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

Geoffroy de Clippel† Kareen Rozen‡

May 2019

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. Three treatments establish the robustness of this qualitative conclusion, while also illustrating how context may impact the parameter estimate.

*First version October 2013. We are grateful to the editor, three referees, Pedro Dal Bó, Mark Dean, Kfir Eliaz, Louis Putterman, Larry Samuelson, Roberto Serrano, Rajiv Vohra, 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 there is to share when they all cooperate, and what surplus would be created should smaller groups form. 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?

Understanding people's views when allocating money in such settings is important, both for its own sake, and to shed light on the right reference point to use when assessing intentions and reciprocity. Intuitively, an agent's role in creating surplus may be judged as less important in the presence of other similar agents, which in turn may impact her reward for collaborating. By contrast, an agent judged as playing a more critical role might be rewarded more. To illustrate, some might allocate more to Musician 1 in the above example, since substituting her with another musician always reduces the surplus. On the other hand, amounts allocated to Musicians 2 and 3 might increase if the worth of their duo increases. For another thought exercise, contrast the two following situations: (a) a positive surplus (of $900) is generated if and only if at least Musicians 1 and 2 participate, *versus* (b) a positive surplus (of $900) is generated if and only if Musician 1 collaborates with at least one other musician. In (b), Musician 2 loses the critical role she had in (a), as she becomes substituable by Musician 3. Would Musician 2's allocated payoff in (b) be smaller than in (a)?

The long economic literature on other-regarding preferences overlooks such complementarity and substitutability among agents.[1] If one were to apply

[1]See Fehr and Schmidt (1999), Bolton and Ockenfels (2000), Andreoni and Miller (2002),

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). Through a series of experimental treatments, we provide 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. Many of the solution concepts and axioms also have a normative interpretation, capturing different ways of rewarding people in settings with complementarity and substitutability.[2] We test whether basic axioms are satisfied, use that information to focus on a meaningful class of solution concepts, and sharply organize the data through these theoretical tools.

Our initial treatment has three subjects designated at the start of each session as *Recipients*, and the remaining subjects 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 Decision Makers to decide how to split the worth of the grand coalition between the three Recipients. At the end of the session, the three Recipients are paid according to one randomly selected decision of one randomly selected Decision Maker. Our experimental design ensures that 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 may be a concern).

Our data in this initial treatment, which comprises six sessions (per a Latin Square design), shows that a large fraction of Decision Makers take the worths

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 from this normative perspective.

of subcoalitions into account when allocating money. While nearly all Decision Makers choose equal split when the characteristic function is symmetric (and all the solution concepts we study agree on this), these same Decision Makers often choose unequal splits when the characteristic function is asymmetric. Analyzing average payoff allocations as a start, we find 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.

From the perspective of theory, we prove that satisfying Symmetry and Additivity (along with Efficiency, which must be satisfied in our experiment) means that Decision Makers' choices must be a linear combination of the equal split solution and Shapley value, with the weights summing to one.[3] This theoretical result suggests a one-parameter model – the *ESS* model – which we fit to the data by linear regression. Despite its simplicity, it does a nice job at explaining average choices. The estimated weight on the Shapley value is around 37% (p-value < 0.001).

We then take a closer look at the individual-level data to examine potential heterogeneity. For each characteristic function, a significant fraction of observed payoff allocations fall on, or near, the line joining the Shapley value to equal split; but differ in how far along the line they go. A fraction of Decision Makers abide by the equal split solution, but a majority take sub-coalition worths into account. We find elucidating evidence of behavioral regularity using the empirical CDF's of money allocated to Recipients, and perform statistical tests of axioms using these distributions. The evidence corroborates that found in the analysis of averages. We estimate the ESS model at the aggregate level, both with and without the sub-population of Decision Makers we call *D-equal splitters* (who opt for equal split whenever the total is divisible by three). In aggregate, D-equal splitters may slightly reward marginal contributions when the total isn't divisible, but to a far lesser extent than non D-equal splitters. Delving more deeply into individual behavior, we run a horse-race based on MSE between the ESS model, the nucleolus, and the

[3]This result holds for the set of 3-player characteristic functions studied here.

strong constrained egalitarian allocation (Dutta and Ray, 1991), which seem to have some adherents in the scatterplots of data. Regardless of whether we allow a continuum of weights in the ESS model, or only four (allowing equal split, Shapley value, and two strict mixtures with weights 1/3 and 2/3, respectively, on Shapley), the analysis finds that the same 93% of subjects are best described by the ESS model. Intermediate weights matter beyond equal split and Shapley: for instance, in the discrete version of the analysis, 38%/27%/24%/5% of subjects are classified as placing a weight of $0/\frac{1}{3}/\frac{2}{3}/1$ on the Shapley value.

Throughout this first stage of the project, we presumed Decision Makers would overlook the worth of sub-coalitions, and simply pick equal splits, if those worths were purely determined by luck. Thus we designed the treatment discussed above to convey a sense of earned worths, while also reflecting a type of real-life uncertainty, whereby we often observe outcomes but not the process by which they were achieved (e.g., we may know little about whether someone's position in life is the outcome of hard work, lucky connections, or initial inheritance; or the relative role of training versus innate talent in someone's demonstrated skill). In this treatment, Decision Makers are told the values of basket combinations and know they were generated based on a quiz, but aren't given information about Recipients' performance or the mapping between performance and basket values.

While the Decision Makers in this treatment could not precisely gauge the level of meritocracy in coalition worths, our guess was that such uncertainty would be inconsequential as far as our qualitative results go. That is, the overall assessment of the axioms and the usefulness of the ESS model should remain valid, though it is plausible that payoff allocations and thus parameter estimates might vary with the context in which the characteristic functions arose.[4] Under the ESS model, all that is needed for accurate predictions

[4]Similarly, expected utility theory can be helpful to explain choices in various contexts of choice under risk, though risk attitude may be context-dependent (see Barseghyan, Prince and Teitelbaum (2011) and Einav, Finkelstein, Pascu and Cullen (2012)). In a similar vein, 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

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.

The Editor and referees rightly pushed us to understand whether conveying a sense of earned worth was even necessary: that is, would Decision Makers really just split equally if coalition worths were randomly assigned? As an analogy, some may want to more greatly reward a band member who plays an important role in drawing audiences, even if that ability is mostly attributable to luck (e.g. appearance, innate vocal talent, etc.). Also, as suggested by the Editor, how to split the pie as an impartial observer may be informed by what would/should happen if the Recipients were actually bargaining, in which case coalition worths matter independently of their origin.

To address this question, we tested a 'No-Quiz' treatment (an additional six sessions per the Latin square design), whose structure differs from our original treatment in only one respect: the same electronic baskets that were earned in the Quiz treatment are now simply assigned randomly, and basket combinations are worth the same as before. Hence the treatments test the same characteristic functions, and are directly comparable. As we anticipated, our earlier qualitative results regarding the axioms and the usefulness of the ESS model are replicated to a large degree by the No-Quiz treatment. To our surprise, 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 treated characteristic functions as if they were randomly assigned.

To further test our conjecture that qualitative results should be 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.

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" (as in the determination of parameter estimates in our context).

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 quite different technique of vignettes, which have a long tradition 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.[5] 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 particular vignette we test is based on the three musicians in our introductory paragraph. 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 (an additional six sessions per the Latin square design), all subjects are Decision Makers and are paid per decision made, as before. The three musicians in each vignette are the hypothetical Recipients. Contrary to the first two 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. In particular, we find extremely similar *qualitative* results, but uncover quantitative differences, with a greater pull towards the Shapley value, away from equal split. In the

[5]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.

aggregate, 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 vignettes treatment nearly first-order stochastically dominates (is dominated by) the corresponding CDFs from the first two treatments. Most of these FOSD rankings are highly statistically significant.

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 also provides some evidence that parameter estimates may vary across contexts. This opens a few promising 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 four distinct categories that impact one's claim to wealth; see also Konow (2003, Section 4.2). Second, one should 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. In addition to exploring this further, testing for cultural differences would be of interest.

Further related literature

We now discuss related literatures that have not already been noted above.

One interpretation of the Shapley value is that it rewards people for their role in creating the surplus, which according to Shapley, should be measured 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 their 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 can explain observed allocations when entitlement follows an accountability principle. Cappelen et al. studies the relative prevalence of fairness ideals beyond liberal egalitarianism, including strict egalitarianism and libertarianism. We study a wide range of scenarios that differ on multiple dimensions, as explained below.

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. As another point of departure, we provide no quantifiable information that would express coalition worths as a precise function of effort and luck parameters. In addition to keeping the analysis focused on our main point of interest – whether and how Decision Makers reward people for their role in creating the surplus – we see it as a realistic feature of some applications. Going back to the musicians, the opportunities available to them are quantifiable in terms of profit, but we would not expect even the musicians themselves, and a fortiori the impartial observers, to understand or agree on the differential impacts of talent and hard work in generating them.

Our work may also be contrasted with the small experimental literature on cooperative games, which allows multiple subjects to bargain with each other given a characteristic function. Kalisch, Milnor, Nash and Nering (1954), one of the earliest papers in the field of experimental economics, inform subjects of their role in a characteristic function and let them interact informally. Other experimental papers 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.

Another paper in that vein is Nash, Nagel, Ockenfels and Selten (2012), who compare outcomes in a 40-times repeated bargaining game to cooperative solutions. Their stage game starts with a coalition formation phase with exogenous breakdown probability, where following their 'agencies' protocol, one of the three players is appointed to allocate the coalition's profit (including to himself). With the grand coalition forming in 91.5% of stage games, they show that their bargaining protocol can successfully yield efficient outcomes when repeated: long-term incentives to maintain cooperation outweigh greedy short-term incentives. They study who is appointed to split the pie (e.g., is it the 'strongest' player in the characteristic function), and how he or she divides it. Their stage game admits multiple equilibria, and the finitely repeated game admits even more. A potential relation to our work is through the notion that a fair allocation can serve as a focal or reference point to select equilibria. If so, then observed outcomes might reflect, at least to some extent, which allocations are viewed as fair. The relation is tenuous, though, as little is known about focal points in complex strategic games, and many conflicting aspects play a role: players' choice may reflect, among other things, negative reciprocity,[6] reputation building, strategic experimentation (e.g., what will others tolerate), or end-game effects (e.g., will the last appointee take everything). They show splits vary widely as a function of the appointee, and in all rounds, the appointee either favors himself or splits equally. Thus, splits depart from all cooperative solutions whenever the appointee is not the 'strongest' player. Considering each characteristic function separately, they compare the mean-squared errors of the average split over 40 rounds (and the average split from the stronger player) to the Shapley value, nucleolus, and equal split solution.

