again refer the reader to Figure 6 in the Online Appendix. This shows the CDFs of differences between symmetric Recipients tend to be symmetric around a mode of zero. Again, noise may be canceled in the aggregate, and Figure 7 in the Online Appendix shows that the CDFs of money allocated to symmetric Recipients are quite similar. See Figure 2 in the text for an example of each plot in the case of CF6, where R2$\sim$R3.

In the imputation triangles, symmetry implies that chosen allocations should lie on the vertical bisector of the triangle for CF1 and CF5, and on the diagonal bisector emanating from the left for CF2, CF3 and CF6. Figure 1 again provides support for this placement of allocations. Among nonequal splits, the percentage of individuals satisfying symmetry is 81.6% for CF1, 57.8% for CF2, 67.6% for CF3, 57.1% for CF5, and 56.9% for CF6.

Additivity. Additivity has two implications for Recipients' payoffs: (1^{st}) payoffs in CF6 should be exactly \$10 higher than in CF2, and (2^{nd}) payoffs in CF3 should be the average of payoffs in CF2 and CF7. Letting $m_{\mathrm{R}i}(k)$ denote Ri's payoff in characteristic function k, we say $m_{\mathrm{R}i}(6) - (m_{\mathrm{R}i}(2) + 10)$ is the deviation from the 1^{st} additivity implication for Recipient i, and $m_{\mathrm{R}i}(3) - 0.5(m_{\mathrm{R}i}(2) + m_{\mathrm{R}i}(7))$ is the deviation from the 2^{nd} additivity implication for Recipient i. Those choosing equal splits in both CF2 and CF6, or both CF2 and CF3, trivially have zero deviation in the corresponding additivity impli-

declared when both one-sided tests are rejected. Fixing a type-I error rate of α, the outcome is equivalent to declaring equivalence when the $(1 - 2\alpha)$% confidence interval for the difference in means is contained in the acceptable interval $[-\Delta_L, \Delta_U]$. Hence the $(1 - 2\alpha)$% confidence interval for the difference is the smallest interval $[-\Delta_L, \Delta_U]$ where equivalence can be declared. With $\alpha = 0.05$, the 90% confidence intervals for the monetary difference between symmetric recipients, conditional on a nonequal split, are [-\$3.08,\$2.88] for CF1, [\$-0.66, \$2.03] for CF2, [\$-1.89,\$0.67] for CF3, [\$-3.44,\$0.09] for CF5 and [\$-0.67,\$2.90] for CF6.

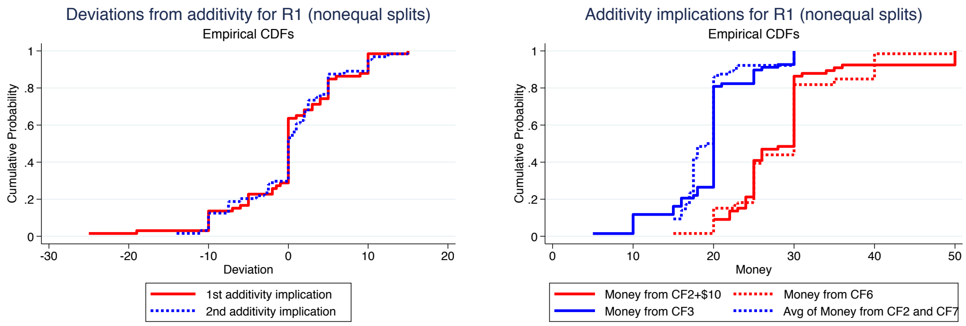


FIGURE 3: Empirical CDFs of additivity deviations for R1 on the left, and CDFs of payoffs for each side of the additivity equations for R1 on right.

cation, so we exclude these from our statistical tests.