[6] Consistent with this, they find: "the more aggressive the demand of one player is, the more aggressive are those of the others."

By contrast to these bargaining experiments, our study has no strategic considerations at play. We focus on one question that is posed to benevolent Decision Makers for each characteristic function: how would you split the pie for the three recipients? The answers allow us to test axioms, narrow down to a meaningful class of solutions, and study behavior as the characteristic function varies. Though it has not been formalized there, our question is thus closer in spirit to the aforementioned experimental literature on distributional justice, while our design and analysis borrow tools from cooperative game theory.

2 Theoretical Benchmark

Before detailing our treatments and results, we provide a quick primer on the theory which is in the background of our design and analysis.

2.1 Solution Concepts

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 coalition S a worth $v(S)$.[7] The amount $v(S)$ represents how much members of S can share should they cooperate. That is, an allocation 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? This is the central question of cooperative game theory.

The *equal-split* solution simply divides $v(I)$ equally among all individuals. 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).

[7]With the convention that $v(\emptyset) = 0$.

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

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

where $p_i(S) = \frac{|S|!(n-|S|-1)!}{n!}$ is the fraction of possible orderings where the set of individuals preceding i is exactly S. This formula also has an axiomatic foundation. The Shapley value is the only single-valued solution that satisfies Efficiency, Symmetric, Additivity and the Dummy Player axiom. Many alternative axiomatic characterizations have been proposed since then. Axioms are defined formally and discussed below, as we explain the rationale behind our selection of characteristic functions for the experiment. We will also test their validity experimentally.

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

The nucleolus. Like the Shapley value, the nucleolus 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 marries egalitarianism with the 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 then 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 be egalitarian themselves. The solution concept has two versions, Strong and Weak, which differ in what 'all prefer' means: in the Strong version (S-CEA), everyone must be strictly better off, while in the Weak version (W-CEA), all must be weakly better off, with at least one strict improvement. This seemingly small difference can yield very different predictions. Note that the S-CEA may be multi-valued and is always nonempty; but the W-CEA, when it exists, selects a unique allocation.

2.2 Normative Principles

We now turn our attention to some normative properties (or axioms) which may guide Decision Makers' choices, even if they do not follow one of the above solution concepts. A significant part of cooperative game theory precisely aims at defining such principles, and understanding which combinations characterize solution concepts. Some properties are satisfied by multiple reasonable 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 solution concepts, and thus sharply capture the essence of what distinguishes some solutions from others. Testing the axioms, in addition to examining the explanatory power and the relative prevalence of a handful of solution concepts, offers a fuller picture of what people view as fair.

Individual i is a *dummy player* if $v(S) = v(S \setminus \{i\})$, for any coalition S containing i. In order to test this property, we included in our study a characteristic function with a dummy player (as will be seen, this is Recipient 3 in CF1). The Dummy Player axiom requires that such individuals receive a zero payoff. It is satisfied by the Shapley value, the core, and thus any selection of it as well (such as the nucleolus for instance). The equal split solution, on the other hand, violates the Dummy Player axiom. Hence characteristic functions with a dummy player offer a stark test of the difference between equal split and most standard solutions from cooperative game theory.

Suppose that for any (non-singleton) coalition containing individual j but not i, replacing j with i strictly increases profit. In this case, we say that individual i is *more desirable* than j. If replacing j with i never makes a difference, we say that 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.[8] The Shapley value respects both Symmetry and Desirability. The core always contains payoff vectors that respect both Symmetry and Desirability, but may contain additional payoff vectors. The equal split solution respects Symmetry, but systematically violates Desirability. The constrained egalitarian allocations may violate both Symmetry and Desirability.

The properties above 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 $i \in S$. The Shapley value selects payoff vectors that systematically respect this property. Young (1985)

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

provides an example of two characteristic functions with singleton cores that violate Monotonicity. However, one can show that the core does admit a single-valued selection (e.g. the nucleolus) that respects Monotonicity for games with only three individuals, as in our experiment. The equal split solution violates Monotonicity since it overlooks the worths of sub-coalitions.

A cornerstone of Shapley's (1953) characterization of his value is the Additivity axiom. 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 $\hat{v}$. Suppose that a single-valued solution concept φ selects the payoff vector x for characteristic function v, and the payoff vector $\hat{x}$ for characteristic function $\hat{v}$. For φ to *respect Additivity*, the allocation selected for the characteristic function $v + \hat{v}$ must be the payoff vector $x + \hat{x}$. That is, $\varphi(v + \hat{v}) = \varphi(v) + \varphi(\hat{v})$. 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 particularly useful for us, and it is easy to see why Additivity implies it. Indeed, since $\varphi(2v) = 2\varphi(v)$, we have

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

3 Design of Treatments and Procedure

Our three treatments test what monetary payments individuals (henceforth called Decision Makers) deem appropriate for three Recipients, in view of how much different coalitions of Recipients would be worth. We describe the treatments below, starting with the initial treatment undertaken (Section 3.1), and then explain how the ensuing two treatments differ (Sections 3.2-3.3). In Section 3.4, we discuss theoretical motivations and implications of the characteristic functions tested. In Section 3.5, we discuss experimental procedures. In Section 3.6, we provide summary information on the subject pool per treatment.

3.1 The 'Quiz' Treatment

At the start of each session in this treatment, three subjects are chosen through uniform randomization and designated Recipients 1, 2 and 3, respectively. Recipients stay in that role for the duration of the session. All other subjects are designated Decision Makers. A session has seven rounds.

At the start of each round, each Recipient receives 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. We momentarily defer discussion of our control over the possible worths of different basket combinations, in order to describe the key role of Decision Makers.

For each round, once the content of the Recipients' baskets has been determined, Decision Makers are informed of the value of different basket combinations. 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 require monetary allocations to be efficient and nonnegative, and allow the Decision Maker to opt out of any given round without making a decision.

At minimum, all subjects receive a five-dollar show up fee. Decision Makers receive one additional dollar for each round in which they participate. 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 their show up fee). Subjects are informed only of their own payoff, and do not learn which roles other subjects played during the experiment.

The treatment was designed with the following considerations in mind. First, the coalitions' worths are "earned" by Recipients, by letting Recipients earn objects by answering quiz questions correctly.

Second, to permit specific tests of solution concepts and axioms, we nonetheless want 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, but were not given information regarding how those objects and their values would be selected. For each round, we pre-selected the available objects and values of object combinations with the following goal in mind: if Recipients earn all the objects available to them in a round, then one of the seven characteristic functions in Table 1 would be generated.[9,10] Section

	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 are described in the rows. The numerical values in the last four columns are the dollar amounts generated by combining the baskets of the Recipients listed, where Recipient i is denoted Ri.

3.4 below details the motivations for, and theoretical implications of, CF1-7.

Third, we run six different sessions of this treatment to be able to test for potential effects from the order in which the characteristic functions are presented, and if needed, help wash these out in the aggregate.[11] We use a Latin square design for characteristic functions one through six. Table 2 details the session-dependent mapping between rounds and characteristic functions. The seventh characteristic function is fully symmetric and all standard solution

[9]We explain below the session-dependent map from rounds to characteristic functions.

[10]Precisely to reduce the probability that some other characteristic functions would be generated, Recipients were afforded multiple opportunities to earn available objects. Incidentally, any superadditive characteristic function can be generated through this process.

[11]As seen in Appendix D, we do not find such order effects.

concepts prescribe an equal split. Consequently, this characteristic function is left as a consistency check in the final round of all sessions, where it cannot affect subsequent behavior.

	Round						
	1	**2**	**3**	**4**	**5**	**6**	**7**
Session 1	1	6	2	5	3	4	7
Session 2	2	1	3	6	4	5	7
Session 3	3	2	4	1	5	6	7
Session 4	4	3	5	2	6	1	7
Session 5	5	4	6	3	1	2	7
Session 6	6	5	1	4	2	3	7

TABLE 2: The ordering of characteristic functions in the six sessions. Round entries identify the characteristic function using the scheme from Table 1. The Latin square design means each possible pair from CF1-CF6 is adjacent in some session.

Fourth, Decision Makers are informed that the Recipient numbers they see on their screen in each round are randomly generated aliases for the true Recipients.[12] That is, the Recipient whose alias is Ri ($i = 1, 2, 3$) on the Decision Maker's screen in a given round is equally likely to be given the alias R1, R2 or R3 in the next round. These random aliases rule out the possibility that a Decision Maker's payoff allocation for a given Recipient is influenced by his or her choice for that Recipient in a previous round.

Finally, related to the point above, we generally tried 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. Moreover, a Decision Maker's chosen payoff allocation need not reflect strategic concerns, both because it cannot influence his or her own payoff, and because Recipients play no further strategic role. Decision Makers are presented *only* with the computed values of different basket combinations.

[12]The characteristic function the Decision Maker sees is permuted accordingly.

They do not learn which objects are in the Recipients' baskets or the values of different object combinations. Similarly, Decision Makers do not see the quiz questions Recipients faced, or how well the Recipients performed.[13] Finally, they do not learn the outcomes of other Decision Makers' choices, and cannot communicate with other subjects. This keeps our setting as close as possible to standard split-the-pie problems. The above features have the added benefit of simplifying the Decision Maker's problem from a computational standpoint.

3.2 The 'No-Quiz' Treatment

Our second treatment, comprising another six experimental sessions following the same Latin square design, differs from the Quiz treatment in only one respect. Instead of having a quiz phase where the objects are earned, the Recipients are *randomly assigned* seven baskets of objects. These baskets are identical to those that were generated in the main treatment, and we use the same monetary values of object combinations. Thus for each round, the same characteristic functions as in the Quiz treatment are generated. The only difference from the Quiz treatment is thus that Recipients play no role in generating these baskets. Recipient aliases are again permuted in each question, as in the Quiz Treatment, and the interface for inputting choices is identical. Payments to subjects are determined just as before.

3.3 The 'Vignettes' Treatment

Our third and final treatment, also comprising six experimental sessions following the same Latin square design, has all subjects in a session serving as Decision Makers. In each session, we present Decision Makers with a sequence of seven vignettes regarding hypothetical musicians, each differing only in the characteristic function it encapsulates.

Each vignette states that "Three musicians can play together as a duo or

[13]Notice in passing that keeping such background information from Decision Makers is not unrealistic outside of the lab, in the sense that one does not necessarily know precisely whether other peoples' successes are due to luck, hard work, nepotism, etc.

trio for an event (but not as soloists)." The vignette specifies the amounts the different duos and the trio would earn. The Decision Maker is then asked, "The musicians will perform as a trio for the event, and ask you to decide on their behalf how to share the $[dollar amount] earned. Which split do you choose?" In each question, the worths of the ensembles correspond to one of CF1-CF7, but with all worths scaled by ten dollars to reflect market values. Musician identifiers ($i = 1, 2, 3$) are randomly permuted in each question, analogously to the prior two treatments, since otherwise Musician 1 (3) would consistently appear strongest (weakest). The interface for inputting choices is identical to the prior treatments. Decision Makers are aware their choices will not be implemented, and their own payments are determined as before.

3.4 CF1-7: Motivations and Theoretical Implications

To ensure that subjects acting as Decision Makers are not overwhelmed by numbers, we tested only characteristic functions for which the monetary payoff of singleton coalitions is zero. We introduce CF1-CF6 to distinguish between some different solution concepts. All solution concepts agree on equal split for CF7; it is useful nonetheless to identify subjects who believe in equal splits for symmetric settings, as our analysis focuses on what these subjects will do in asymmetric ones. We introduce CF1-CF6 to distinguish between some different solution concepts. Table 3 details the payoff allocations selected in those characteristic functions.

Since the Shapley value need not belong to the core, we can test the relative prevalence of these competing norms. To make this comparison 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 (since the core is single-valued, it also coincides with the nucleolus).

We also include two characteristic functions with multi-valued cores (CF1,

	CF1	CF2	CF3	CF4	CF5	CF6
Shapley	$(30, 30, 0)$	$(\frac{80}{3}, \frac{40}{6}, \frac{40}{6})$	$(\frac{70}{3}, \frac{40}{3}, \frac{40}{3})$	$(40, 30, 20)$	$(\frac{25}{2}, \frac{25}{2}, 5)$	$(\frac{110}{3}, \frac{50}{3}, \frac{50}{3})$
Core	P_1	$\{(40, 0, 0)\}$	$\{(30, 10, 10)\}$	$\{(50, 30, 10)\}$	$\{(15, 15, 0)\}$	P_2
Nucleolus	$(30, 30, 0)$	$(40, 0, 0)$	$(30, 10, 10)$	$(50, 30, 10)$	$(15, 15, 0)$	$(40, 15, 15)$
W-CEA	$(30, 30, 0)$	-	-	$(40, 40, 10)$	$(15, 15, 0)$	$(\frac{70}{3}, \frac{70}{3}, \frac{70}{3})$
S-CEA	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 different 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 $P_3 = \{(30, 15, 15), (15, 30, 15)\}$. All the solution concepts prescribe equal split in CF7. W-CEA does not exist in CF2-CF3.

CF6). The Dummy Player axiom can be tested in CF1 (where Recipient 3 plays the dummy role). The worth of the grand coalition in CF1 is the same as in the fully symmetric CF7, since it interesting to see whether the choices in these two cases differ. The Monotonicity axiom can be tested by comparing the choices in CF2 with those in CF3 and CF6. Indeed, Monotonicity requires that the payoffs of Recipients 2 and 3 are greater in CF3 than in CF2; and that the payoffs of all three Recipients are greater in CF6 than in CF2.

We have two ways of testing Additivity, even though no two of our characteristic functions directly 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. Second, as noted earlier, Additivity is equivalent to linearity with rational coefficients, which can be tested directly using the fact that CF3 is the average of CF2 and CF7.

In each of CF1-7, every pair of Recipients can be ranked in terms of either symmetry or desirability. In particular, Recipient i is more desirable than (symmetric to) Recipient j if and only if $v(\{i, k\}) > v(\{j, k\})$ (resp., $v(\{i, k\}) = v(\{j, k\})$). Table 4 shows the ranking of Recipients in each of our seven characteristic functions. Symmetry and Desirability have implications within each characteristic function, with the exceptions of CF4 (only Desirability applies, as it is fully asymmetric) and CF7 (only Symmetry applies, as it is fully symmetric). Notice that Recipient 1 is always more desirable than, or symmetric to, Recipient 2; and in turn, Recipient 2 is always more desirable than, or symmetric to, Recipient 3. This was only for the purpose of normalization when designing the characteristic functions. As discussed in Section 3.1, Recipients' true identities (as R1, R2 or R3) are masked by a randomly generated alias in each round (with the characteristic function permuted accordingly), so Decision Makers cannot identify a pattern.

	CF 1 and 5	**CF 2, 3 and 6**	**CF4**	**CF7**
Rankings	R1$\sim$R2$\succ$R3	R1$\succ$R2$\sim$R3	R1$\succ$R2$\succ$R3	R1$\sim$R2$\sim$R3

TABLE 4: The ranking of Recipients in each of the seven characteristic functions, where Ri$\succ$Rj (Ri$\sim$Rj) means that Ri is more desirable than (symmetric to) Rj.

The imputation triangles representing our data in Figures 1, 3 and 4 include, as a reference, the payoff allocations corresponding to the different solution concepts in CF1-7. Since R1 is either symmetric to, or more desirable than R2, most solution concepts would require that R1's payoff is at least as high as that of R2. In those figures, this corresponds to a payoff allocation in the "left" half of each triangle (that is, left of the vertical line which bisects the bottom edge). Similarly, since R2 is either symmetric to, or more desirable than R3, this corresponds to a payoff allocation in the "bottom" half of each triangle (that is, below the diagonal line which bisects the right edge). Given our normalization of Recipient rankings, nearly all the solution concepts in our setting prescribe choosing an allocation in the "bottom-left" subtriangle. Although these imputation triangles represent different total monetary amounts, allocations can be compared even *across* triangles as describing the percentages allotted to different Recipients. We picked CF1-CF6 to make sure there was enough variation in the solution concepts.

3.5 Experimental Procedure

Subjects were allowed to participate in at most one session, across all treatments. The treatments thus comprise disjoint sets of subjects. Subjects were recruited via the BUSSEL (Brown University Social Science Experimental Laboratory) website.[14] The six sessions of the Quiz treatment were conducted in April and May 2013. The six sessions of the No-Quiz treatment were conducted in April, May and November of 2017. The six sessions of the Vignettes treatment were conducted in December 2018. All sessions were held at Brown University.[15] The interface for the experiment was programmed by Possible Worlds Ltd. to run through a web browser.[16]

Sessions lasted approximately thirty to forty minutes. At the start of each session, the supervisor read aloud the experimental instructions, which were simultaneously available on each subject's computer screen. The onscreen instructions for each treatment, available in Appendix E, contained a practice screen for inputting Recipients' payoffs, to get accustomed to the interface. For the Quiz and No-Quiz treatments, the session supervisor then summarized how subjects are selected into roles and how baskets values are constructed using a presentation projected onto a screen (see Appendix F). In those two treatments, subjects learned their role as Decision Maker or Recipient only after going through all of the instructions. In all three treatments, each subject

[14]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.

[15]The Quiz treatment was held at a Brown University computer laboratory, which was used by BUSSEL for economic experiments in that time period. The more recent two treatments were held at the new laboratory space designated for BUSSEL, which is comparable to the previous laboratory in size and location on campus.

[16]In the first couple of sessions of the Quiz treatment, after all but 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 then identified and corrected.

received payment in cash at the end of the session, including a $5 show-up fee.

After completing all seven rounds, 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.

3.6 Our subject pool

A total of 107 subjects (and thus 89 Decision Makers, given the three Recipients per session) participated in the Quiz treatment; 130 subjects (and 112 Decision Makers) participated in the No-Quiz treatment; and 85 subjects (and thus 85 Decision Makers) participated in the Vignettes treatment. Demographic details are provided in Appendix D.

Almost all Decision Makers chose to actively participate in each round (see Table 5 for the number of observations per characteristic function).[17] In all sessions of the Quiz treatment, recipients answered sufficiently many quiz questions per round to generate the desired characteristic functions, CF1-7. Across all treatments, only 12 out of the 286 Decision Makers opted for an unequal split in the fully symmetric CF7 (5 in the Quiz treatment, and 7 in the No-Quiz Treatment).[18] CF7, which is always the last characteristic function, serves a purpose as a screening device: our study aims to understand what individuals who believe in an equal split for symmetric settings do in asymmetric settings. As such, we drop these 12 subjects from all ensuing analysis, leaving 274 Decision Makers.

[17]In the No-Quiz treatment, all but 4 of the 112 Decision Makers opted to answer all characteristic functions; one Decision Maker answered 5 out of the 7, and 3 Decision Makers answered 6 out of the 7. In the Vignettes treatment, one of the 85 Decision Makers answered 6 out of the 7 characteristic functions, with the others answering all of them. For the Quiz treatment, two Decision Makers in the Quiz treatment chose to opt out of one characteristic function, and one chose to opt out of three.

[18]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).

4 Analysis of Population Averages

For a first look at the data, we analyze (for each treatment) the average payoff allocations to the three recipients in each characteristic functions. These are provided in Table 5, and depicted in the 'imputation triangles' of Figure 1.

It will prove helpful to look at those payoff allocations in terms of percentage departures from equal split. These numbers are provided in Table 6; underset 'fitted' values will be explained further below.[19] Percentages are easier to compare across characteristic functions, as the worth of the grand coalition may vary. The table reveals that Recipient 1 systematically gets a positive transfer compared to equal split, while Recipient 3 always suffers a net loss. Whether Recipient 2 benefits or loses against the equal-split benchmark depends on the characteristic function being tested. The analysis below culminates in a simple, theoretically founded model that can explain not only the sign of those transfers, but also, to a remarkable degree, their magnitude.

Standard theories of other-regarding preferences would overlook the worth of pairwise coalitions. If so, then average payoff allocations should also be insensitive to such worths, which is not the case.