Average payoffs appear to satisfy the additivity implications even after dropping such equal splits, with the largest average deviation around fifty cents. For each Recipient and additivity implication, a paired t-test cannot reject the null that the average deviation is zero (p-values ranging from 0.3482 to 0.9930).[19] The effect sizes we find are all very small ($d = 0.0011/0.0336/0.0345$ for R1/R2/R3 in the 1^{st} additivity implication, and $d = 0.0291/0.1181/0.0648$ for R1/R2/R3 in the 2^{nd}). Taking into account the observed standard deviations and the number of observations remaining in the sample,[20] our minimum detectable effect sizes with 80% power are $d = 0.3501$ for each Recipient in the 1^{st} implication, and $d = 0.3557$ for each Recipient in the 2^{nd} implication.[21]

We next consider distributions of choices. Even among only those Decision Makers who choose an unequal split in at least one of CF2 or CF6, the Wilcoxon signed-ranks test cannot reject for any $i = 1, 2, 3$ the null hy-

[19] One can also examine the joint null hypothesis that all deviations are zero using the paired-sample Hotelling T-square test, conditional on unequal splits in at least two of CF2, CF3 and CF6 (so neither implication holds trivially). This too cannot reject the null ($p = 0.7838$).

[20] We have $n = 66$ for the 1^{st} additivity implication and $n = 64$ for the 2^{nd} implication.

[21] When the null is the hypothesis of interest we can also use Schuirmann (1987)'s two one-sided tests, as in Footnote 18. With $\alpha = 0.05$, the 90% confidence intervals for the deviation give the smallest bounds where equivalence may be declared. For the 1^{st} implication, these are [-$1.42,$1.44] for R1, [$-1.20, $1.67] for R2, and [$-1.68,$1.20] for R3; for the 2^{nd}, these are [$-1.11,$1.47] for R1, [$-1.40, $0.39] for R2, and [$-0.73,$1.39] for R3.

pothesis that $m_{\mathrm{R}i}(6)$ and $m_{\mathrm{R}i}(2)+10$ come from the same distribution; and similarly, it cannot reject for any $i = 1, 2, 3$ the null hypothesis that $m_{\mathrm{R}i}(3)$ and $0.5(m_{\mathrm{R}i}(2)+m_{\mathrm{R}i}(7))$ come from the same distribution, among those who choose an unequal split in at least one of CF2 or CF3. The p-values for these tests range from 0.2297 to 0.9080. For these subjects, Figure 8 in the Online Appendix visualizes the empirical CDFs of both the deviations and payoffs for each additivity implication; see Figure 3 in the text for examples of these plots for R1. We see some heterogeneity or noise in terms of the deviations, though there are 17 (7) subjects who satisfy the 1^{st} (respectively, the 2^{nd}) implication with exact equality for *all* recipients. These deviations mostly wash out in the aggregate, as we see that the empirical CDF of $m_{\mathrm{R}i}(6)$ is similar to that of $m_{\mathrm{R}i}(2)+10$, and the CDF of $m_{\mathrm{R}i}(3)$ is similar to that of $0.5(m_{\mathrm{R}i}(2)+m_{\mathrm{R}i}(7))$.

Monotonicity. The monotonicity axiom has multiple implications here: $m_{\mathrm{R}i}(6)-m_{\mathrm{R}i}(2) > 0$ for $i = 1, 2, 3$; $m_{\mathrm{R}i}(3)-m_{\mathrm{R}i}(2) > 0$ and $m_{\mathrm{R}i}(7)-m_{\mathrm{R}i}(2) > 0$ for $i = 2, 3$; $m_{\mathrm{R}1}(6) - m_{\mathrm{R}1}(3) > 0$ and $m_{\mathrm{R}1}(2) - m_{\mathrm{R}1}(5) > 0$; and finally $m_{\mathrm{R}3}(7) - m_{\mathrm{R}3}(1) > 0$. Average payoffs do satisfy these restrictions. The equal split solution happens to satisfy all these instances of monotonicity except for the last one (R3 gets \$20 in both cases). Even including those who split equally in CF1, and excluding those who split equally in CF2 and either CF3 or CF6 (so that none of the ten implications can be trivially satisfied), we can reject the joint null hypothesis that the payoff differences are all zero using the paired-sample Hotelling T-square test ($p = 0.0000$). Delving into the ten implications separately using paired t-tests to see which differences are nonzero, we reject every null ($p = 0.0000$ in all cases). Seven out of the ten effect sizes are larger than 1, with the smallest effect size being 0.6381. Taking a closer look at individual behavior, Figure 9 in the Online Appendix shows that the empirical CDFs of the above payoff differences are very heavily skewed towards positive numbers. Unsurprisingly then, for each combination Ri, CFj and CFk where monotonicity applies, the Wilcoxon signed-ranks test rejects at all standard significance levels the null hypothesis that $m_{\mathrm{Ri}}(j)$ and $m_{\mathrm{Ri}}(k)$ come from the same distribution.