Result 1. *Average payoff allocations vary with the worth of sub-coalitions.*

Support: If not, then percentage departures from equal split should be independent for each treatment across characteristic functions. Table 6 clearly shows that, on the contrary, there are sizable variations. For Recipient 2, for instance, those percentage departures vary from -14.3% to $+21.8\%$ in the Quiz treatment, from -16.4% to $+20.3\%$ in the No-Quiz treatment, and from -33% to $+28.1\%$ in the Vignettes treatment. From a statistical standpoint, a test of means based on Hotelling's T-squared statistic strongly rejects, for each recipient and each treatment, the joint equality (across characteristic functions) of the percentage by which the average allocation departs from equal split (with

[19] To understand how these percentages are derived, consider for instance Recipient 1's average payoff in CF1 in the Quiz treatment: \$24.30, as reported in Table 5. In that characteristic function, there are \$60 to share. Thus the equal split solution gives \$20 to each recipient. Recipient 1's percentage departure from equal split is thus equal in that case to $\frac{24.30-20}{20} = +21.5\%$ as reported in Table 6.

Quiz Treatment	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)	$10.50 (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)	$11.08 (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)
Observations	83	84	83	83	81	82	79
No-Quiz Treatment	**CF1**	**CF2**	**CF3**	**CF4**	**CF5**	**CF6**	**CF7**
Recipient 1	$22.94 (0.78)	$18.33 (0.88)	$19.64 (0.58)	$33.84 (0.97)	$10.73 (0.22)	$28.81 (1.02)	$20 (0)
Recipient 2	$24.06 (0.84)	$11.15 (0.53)	$15.05 (0.45)	$29.96 (0.87)	$11.09 (0.24)	$19.78 (0.57)	$20 (0)
Recipient 3	$13.00 (1.00)	$10.53 (0.50)	$15.31 (0.42)	$26.20 (0.92)	$8.18 (0.38)	$21.41 (0.61)	$20 (0)
Observations	105	104	104	104	104	105	104
Vignettes Treatment	**CF1**	**CF2**	**CF3**	**CF4**	**CF5**	**CF6**	**CF7**
Recipient 1	$248.46 (6.38)	$216.87 (8.84)	$222.32 (5.90)	$373.53 (8.19)	$114.89 (2.70)	$330.58 (11.92)	$200 (0)
Recipient 2	$256.11 (7.15)	$89.33 (4.87)	$141.07 (3.86)	$295.88 (5.04)	$116.18 (2.64)	$182.32 (6.26)	$200 (0)
Recipient 3	$95.44 (8.32)	$93.80 (5.47)	$136.61 (3.26)	$230.59 (7.87)	$68.93 (3.71)	$187.10 (7.04)	$200 (0)
Observations	85	85	85	85	85	84	85

TABLE 5: Average amounts allocated to Recipients per characteristic function, in each treatment (after dropping the 5/7/0 outliers in the Quiz/No-Quiz/Vignettes treatments), with standard errors in parentheses.

eight out of the nine p-values below 0.001, and a p-value of 0.019 for Recipient 1 in the No-Quiz treatment). The additional p-values for tests in this section appear in Appendix B. □

Figure 1 also depicts the standard solution concepts presented in Section 2. Clearly, none of them provide a good description of average choices. Yet some striking regularity can be found. A first, obvious feature is that average payoff allocations systematically fall very near the line passing through the equal split solution and the Shapley value. Next, notice that both average payoff allocations and the Shapley value are sometimes closer to equal split and

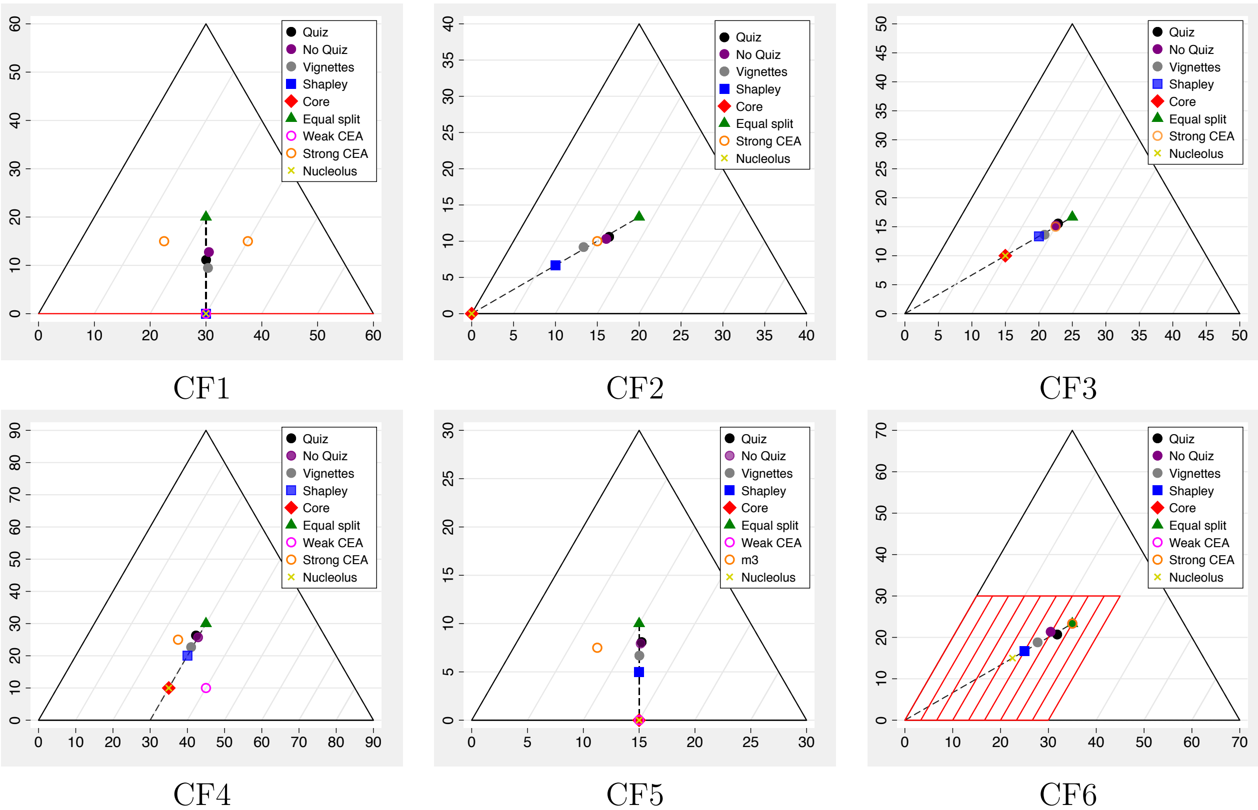


FIGURE 1: Averages choices per treatment, and solution concepts. R3's payoff is read from the vertical axis, R2's payoff is read from the diagonal isoprofit lines, and R1's payoff is what remains from the total. The top (bottom right, bottom left) corner of the simplex corresponds to giving everything to R3 (R2, R1). The dashed line connects Shapley and equal split in the region satisfying Desirability

sometimes further out (for each of the three treatments, compare for instance where the corresponding dot falls in CF2 versus CF4). Taking a closer look at the graphs, the two payoff allocations (average choices and Shapley) move inward or outward in a covariant way. In fact, as we will see, the relation is essentially linear: the distance from equal split to the average payoff allocation divided by the distance from equal split to the Shapley value is nearly constant across characteristic functions.

Quiz	CF1	CF2	CF3	CF4	CF5	CF6
Recipient 1	21.5%	32.8%	14.4%	13.4%	5.1%	18.8%
	18.4%	36.8%	14.7%	12.3%	9.2%	21.0%
Recipient 2	21.8%	−14.3%	−8.7%	−3.2%	10.9%	−7.6%
	18.4%	−18.4%	−7.4%	0%	9.2%	−10.5%
Recipient 3	−43.3%	−18.4%	−5.8%	−10.2%	−15.9%	−11.2%
	−36.8%	−18.4%	−7.4%	−12.3%	−18.4%	−10.5%
No Quiz	**CF1**	**CF2**	**CF3**	**CF4**	**CF5**	**CF6**
Recipient 1	14.7%	37.4%	17.8%	12.8%	7.3%	23.5%
	18.4%	36.9%	14.7%	12.3%	9.2%	21.1%
Recipient 2	20.3%	−16.4%	−9.7%	−0.1%	10.9%	−15.2%
	18.4%	−18.4%	−7.4%	0%	9.2%	−10.5%
Recipient 3	−35.0%	−21.0%	−8.1%	−12.7%	−18.2%	−8.2%
	−36.9%	−18.4%	−7.4%	12.3%	−18.4%	−10.5%
Vignettes	**CF1**	**CF2**	**CF3**	**CF4**	**CF5**	**CF6**
Recipient 1	24.2%	63.7%	33.4%	24.5%	14.9%	41.7%
	30.6%	61.3%	24.5%	20.4%	15.3%	35.1%
Recipient 2	28.1%	−33.0%	−15.4%	−1.4%	16.2%	−21.9%
	30.6%	−30.6%	−12.3%	0%	15.3%	−17.5%
Recipient 3	−52.3%	−29.6%	−18.0%	−23.1%	−31.1%	−19.8%
	−61.3%	−30.6%	−12.3%	−20.4%	−30.6%	−17.5%

TABLE 6: Percent departures from equal split, rounded to one decimal point. Entries show actual values, with fitted values underset.

To understand average payoff allocations, we start by checking the empirical validity of the axioms presented in Section 2.

Result 2. *Overall, there is strong evidence for Additivity, Desirability, Monotonicity, and Symmetry for average payoff allocations. On the other hand, Dummy Player is clearly rejected.*

Support: A casual look at Table 5 suggests that average payoff allocations respect Symmetry and Desirability comparisons listed in Table 4, with symmetric Recipients allocated approximately equal average payoffs, and more desirable Recipients allocated seemingly higher average payoffs. For each characteristic function and each applicable symmetry comparison Ri$\sim$Rj, the null hypothesis that the average payoffs of Ri and Rj are equal cannot be rejected by a paired t-test. Moreover, for each applicable desirability comparison Ri$\succ$Rj, the null hypothesis that the payoffs of Ri and Rj are equal is rejected by a paired t-test at all conventional levels of significance ($p \leq .001$), with the exception of a p-value of 0.0138 for the payoffs of R2 and R3 in CF4.

In the characteristic functions tested here, Monotonicity has implications when moving from CF2 to either CF3 or CF6. In the former case, Recipient 2 and 3's payoffs should increase because the worths of both the grand coalition and $\{2, 3\}$ increase. In the latter case, all three recipients' payoffs should increase because the worth of the grand coalition increases. The average payoffss in Table 5 appear to confirm these comparisons. For each applicable Recipient and each characteristic function, a paired t-test rejects the null hypothesis that the Recipient's average payoffs are the same (all p-values are 0.0000).

Recall that Additivity has two testable implications for our characteristic functions. We discuss each of these testable implications in turn. Note that the following statements pertain to all three treatments, with the understanding that numbers should be multiplied by 10 when discussing the Vignettes treatment. First, remember that CF6 can be written as the sum of CF2 and the characteristic function given by $v(\{1, 2, 3\}) = 30$ and $v(S) = 0$ for all other coalitions S. Under the relatively safe assumption that Decision Makers would allocate \$10 to each Recipient in v, Additivity can be tested by checking whether each Recipient is allocated an extra \$10 when moving from CF2 to CF6. With three treatments and three recipients per treatment, there are thus nine equations to check. Average payoff allocations in Table 5 suggest that Additivity holds. All but one paired t-test cannot reject the null hypothesis that Ri's payoff in CF6 is exactly ten dollars larger than that in CF2, for any $i = 1, 2, 3$. The only rejection concerns Recipient 2 in the No-Quiz treatment (\$11.15 in CF2 and \$19.78 in CF6, with a p-value of 0.0189).

As a second test of Additivity, notice that CF3 is the average of CF2 and CF7. Additivity implies that the solution for CF3 should be the average of solutions for CF2 and CF7. Again, there are nine equations to check, and average payoff allocations in Table 5 suggest that Decision Makers' decisions respect linearity. To confirm this, we test the null hypotheses that each Recipient's average payoff in CF3 is exactly the average of those in CF2 and CF7. The null for all but two comparisons cannot be rejected using a paired t-test.

The two cases where the hypothesis is rejected concern Recipients 1 and 3 in the Vignettes treatment (who get \$222.32 and \$136.61 in CF3, versus an average over CF2 and CF7 of \$208.44 and \$146.90, with p-values of 0.0107 and 0.0016 respectively). As we take a closer look at the data in the next section, we will identify a possible cause for this specific violation of Additivity.

Finally, CF1 is the only characteristic function among those we tested which has a dummy player (Recipient 3). It is clear at once from Table 5 that the dummy player is getting an average payoff that is strictly positive (p-value of 0.0000 in each treatment). Beyond statistical relevance, the magnitude of Recipient 3's average payoff is also noteworthy. □

The classic characterization of the Shapley value is based on the axioms of Additivity, Efficiency, Symmetry, and Dummy Player. Having to allocate all the money means Efficiency is automatically satisfied in our setting. In view of Result 2, it is natural to ask which class of solution concepts emerges if we drop the Dummy Player axiom from the above characterization. A clean, theoretical characterization emerges for the domain V of three-player characteristic functions for which the worth of each coalition is a rational number, and singleton coalitions are worth nothing. Naturally, V contains all seven characteristic functions we tested. The following proposition is proved in Appendix A.

Proposition 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 axioms, which our averaged data seems to corroborate, singles out a simple, one-parameter solution concept. Under this model, payoffs for all recipients are determined in all characteristic functions by a fixed affine combination (i.e., independent of Recipients and characteristic functions) of equal split and the Shapley value. We will call this the *ESS model*. Recipients start on equal footing, and then gain (lose) δ dollars for each dollar by which the

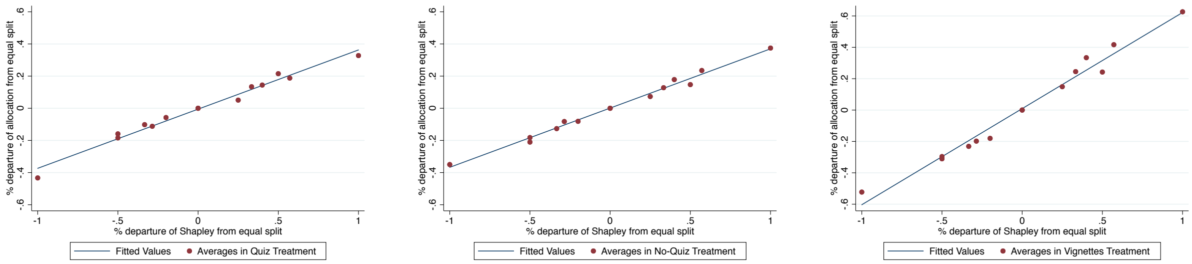


FIGURE 2: The ESS model fitted to the average payoffs in the three treatments.

Shapley value is larger (smaller) than equal split: $\delta = \frac{\sigma^i(v)-ES^i(v)}{Sh^i(v)-ES^i(v)}$, for *each* characteristic function v and *each* Recipient i such that $Sh^i(v) \neq ES^i(v)$. Equivalently, the percentage departure of the solution from equal split coincides with δ times the percentage departure of the Shapley value from equal split for all recipients in all characteristic functions:

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

for each Recipient i and each characteristic function 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 in Proposition 1 relies on the axioms being satisfied universally by the average choices, which is not directly testable. For this purpose, we fit the model to the data and see that it provides a close match. The next result also provides an empirical estimate of the parameter δ in each of the three treatments.

Result 3. *The ESS model successfully captures average payoff allocations. The estimated weight on the Shapley value is* 0.368 *for the Quiz treatment,* 0.369 *for the No-Quiz treatment, and* 0.613 *for the Vignettes treatment.*

Support: Accounting for noisy departures, the ESS model lends itself to at least two possible methods of estimation through linear regression. In one possibility, the average amount $\bar{m}^i(v) - ES^i(v)$ a Recipient receives net of equal split is proportional, by δ, to the amount $Sh^i(v) - ES^i(v)$ that the Shapley

	Quiz Treatment	No-Quiz Treatment	Vignettes Treatment
	$\frac{\bar{m}-\text{ES}}{ES}$	$\frac{\bar{m}-\text{ES}}{ES}$	$\frac{\bar{m}-\text{ES}}{ES}$
$\frac{\text{Sh}-\text{ES}}{ES}$	0.368^{***}	0.369^{***}	0.613^{***}
	(0.021)	(0.010)	(0.029)
constant	-0.005	0.001	0.009
	(0.007)	(0.005)	(0.013)
num. obs.	14	14	14
R^2	0.9799	0.9901	0.9786

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

TABLE 7: Per treatment, regression of the average percentage departure from equal split of the allocations for R1 and R3, against percentage departure of the Shapley value from equal split. Robust standard errors in parentheses. Estimates rounded to 3 decimal places.

value offers him, net of equal split. This would lead to a regression in levels, potentially allowing characteristic functions with larger monetary amounts to unduly influence our estimates. To account for this, we follow a natural alternative instead, which is to divide both sides of the solution in Proposition 1 by the equal split solution. Note that this does not change the solution concept: like all the other solution concepts, it is invariant to normalization. The resulting interpretation of variables is then in terms of percent departure from equal split, just as in Table 6.[20] That is, under this model, the percentage by which a Recipient's allocation departs from equal split is proportional, by δ, to the percentage by which the Shapley value departs from equal split. Because the ESS model has the same δ applying to each Recipient, we can pool the data across Recipients to estimate δ. However, since the payoff of a Recipient can be inferred from the payoffs of the other two Recipients (the sum of all three payoffs is fixed per characteristic function). Thus we only take into consideration the average choices for Recipients 1 and 3 in each characteristic function (we drop Recipient 2 since the Shapley value for R2 coincides with equal split in CF4, leading to less variation that we can exploit).

Table 7 shows the regression results for each treatment, using robust standard errors. As can be seen there, the estimate of δ, which is the coefficient

[20] This normalization is mathematically equivalent to normalizing by the total amount $v(\{1, 2, 3\})$ available, which would instead be interpreted as the departure from equal split as a percentage of the total pie.

of the variable $(Sh - ES)/ES$, is highly significant in each treatment, and the *smallest* R^2 value across treatments is 0.9786. There is indeed a striking linear relationship apparent in Figure 2, where we plot for each treatment the averaged data together with the fitted regression line. Table 6 specifies these estimated values beneath the observed ones, making it easy to quantify the close fit. □

Results 1-4 show that our qualitative findings replicate to a large extent across the three treatments. As noted in the Introduction, an implication is that sub-coalition worths matter for average payoff allocations even when there is no sense conveyed that they were earned (as in the No-Quiz treatment). As for quantitative comparisons, the average payoffs in the Quiz and No-Quiz treatments are strikingly similar, and nearly overlay each other in the imputation triangles of Figure 1. By contrast, average payoffs in the Vignettes treatment, where Recipient values are generated in the context of professional service, suggest that marginal contributions there are weighted more heavily than in the prior treatments. The next result confirms these statements.

Result 4. *Average payoff allocations are not significantly different in the Quiz and No-Quiz treatments. By contrast, average payoff allocations are significantly different, moving further away from Equal Split, in the Vignettes treatment than in both the Quiz and No-Quiz treatments.*

Support: Going back to the estimations in Table 2, we cannot reject the null hypothesis that the estimated weights $\delta = 0.368$ and $\delta = 0.369$ for the Quiz and No-Quiz treatments, respectively, are the same (p-value 0.9878). However, the null hypothesis that the estimated $\delta = 0.613$ for the Vignettes treatment is the same as for the Quiz Treatment, or the No-Quiz treatment, is rejected at all conventional significance levels (both p-values are 0.0000). □

Average payoffs provide an elucidating, birds-eye view of the data, which is the primary purpose of this section. Beyond this, the simple average is a quite natural means of aggregating opinions,[21] and in this case average payoff

[21] Applying Rubinstein and Fishburn (1986)'s result to our setting, it is the only aggregator

allocations could also reflect a societal view of how to split the money available. Of course, taking averages is purely an ex-post exercise: recipients were not paid according to such averages, nor were they mentioned to subjects.

5 A Closer Look at the Data

Average payoffs have the advantage of partially canceling noise, which makes the analysis more transparent; but they miss additional information and nuance found in the data. In this section, we revisit Results 1-4 at a deeper level, by examining Decision Makers' choices from different perspectives. How many Decision Makers overlook the worths of sub-coalitions, and actually choose equal splits? Are axioms like Additivity and Symmetry valid on average due to the presence of some equal splitters, or do they still hold among non-equal splits? Do the δ's identified in Result 3 reflect a homogenous population or the average of a heterogenous one? Could it be that other models are useful for explaining some individuals' choices, but are not prevalent enough to survive (or cancel each other out) when averaging? How do the distributions of choices compare across treatments?

5.1 Description of the Data

By depicting a Decision Maker's allocation for the three Recipients in imputation triangles (as standard in the cooperative games literature), Figures 3-4 provide a visualization of all Decision Makers' choices for each characteristic function in each treatment.[22] Within each simplex, a ball's radius is proportional to the fraction of Decision Makers who picked its center.