4.2 Emerging solution concept

The classic characterization of the Shapley value is based on Additivity, Efficiency, Symmetry, and Dummy Player (Shapley, 1953). Having to allocate all the money means Efficiency is automatically satisfied here. In view of Result 1, it is natural to ask which class of solution concepts emerges if we drop the Dummy Player axiom from Shapley's characterization. A clean characterization emerges for the domain V of three-player characteristic functions where the worth of each coalition is a rational number, and singleton coalitions are worth nothing (naturally, V contains CF1-7). The proof of the following observation is straightforward, and appears in Online Appendix B.

Observation 1. *A single-valued solution concept $\sigma : V \to \mathbb{R}^3$ is Additive, Symmetric, and Efficient if and only if σ is a linear combination of the Shapley value and the equal split solution, that is, $\sigma = \delta Sh + (1-\delta)ES$. Moreover, δ is positive if and only if σ satisfies either Monotonicity or Desirability.*

Thus the data singles out a simple, one-parameter solution concept. Under this model, payoffs are determined by a fixed affine combination of equal split and the Shapley value, *with δ independent of Recipients and characteristic functions*. We will call this the *Equal-Split Shapley (ESS) model*. Recipients start on equal footing, and then gain (lose) δ dollars for each dollar by which the Shapley value is larger (smaller) than equal split: $\delta = \frac{\sigma^i(v)-ES^i(v)}{Sh^i(v)-ES^i(v)}$, for any characteristic function v and any Recipient i such that $Sh^i(v) \neq ES^i(v)$.

The tests of axioms performed earlier check whether the data is consistent with particular *instances* of these axioms, as they apply to the characteristic functions studied here. On the other hand, the ESS model relies on the axioms being satisfied universally by choices, which is not directly testable. For this purpose, we fit the model to the data and see that it provides a nice match. To estimate δ, it is useful to avoid a regression in levels (so CFs with larger monetary amounts do not unduly influence our estimates) by using an equivalent rephrasing the model in terms of percentage departure from equal

	(all) $\frac{m-\text{ES}}{ES}$	(no D-equal splitters) $\frac{m-\text{ES}}{ES}$	(D-equal splitters) $\frac{m-\text{ES}}{ES}$
$\frac{\text{Sh}-\text{ES}}{ES}$	0.368^{***}	0.521^{***}	0.065^{*}
	(0.039)	(0.044)	(0.028)
constant	-0.005	-0.012	0.008^{*}
	(0.005)	(0.007)	(0.003)
No. subjects	84	56	28
Observations	1150	766	384
R^2	0.3003	0.4297	0.0450

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

TABLE 6: Regressions of the percentage departure from equal split of allocations to R1 and R3, against percentage departure of the Shapley value from equal split. Huber-White heteroscedasticity-robust standard errors in parentheses, clustered by Decision Maker.

split:

$$\frac{\sigma_i(v) - ES_i(v)}{ES_i(v)} = \delta \left(\frac{Sh_i(v) - ES_i(v)}{ES_i(v)} \right),$$

for any Recipient i and any characteristic function v.[22] Because the ESS model applies the same δ to all recipients, we can pool data across them to estimate δ. Since the payoff of a Recipient can be inferred from the payoffs of the other two (the sum of all three payoffs is fixed per characteristic function), we only consider choices for R1 and R3 per characteristic function (we drop R2 since their Shapley value coincides with equal split in CF4, leading to less exploitable variation). We use a generalization of the Huber-White sandwich estimator of errors that is not only robust to heteroscedasticity, but also clustered at the level of the Decision Maker to permit for correlation across his or her choices (Rogers, 1993). The first column in Table 14 provides the estimation results among all Decision Makers, with the weight on the Shapley value $\delta = 0.368$ significantly different from zero.