Table 8 shows the percentage of equal splits in CF1-CF6. Observe that in CF2, CF3 and CF6, the worth of the grand coalition is not divisible by

that picks the common opinion when all Decision Makers agree, that is efficient, and for which a Recipient's payoff depends only on the amounts Decision Makers' allocated to him. For example, the aggregation method giving each Recipient the median payoff chosen for him would satisfy the first and last properties, but violate the second.

[22] We include in these scatterplots the 5 subjects in the Quiz treatment and the 7 subjects in the No-Quiz treatment who select an unequal split in CF7.

	CF1	CF2	CF3	CF4	CF5	CF6	CF7
Quiz	40.96%	23.81%	18.07%	57.83%	65.43%	20.73%	100%
No Quiz	51.43%	38.46%	35.58%	50.96%	60.58%	39.05%	100%
Vignettes	25.88%	12.94%	11.76%	31.76%	35.29%	20.24%	100%

Table 8: Percent of payoff allocations that are 'equal splits'.

three. Decision Makers can input numbers with decimal places, but may find payments in whole dollars to be simpler. Throughout the paper, we will thus count a Decision Maker's chosen allocation in the Quiz and No-Quiz treatments as an *equal split* if payoffs across Recipients differ by at most one dollar. For the Vignettes treatment, where all values are scaled by ten, we count an allocation as an equal split if payoffs across Recipients differ by at most ten dollars.

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. Reinforcing Result 1, 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.

5.2 Testing Axioms

Result 2 established that average payoff allocations satisfy all the axioms listed in Section 2.2, except Dummy Player. We now take a closer look at the data, to gauge the extent to which individual choices satisfy them.

A good understanding of individual choices is more informative than an understanding of average choices. However, since averaging may help cancel noise, individual choices can be more difficult to understand. Satisfying Additivity, for instance, requires verifying a knife-edge equality. Even if all Decision Makers abide by the axiom, most will appear to fail it individually, if noise is added to their choices.

Thus we also explore axioms at an intermediate level of aggregation, by performing tests on the distributions of money allocated to recipients. Certain axioms are trivially satisfied for Decision Makers who split equally in the relevant characteristic functions. In those cases, we focus on testing the axiom among the subpopulation that chose otherwise. Additional figures and p-values for tests in this section appear in Appendix C.

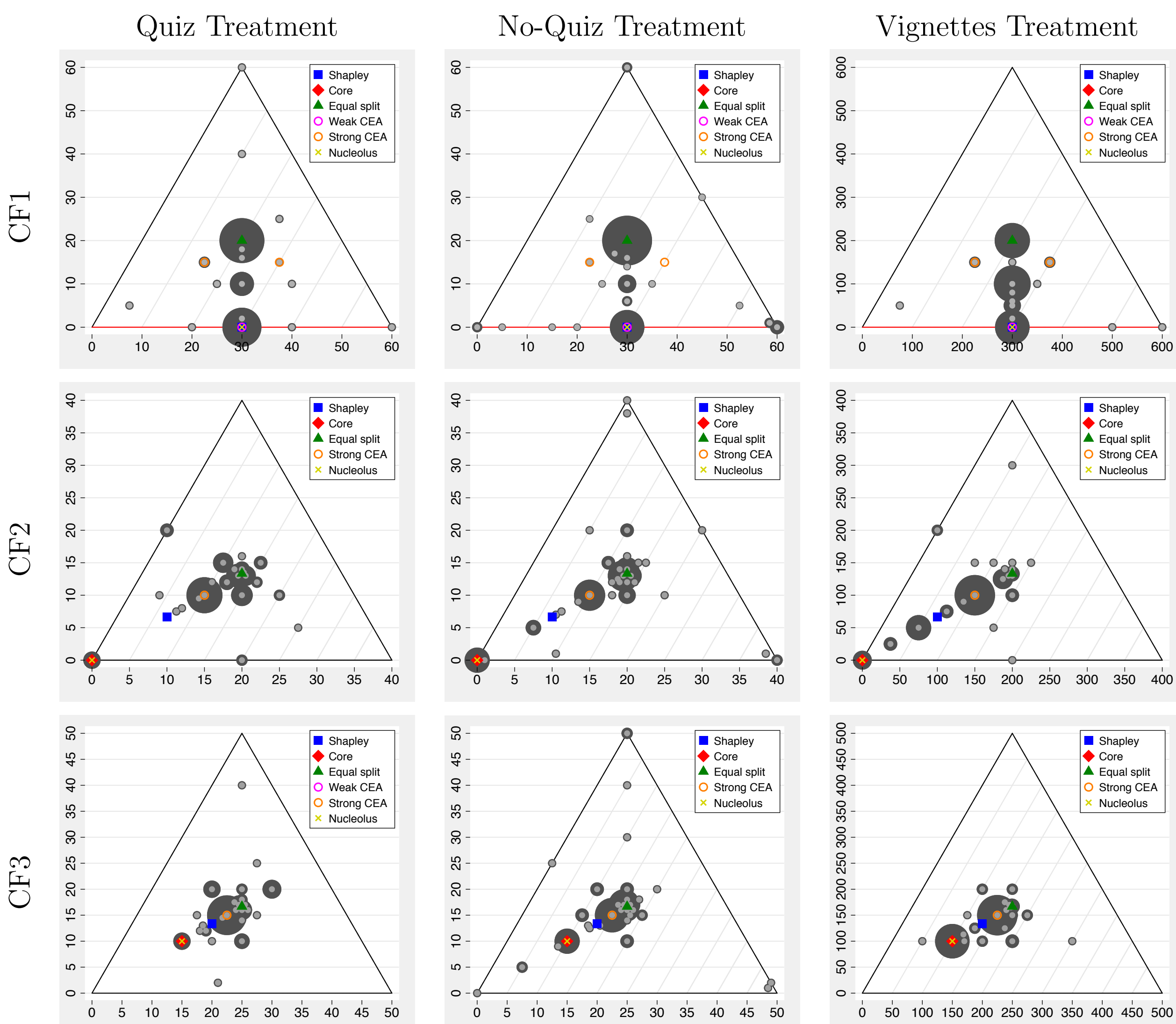


FIGURE 3: Frequency-weighted scatterplots of all choices in CF1-CF3.

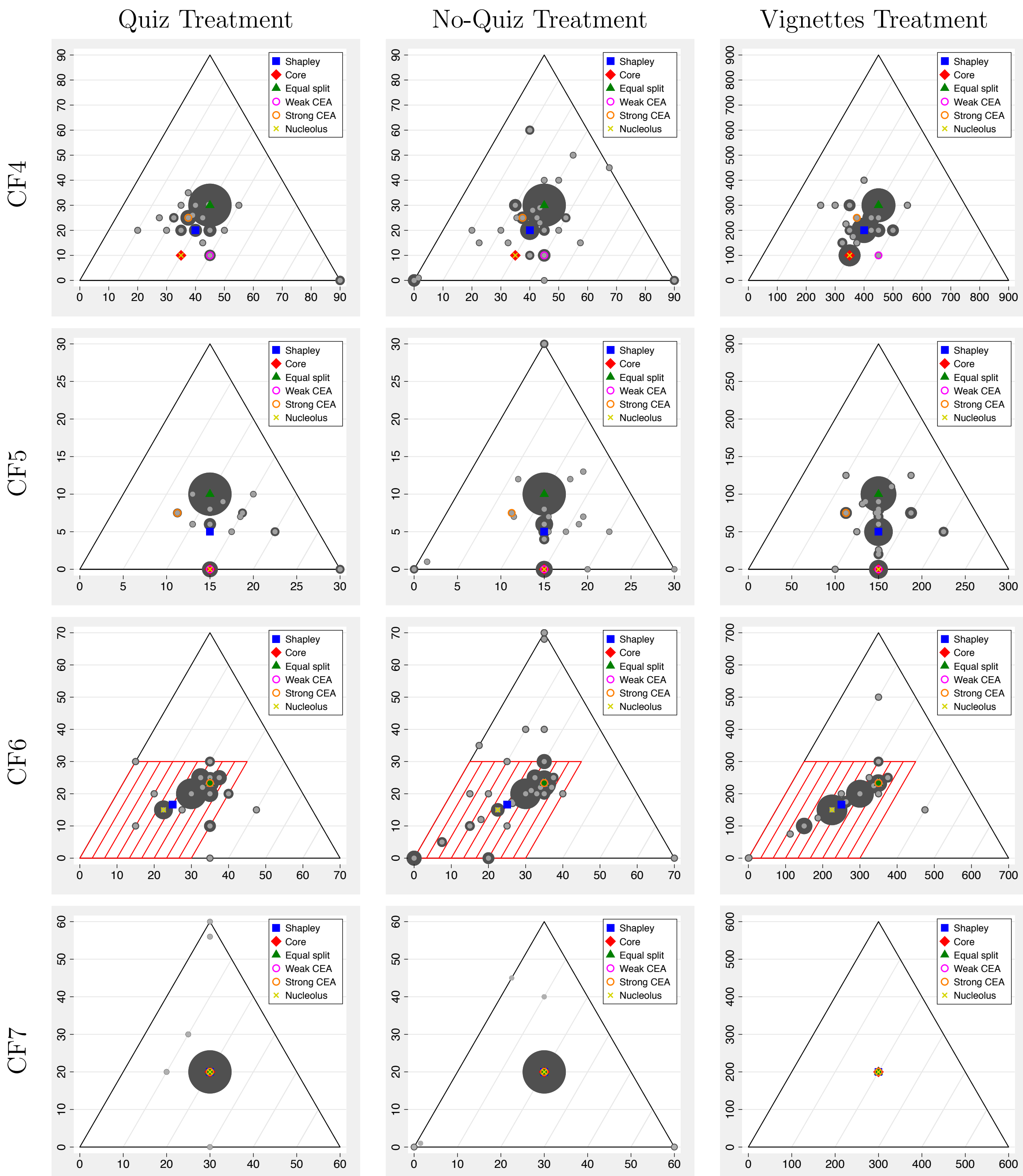


FIGURE 4: Frequency-weighted scatterplots of choices in CF4-CF7. There is no frequency-weighting for CF7 of the Vignettes treatment, as all split equally.

Dummy player. CF1 is the only characteristic function tested which has a dummy player (Recipient 3). A substantial fraction of subjects satisfy the Dummy Player (34.9%/29.5%/27.1%),[23] and a substantial fraction violate it, either by choosing an equal split (41%/51.4%/25.9%) or by picking an allocation that is a strict convex combination of the equal split solution and the Shapley value (15.7%/12.4%/41.2%). Many in the last category gave $10 to Recipient 3 (or $100 in the Vignettes treatment). There are many reasons why one may see few norms here; for instance, the Shapley value is an element of the core, and coincides with the nucleolus. A more complex picture arises in other characteristic functions.

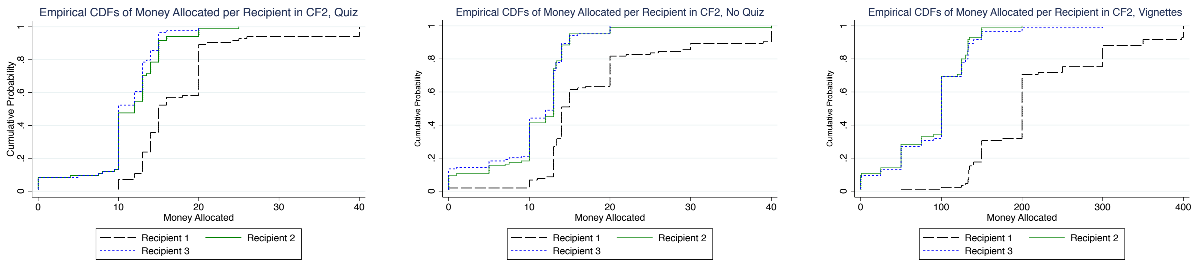
Symmetry and Desirability. By its nature, the equal split solution respects Symmetry but violates Desirability. As we will show, Symmetry appears to be respected *even among non-equal splits*, and as we have already established, Desirability appears to be respected *even when including them.*

We quickly revisit our analysis of average payoffs, among non-equal splits. For Symmetry, the null hypothesis that average payoffs of Ri and Rj are equal cannot be rejected by a paired t-test for any treatment or applicable symmetry comparison, with only one exception: in CF6 of the No-Quiz Treatment, the p-value for symmetry of average payoffs of R2 and R3 is 0.0091. For Desirability, our earlier conclusion that it is generally satisfied was reached *even when including* equal splits. The evidence is yet stronger when dropping them. For each applicable desirability comparison, the null hypothesis that the average payoffs are equal is rejected by a paired t-test at all conventional levels of significance ($p \leq .001$), with the exception of three comparisons involving CF4, which are rejected only at the 5% significance level.[24]

A panoramic view of Symmetry and Desirability at the aggregate level is provided by the empirical CDFs of money allocated per Recipient in each characteristic function. Theoretically, average payoffs of symmetric agents could be near each other, even while the CDFs differ widely. For each treatment, Fig-

[23] The three figures refer to percentages in the Quiz/No-Quiz/Vignettes treatments.

[24] The p-values are 0.0167 for R1$\succ$R2 in No-Quiz, 0.0109 for R2$\succ$R3 in Quiz, and 0.0138 for R2$\succ$R3 in No-Quiz.



Quiz treatment | No-Quiz treatment | Vignettes treatment

FIGURE 5: Symmetry and Desirability. Empirical CDFs per Recipient in CF2.

ure 5 superimposes the empirical CDF's of money allocated to the recipients in CF2; see Figure 12 in the Appendix for the other characteristic functions. These figures paint a suggestive picture of Symmetry and Desirability.

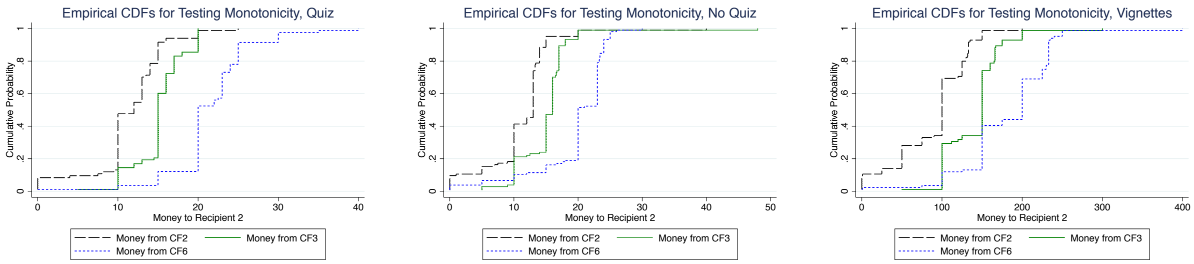
Indeed, *even among non-equal splits*, the Wilcoxon[25] signed-ranks test cannot reject any of the null hypotheses that the money allocated to symmetric Recipients come from the same distribution (although the p-value of 0.0540 for $R2 \sim R3$ in CF6 of the No-Quiz Treatment is marginal). Moreover, *even when including equal splits*, the Wilcoxon signed-ranks test rejects all the null hypotheses that the money allocated to two Recipients ranked by desirability in a characteristic function come from the same distribution. These null hypotheses are rejected at all conventional levels of significance ($p \leq .001$), with the exception of the desirability comparison R2$\succ$R3 in CF4, which is rejected at a 5% significance level in the Quiz treatment (p-value 0.0387) and at the 1% level in the No-Quiz treatment (p-value 0.0029).

[25]Note that the Kolmogorov-Smirnov test, which we will use later for comparing distributions *across* treatments, is not the most appropriate test for the current null hypotheses, which are *within* treatments: unlike the Wilcoxon test, it does not take into account that these are matched samples (i.e., from the same Decision Makers). However, it may be worth noting that the Kolmogorov-Smirnov test makes nearly identical conclusions here.

Quiz	CF1	CF2	CF3	CF4	CF5	CF6
Symmetry	81.6%	57.8 %	67.6%	n/a	57.1%	56.9%
Desirability	85.7%	56.3%	63.2%	31.4%	67.9%	55.4%
Both	79.6%	53.1%	58.8%	31.4%	57.1%	50.8%
No-Quiz	**CF1**	**CF2**	**CF3**	**CF4**	**CF5**	**CF6**
Symmetry	74.5%	64.1 %	68.7%	n/a	78.0%	62.5%
Desirability	84.3%	62.5%	71.6%	35.3%	85.4%	65.6%
Both	76.5%	59.4%	65.7%	35.3%	75.6%	59.4%
Vignettes	**CF1**	**CF2**	**CF3**	**CF4**	**CF5**	**CF6**
Symmetry	87.3%	83.8%	74.7%	n/a	72.7%	89.6%
Desirability	90.5%	85.1%	82.7%	96.6%	78.2%	88.1%
Both	87.3%	82.4%	74.7%	96.6%	63.6%	86.6%

Table 9: Symmetry and Desirability in each treatments. For each CF, the percentage of chosen allocations (among non-equal splits) respecting these axioms.

Finally, we examine Symmetry and Desirability at the individual level. Decision Makers opting for equal splits clearly respect all symmetry comparisons, but violate all desirability comparisons. Among Decision Makers who split unequally in a given characteristic function, Table 9 shows that a substantial portion respect all applicable symmetry and desirability comparisons. One should keep in mind that it is nontrivial to assess symmetry and desirability in each characteristic function; and due to our randomly generated aliases for recipients, Decision Makers cannot detect or rely on any patterns. The table allows for differences of at most one dollar in payoffs in assessing symmetry (ten for Vignettes). Note that in CF4, no two players are symmetric. As also seen at the aggregate level, this feature may have complicated the problem in Quiz and No-Quiz (though, interestingly, not in Vignettes, which might suggest that the presence of a relatable story in the vignette did help subjects think through the problem at hand), adding some noise. However, 94.3%/84.3%/96.6% of subjects respect at least two out of the three rankings.



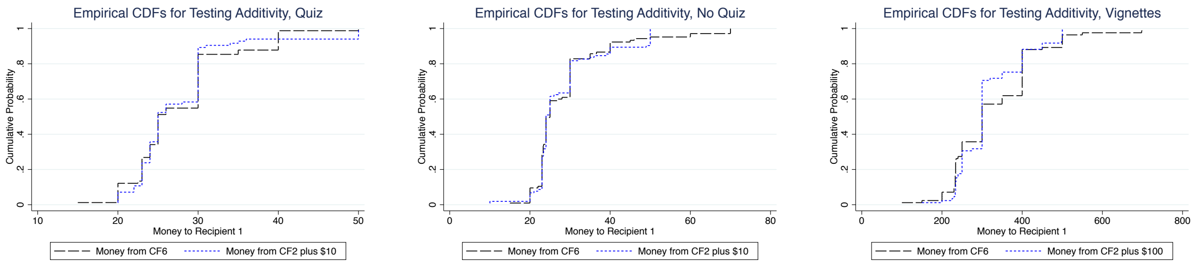
Quiz treatment　　No-Quiz treatment　　Vignettes treatment

FIGURE 6: Monotonicity. Empirical CDFs for Recipient 2 in CF2, CF3 and CF6.

Monotonicity Among those characteristic functions tested here, Monotonicity has implications only when moving from CF2 to either CF3 or CF6. In the former case, Recipient 2 and 3's payoffs should increase because the worths of both the grand coalition and $\{2, 3\}$ increase. In the latter case, all Recipients' payoffs should increase because the worth of the grand coalition increases.

Figure 6 superimposes, for each treatment, the relevant empirical CDFs for Recipient 2. As seen there, his payoff in both CF3 and CF6 first-order stochastically dominates (or nearly does) his payoff from CF2. The analogous graphs for the other recipients, shown in Figure 13, similarly corroborate Monotonicity. Because the value of the grand coalition happens to also increase relative to CF2, the equal split solution also satisfies these instances of Monotonicity. Even among those who split unequally in at least one of the applicable characteristic functions, t-tests of the average payoffs of a recipient, and Wilcoxon signed-ranks tests of the distributions of payoffs to a recipient, all soundly reject the null hypotheses of equality when moving from CF2 to CF3/CF6 (all the p-values are 0.0000, for all recipients and treatments).

At the individual level, any Decision Maker splitting equally in both characteristic functions would be consistent with these particular instances of Monotonicity. Among all *other* Decision Makers, 75.8%/68.6%/74.3% allocated strictly more money to all three Recipients when moving from CF2 to CF6, and 57.4%/66.7%/71.1% allocated strictly more money to Recipients 2 and 3 when moving from CF2 to CF3.