For each CF, we know there is a fraction of Decision Makers who split equally. We can define a Decision Maker as an *equal splitter* if she picks equal split in *every* CF. By definition, the δ of these individuals under the ESS model will be near zero. We can also single out a larger class of Decision Makers: say a subject is a *D-equal splitter* if she splits *exactly* equally in *all four* characteristic

[22]This is also equivalent to normalizing by the total $v(\{1,2,3\})$ available; that would be interpreted as the departure from equal split as a percent of the pie.

functions where the total worth is divisible by three. Through their choices, D-equal splitters reveal themselves as having a strong tendency towards equal splits. We find 10 equal splitters and 28 D-equal splitters, with the former set nested in the latter: they split *exactly* equally, not just within \$1, in CF1, CF4, CF5 and CF7. Some of these subjects may round payments by multiples of \$5 instead of \$1 when the worth of the grand coalition is not divisible by three, but others may choose with subcoalitional worths in mind. The third column in Table 14 shows the estimated δ among D-equal splitters is small but different from zero at the 5% level. A few of these subjects seem to follow a more intricate model of choice than ESS: they sometimes select reasonable payoff allocations that are far from equal splits when the worth of the grand coalition is not divisible by three.[23] We find it interesting to document these behaviors, though they are unusual and have limited impact on our analysis.

Given that D-equal splitters comprise a sizable fraction of subjects with δ near zero, it must be that other Decision Makers use a larger δ than appears in the first column of Table 14. Indeed, the estimate of δ among those subjects is close to one-half, suggesting that they are well-described by a simple average of equal split and the Shapley value.

Given the heterogeneity uncovered above, one may be interested in aggregating opinions to determine what the society *as a whole* views as an appropriate allocation in each CF.[24] The simple average is perhaps the most natural means of aggregation: Rubinstein and Fishburn (1986) show it is the *only* aggregator that picks the common opinion when all Decision Makers agree, that

[23] One such D-equal splitter is within \$5 in all other characteristic functions, with one exception: they choose $(\$40, \$0, \$0)$ in CF2, following the nucleolus in respecting the extreme competition between R2 and R3 for cooperation with R1.

[24] In Online Appendix H we consider some possible sources (or correlates) of heterogeneity in δ: interaction effects with a Decision Maker's major, age, gender, and number of siblings, or session effects (e.g., arising from the ordering of characteristic functions in the Latin square design). We find that a Decision Maker's gender has no statistically significant impact on δ; nor does their number of siblings. Being an economics-related major may have some impact: an increase in δ of about 0.2, significant only at the 5% level. Further interaction with age reveals that the effect is significant only for subjects who are at least 20 years old, and presumably more advanced in their studies. We thus suspect that the effect has more to do with coursework in economics than personal traits, but cannot draw any definitive conclusions using the sparse education data we collected.

is efficient, and for which a Recipient's payoff depends only on the amounts Decision Makers' allocated to him. The ESS model remains useful in this case. Indeed, if each Decision Maker j (noisily) follows the ESS model, then the *average allocation* also follows the ESS model, with the societal δ given by the average of individual δ_j's.[25] The close adherence of average payoffs to the axioms provides further support for an ESS model of societal opinion.

5 Concluding Remarks

We conclude by examining whether the qualitative assessment of the axioms, and the usefulness of the ESS model, remain valid in other contexts. Payoff allocations, and thus parameter estimates, could plausibly vary with the context in which the characteristic functions arose. As an analogy, expected utility theory can be helpful to explain choices in various contexts of choice under risk, though risk attitude may be context-dependent.[26] In his survey of positive analyses of distributive justice, Konow (2003) argues that justice is "context dependent, but not context specific." General principles hold widely (qualitative results in our context), while "context is the indispensable element that supplies the people, variables, time framework and weighting of principles that result in justice preferences" (parameter estimates in our context). For the ESS model, all that is needed for accurate predictions in a given context is to test choices in just a few–or even as little as one–characteristic functions, *within that same context*, to assess the weight on the Shapley value.

These considerations lead to some interesting questions. First, if coalitional worths were determined randomly instead of through a quiz, would Decision Makers simply split the pie equally, or would they still take coalition worths into account? Though we had conjectured the former, there are arguments for the latter. For instance, one may want to more greatly reward a band member who plays an important role in drawing audiences, even if that ability is mostly

[25]We note averaging is purely an ex-post exercise: recipients were not paid according to such averages, nor were they mentioned to subjects.

[26]See Barseghyan, Prince and Teitelbaum (2011) and Einav, Finkelstein, Pascu and Cullen (2012).

attributable to luck (e.g. appearance, innate vocal talent). Decision Makers may also select what they think would be the likely outcome if Recipients were to bargain. In that case, coalitional worths represent outside options and their origins are irrelevant. Second, given that Decision Makers are not informed of Recipients' quiz performance, prizes and their worths, did our design create a sense of earned worths that impacted Decision Makers' choices? Third, does luck versus desert play a role in our allocation problems (as it does in other settings, see e.g. discussion in Section 4 of Konow (2003))? We subsequently created another two related designs to shed light on these questions. The data analysis supporting the discussion below is available in Online Appendix G.

The 'No-Quiz' treatment differs from our original experiment – from now on called the 'Quiz' treatment – in only one respect: the same electronic baskets that were earned in the Quiz treatment are simply assigned randomly, leading to the same subcoalitional worths. These two treatments are thus directly comparable. By eliminating effort entirely, the No-Quiz treatment eliminates any uncertainty about the extent of meritocracy. As we anticipated, our qualitative results regarding the axioms and the usefulness of the ESS model are replicated to a large degree by the No-Quiz treatment. Perhaps surprisingly, the quantitative results are remarkably similar too. Theoretically, it could mean that estimated parameters are context independent. Alternatively, not knowing how challenging the quiz was, nor the precise mapping between earned fictitious objects and coalition worths, it could be that many Decision Makers in the quiz treatment viewed characteristic functions as if they were randomly assigned. Still, this would show that many Decision Makers take coalition worths into account independently of their origin.

To further test whether our qualitative results are portable across a wide variety of contexts, we designed a third, more radically different treatment. Both the Quiz and No-Quiz treatments generate coalition worths somewhat abstractly through baskets combinations of fictitious objects. Would we still find that the ESS model, and its underlying axioms, help organize choices if coalition worths arise in a context more relatable to real-life situations? And if so, would the pull towards the Shapley value be quantitatively different? To study these questions, we turn to the long tradition of vignettes in a strand of the experimental literature on distributive justice: see, for instance, the classic papers of Yaari and Bar-Hillel (1984), Kahneman, Knetsch and Thaler (1986), Levine (1993), or many other papers reviewed in Konow (2003)'s survey, which also discusses benefits and drawbacks of the method.[27] A vignette provides subjects with contextual information on a realistic problem, and asks them to make a decision for that circumstance. They are intended to help participants understand, relate and think through a problem. In our setting, the hope is to make the characteristic function come to life in a practical problem.

The vignette we test is based on the musicians in the first paragraph of the introduction. We use the 'same' characteristic functions as in the first two treatments, but multiply all coalition worths by 10 for the vignette to be plausible. In our 'Vignettes' treatment, all subjects are Decision Makers (and paid per decision, as before). The three musicians in a vignette are the hypothetical Recipients. Unlike our other treatments, Decision Makers' choices are never implemented. Since their choice matters to no one but themselves, and they are paid a fixed amount regardless of the allocation selected, some might expect Decision Makers to avoid thinking costs: for instance, simply allocating the entire amount to one musician, or always splitting equally. However, it is well documented that subjects take vignettes seriously (Konow, 2003).

Indeed, we again find that subjects take coalition worths into account, and that cooperative game theory provides a useful way to organize the data. We find extremely similar *qualitative* results, but uncover quantitative differences, with a greater pull away from equal split: the estimated weight on the Shapley value in the Vignettes treatment is about 50% larger than in the Quiz and No-Quiz treatments. This is reflected in a comparison across treatments of the CDFs of money allocated to each Recipient. For each characteristic function and Recipient whose Shapley value is greater than (smaller than) the amount they would receive from equal split, the CDF of money allocated to them in the

[27]This treatment thus contributes to that literature, which also employs an impartial observer approach (often using the terminology 'benevolent dictator'). Unlike our work, this experimental literature does not consider sub-coalition worths, and thus overlooks potential complementary or substitutability of agents.

Vignettes treatment nearly first-order stochastically dominates (is dominated by) the CDFs from the other treatments. Most of these rankings are highly statistically significant. Thus the fact that behavior was overall similar in the Quiz and No-quiz treatments suggests that coalition worths were mostly interpreted as determined by luck in our Quiz treatment.

Our three treatments provide robust evidence that coalition worths do matter in settings where agents can vary in how substitutable or complementary they are, and that the ESS model and its underlying axioms are important tools for organizing the data. The Vignettes treatment provides some evidence that parameter estimates may vary across contexts. This opens directions for future research. First, one may want to better understand how parameter estimates might vary across contexts, by drawing connections to theories of desert in the distributive-justice literature. For instance, Buchanan (1986) contrasts luck, choice, effort, and birth as distinct categories that impact one's claim to wealth; see also Konow (2003, Section 4.2). Second, one could test and calibrate the ESS model with different subject pools. Interestingly, while Croson and Gneezy (2009)'s survey highlights robust gender differences in risk, other-regarding and competitive preferences, we find no statistically significant differences in the parameter estimates across men and women. Exploring this further, and testing for cultural differences, would be of interest.

References

Ambuehl, Sandro, B. Douglas Bernheim and Axel Ockenfels, 2020. What Motivates Paternalism? An Experimental Study. *American Economic Review*, forthcoming.

Andreoni, James, and John H. Miller, 2002. Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism. *Econometrica* **70**, 737-753.

Barseghyan, Levon, Jeffrey Prince and Joshua C. Teitelbaum, 2011. Are Risk Preferences Stable across Contexts? Evidence from Insurance Data. *American Economic Review* **101**, 591-631.

Bloch, Francis, and Geoffroy de Clippel, 2010. Cores of Combined Games. *Games and Economic Behavior* **145**, 2424-2434.

Bolton, Gary E., Kalyan Chatterjee, and Kathleen L. McGinn, 2003. How Communication Links Influence Coalition Bargaining: A Laboratory Investigation. *Management Science* **49**, 583-598.

Bolton, Gary E., and Axel Ockenfels, 2000. ERC: A Theory of Equity, Reciprocity, and Competition. *American Economic Review* **90**, 166-193.

Buchanan, James M., 1986. Liberty, Market and State: Political Economy in the 1980s. NY: NYU. Press, Columbia U. Press.

Camerer, Colin F., 2003. Behavioral Game Theory: Experiments in Strategic Interaction. Princeton University Press.

Cappelen, Alexander W., Astri D. Hole, Erik Ø. Sørensen, and Bertil Tungodden, 2007. The Pluralism of Fairness Ideals: An Experimental Approach. *American Economic Review* **97**, 818-827.

Cappelen, Alexander W., James Konow, Erik Ø. Sørensen, and Bertil Tungodden, 2013. Just Luck: An Experimental Study of Risk-Taking and Fairness. *American Economic Review* **103**, 1398-1413.

Cappelen, Alexander W., John List, Anya Samek, and Bertil Tungodden, 2020. The Effect of Early-Childhood Education on Social Preferences. *Journal of Political Economy* **128**, 2739-2758.

Charness, Gary, and Matthew Rabin, 2002. Understanding Social Preferences with Simple Tests. *Quarterly Journal of Economics* **117**, 817-869.

Casajus, André and Frank Huettner, 2013. Null Players, Solidarity, and Egalitarian Shapley Values. *Journal of Mathematical Economics* **49**, 58-61.

Croson, Rachel and Uri Gneezy, 2009. Gender Differences in Preferences.*Journal of Economic Literature* **47**, 448-474.

Dutta, Bhaskar, 1990. The Egalitarian Solution and Reduced Game Properties in Convex Games. *International Journal of Game Theory* **19**, 153-169.

Dutta, Bhaskar, and Debraj Ray, 1989. A Concept of Egalitarianism Under Participation Constraints. *Econometrica* **57**, 615-635.

Dutta, Bhaskar, and Debraj Ray, 1991. Constrained Egalitarian Allocations. *Games and Economic Behavior* **3**, 403-422.

Einav, Liran, Amy Finkelstein, Iuliana Pascu, and Mark R. Cullen, 2012. How General Are Risk Preferences? Choices under Uncertainty in Different Domains. *American Economic Review* **102**, 2606-2638.

Fehr, Ernst, and Klaus M. Schmidt, 1999. A Theory of Fairness, Competition and Co-operation. *Quarterly Journal of Economics* **114**, 817-868.

Fisman, Raymond, Shachar Kariv, and Daniel Markovits, 2007. Individual Preferences for Giving. *American Economic Review* **97**, 1858-1876.

Gillies, Donald B., 1959. Solutions to General Non-Zero-Sum Games. In: Contributions to the Theory of Games (Vol. IV), ed. by A.W. Tucker and R.D. Luce, 47-85. Princeton University Press.

Kahneman, Daniel, Jack L. Knetsch, and Richard Thaler, 1986. Fairness as a Constraint on Profit Seeking: Entitlements in the Market. *American Economic Review* **76**, 728-741.

Kalisch, Gerhard K., John W. Milnor, John F. Nash, and Evar D. Nering, 1954. Some experimental n-person games. Decision Processes, ed. by R.M. Thrall, C.H. Coombs and R.L. Davis, 301-327. Wiley, NY.

Karni, Edi, and Zvi Safra, 2002. Individual Sense of Justice: A Utility Representation. *Econometrica* **70**, 263-284.

Konow, James, 2000. Fair Shares and Cognitive Dissonance in Allocation Decisions. *American Economic Review* **90**, 1072-1091.

Konow, James, 2003. Which Is the Fairest One of All? A Positive Analysis of Justice Theories. *Journal of Economic Literature* **41**, 1188-1239.

Levine, David I., 1993. Fairness, Markets, and Ability to Pay: Evidence from Compensation Executives. *American Economic Review* **83**, 1241-1259.

List, John A. and Sadoff, Sally and Wagner, Mathis, 2011. So you want to run an experiment, now what? Some simple rules of thumb for optimal experimental design. *Experimental Economics* **14**, 439-457.

Maschler, Michael, and Bezalel Peleg, 1966. A Characterization, Existence Proof and Dimension Bounds for the Kernel of a Game. *Pacific Journal of Mathematics* **18**, 289-328.

Moulin, Hervé, 2003. Fair Division and Collective Welfare. MIT Press.

Murnighan, J. Keith, and Alvin E. Roth, 1977. The Effects of Communication and Information Availablity in an Experimental Study of a Three-Person Game. *Management Science* **23**, 1336-1348.

Nash, John F., Rosemarie Nagel, Axel Ockenfels, and Reinhard Selten, 2012. The Agencies Method for Coalition Formation in Experimental Games. *PNAS* **109**, 20358-20363.

Peleg, Bezalel, 1986. On the Reduced Game Property and its Converse. *International Journal of Game Theory* **15**, 187-200.

Potters, Jan A., 1991. An Axiomatization of the Nucleolus. *International Journal of Game Theory* **19**, 365-373.

Rogers, W. H., 1993. sg17: Regression Standard Errors in Clustered Samples. *Stata Technical Bulletin* **13**, 19-23.

Rubinstein, Ariel, and Peter C. Fishburn, 1986. Algebraic Aggregation Theory. *Journal of Economic Theory* **38**, 63-77.

Schmeidler, David, 1969. The Nucleolus of a Characteristic Function Game. *SIAM Journal of Applied Mathematics* **17**(6), 1163-1170.

Schuirmann, Donald, 1987. A comparison of the Two One-Sided Tests Procedure and the Power Approach for assessing the equivalence of average bioavailability. *Journal of Pharmacokinetics and Biopharmaceutics* **15**, 657–680.

Shapley, Lloyd S., 1953. A Value for n-Person Games. In: Contributions to the Theory of Games (Vol. II), ed. by A. W. Tucker and R. D. Luce, 307-317. Princeton University Press.

Sobel, Joel, 2005. Interdependent Preferences and Reciprocity. *Journal of Economic Literature* **43**, 392-436.

von Neumann, John, and Oskar Morgenstern, 1944. Theory of Games and Economic Behavior. Princeton University Press.

van den Brink, René, Yukihiko Funaki and Yuan Ju, 2013. Reconciling Marginalism with Egalitarism: Consistency, Monotonicity, and Implementation of Egalitarian Shapley Values. *Social Choice and Welfare* **40**, 693-714.

Winter, Eyal, 2002. The Shapley Value. *Handbook of Game Theory with Economic Applications*, **3**, 2025-2054.

Yaari, Menahem E., and Maya Bar-Hillel, 1984. On Dividing Justly.

Social Choice and Welfare **1**, 1-24.

Yan, Huibin, Daniel Friedman and David Munro, 2016. An experiment on a core controversy. *Games and Economic Behavio* **96**, 132-144.