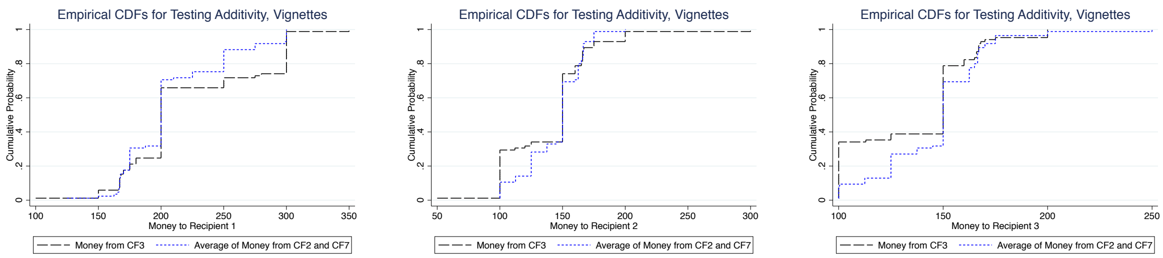
Quiz treatment   No-Quiz treatment   Vignettes treatment

FIGURE 7: 1^{st} instance of Additivity. The relevant empirical CDFs for Recipient 1.

Additivity Recall that Additivity has two testable implications for our characteristic functions. We discuss each of these testable implications in turn.

First, remember CF6 can be written as the sum of CF2 and the characteristic function given by $v(\{1,2,3\}) = 30$ and $v(S) = 0$ for all other coalitions S (all amounts should be multiplied by 10 for the Vignettes treatment). Under the relatively safe assumption that Decision Makers would allocate \$10 to each Recipient in v, Additivity can be tested by checking whether each recipient is allocated an extra \$10 when moving from CF2 to CF6. Additivity was first discussed in Result 2, but when including equal splits. The averages remain strongly suggestive of Additivity even when dropping those Decision Makers who split equally in both CF2 and CF6: paired t-tests cannot reject the null hypothesis that Ri's payoff in CF6 is exactly ten dollars larger than that in CF2, for any $i = 1, 2, 3$, with the only exception of R2 in No-Quiz (the payoff difference is statistically significant with a p-value of 0.0180, though amounts to a discrepancy of only \$1.37).

For a broader picture, we consider again the empirical CDFs. As seen in Figure 7, after translating the CDF of money allocated to R1 in CF2 by \$10, the resulting CDF is close to the empirical CDF of money allocated to R1 in CF6. Corresponding graphs for R2 and R3 appear in Figure 14 of the Appendix, and overall suggest a strong support for Additivity. Even among only those Decision Makers who choose an unequal split in at least one of CF2 or CF6, the Wilcoxon signed-rank test cannot reject the null hypothesis that the data in each case comes from the same distribution, with the only



Recipient 1 Recipient 2 Recipient 3

FIGURE 8: 2^{nd} instance of Additivity. Empirical CDFs in Vignettes treatment.

exception of R2 in the No-Quiz treatment (p-value 0.0325).

At the individual level, even among those choosing an *unequal split* in at least one of CF2 or CF6, we find 17/14/24 Decision Makers satisfy this instance of Additivity with exact equality for all three Recipients.

As a second test of Additivity, remember that CF3 is the average of CF2 and CF7. Additivity implies that the solution for CF3 should be the average of solutions for CF2 and CF7. For the Quiz and No-Quiz treatments, none of the null hypotheses corresponding to this second instance of Additivity can be rejected, whether we examine equality of average allocations among non equal-splitters using paired t-tests, or equality of underlying distributions using Wilcoxon signed-rank tests. For the Vignettes Treatment, however, these null hypotheses are rejected for both R1 and R3 (which reflects the fact that Additivity was not satisfied for average payoffs in that case either, see the analysis in support of Result 2). The empirical CDFs in Figure 8 show the regions where these (translated) empirical CDFs in the Vignettes treatment depart, and shed light on a likely cause. Namely, in CF3, there are 21 Decision Makers who allocate (\$300, \$100, \$100), which is precisely the singleton core in that characteristic function, and thus also the nucleolus. To satisfy Additivity, those same Decision Makers would need to allocate \$0 to both Recipients 2 and 3 in CF2, since they allocate \$200 to all Recipients in CF7. Interestingly, all these subjects do treat Recipients 2 and 3 symmetrically in CF2, but only 5 subjects out of these 21 give them \$0.[26] Repeating the paired t-tests and

[26] Among the other 16 subjects, 1 subject gives them \$1; 1 subject gives them \$25; 6

Wilcoxon signed-rank tests after dropping the above 21 (even dropping the 5 among them who do satisfy Additivity), we find that *Additivity is satisfied by the remaining 55 subjects*: none of the null hypotheses corresponding to this second instance of Additivity can be rejected.

At the individual level, even among those choosing an *unequal split* in at least one of CF2 or CF3, we find 7/15/19 Decision Makers satisfy this instance of Additivity with exact equality for all three Recipients.

5.3 ESS and Other Models

We now turn our attention to Result 3. Recall that average payoff allocations are successfully explained by the ESS model with $\delta = 0.368/0.369/0.613$ for the Quiz/No Quiz/Vignettes treatment. What does this mean at the individual level? Does the ESS model provide a better fit to individual choices compared to alternatives like the nucleolus or the strong-CEA solution? Moreover, should the ESS model prove dominant, is the δ identified for each treatment in Result 3 a reflection of mostly homogenous opinions, or is it the average of heterogenous ideals?

Section 5.1 revealed that for each characteristic function, there is a fraction of Decision Makers who select an equal split. As we now consider choices across characteristic functions, we will define a Decision Maker as an *equal splitter* if she picks equal splits in *all* characteristic functions. There are 10/22/6 such Decision Makers in the Quiz/No-Quiz/Vignettes treatment. By definition, the δ of these individuals under the ESS model will be close to zero, providing a first sense of heterogeneity in the population.

We can single out a larger class of Decision Makers. Say that a Decision Maker is a *D-equal splitter* if she splits the money *exactly* equally in *all four* characteristic functions where the total worth is divisible by three. Through their choices, D-equal splitters reveal themselves as having a strong tendency

subjects give them \$50; 2 subjects give them \$75; and 6 subjects give them \$100. This can either be noise or a disinclination to give Recipient 1 all the money in CF2, which is theoretically different than violating the Dummy Player axiom. In fact, half of these 16 subjects do give \$0 to the dummy player in CF1.

towards equal splits. In our data, equal splitters are a subset of D-equal splitters (they do divide the money exactly equally, not just within \$1, in CF1, CF4, CF5 and CF7), but there are many more D-equal splitters (28/42/15 in Quiz/No-Quiz/Vignettes). Looking at the data, some of these subjects seem to round payments by multiples of \$5 instead of \$1 when the worth of the grand coalition is not divisible by three.[27] Those Decision Makers, too, would be well-captured by the ESS model with δ close to zero. Interestingly, a few other D-equal splitters seem to follow a more intricate model of choice: they sometimes select reasonable payoff allocations that are far from equal splits when the worth of the grand coalition is not divisible by three.[28] The ESS model is not a good description of such choices. We find it interesting to document these behaviors, though they are rather unusual and have limited impact on our analysis.

As D-equal splitters comprise a sizable fraction of subjects and most are well described by the ESS model with δ's near zero, it must be that other Decision Makers use much larger δ's to obtain the average δ uncovered in Result 3. Addressing this question, Table 10 provides the results of linear regressions of the same normalized form as in Result 3. Unlike Result 3, the choices are not averaged. This allows us to keep track of each Decision Maker's identity, and 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 regressions in the leftmost column consider all Decision

[27] It would be inadequate, though, to redefine the notion of equal split in a given characteristic function by allowing differences of up to \$5. For instance, 53 subjects in the Quiz treatment would pass the \$5 test in CF3, but only 23 of them are D-equal splitters. Instead, the large majority (73.7%) of those who are within \$5 but not within \$1 select the allocation $(\$20, \$15, \$15)$, which is consistent with rewarding R1, who is most desirable, while giving equal payoffs to the symmetric R2 and R3.

[28] One such D-equal splitter in the Quiz treatment 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. Another interesting D-equal splitter, this time in the Vignettes treatment, chooses exactly the allocations prescribed by the nucleolus in CF3 and CF6, and exactly that prescribed by the strong-CEA in CF2; hence they split equally when the total is divisible by three, and otherwise choose allocations with differences ranging from \$100 to \$250.

Quiz Treatment	(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^{***} (0.039)	0.521^{***} (0.044)	0.065^{*} (0.028)
constant	-0.005 (0.005)	-0.012 (0.007)	0.008^{*} (0.003)
No. subjects	84	56	28
Observations	1150	766	384
R^2	0.3003	0.4297	0.0450
No-Quiz Treatment	(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^{***} (0.047)	0.595^{***} (0.064)	0.027^{*} (0.011)
constant	0.001 (0.006)	-0.002 (0.009)	0.005^{**} (0.001)
No. subjects	105	63	42
Observations	1460	874	586
R^2	0.2085	0.3305	0.0384
Vignettes Treatment	(all) $\frac{m-\text{ES}}{ES}$	(no D-equal splitters) $\frac{m-\text{ES}}{ES}$	(D-equal splitters) $\frac{m-\text{ES}}{ES}$
$\frac{\text{Sh}-\text{ES}}{ES}$	0.613^{***} (0.049)	0.733^{***} (0.048)	0.054 (0.036)
constant	0.009^{*} (0.004)	0.008 (0.005)	0.014 (0.007)
No. subjects	85	70	15
Observations	1188	978	210
R^2	0.4774	0.5752	0.0285

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

TABLE 10: Regressions of the percentage departure from equal split of allocations to R1 and R3, against percentage departure of the Shapley value from equal split. Robust standard errors in parentheses, clustered by Decision Maker. Estimates rounded to 3 decimal places.

Makers. We get about the same estimates for δ in each treatment as we did in Result 3, but naturally with lower R^2's as the noise from the variety of choices hasn't been canceled out by averaging. The middle column estimates the same regression model, but among those who are not D-equal splitters. The estimates of δ significantly increase for each treatment (p-values all less than 0.0001). The rightmost column estimates the model among D-equal splitters, who are captured by a very small, positive δ. We will discuss comparisons of δ's across treatments in Section 5.4.

In Appendix D 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 economics education than personal traits, but cannot draw any definitive conclusions using the sparse education data we collected. We find no session effects.

Next, we delve more deeply into Decision-Maker behavior, considering individual-level regressions to estimate each Decision Maker's δ under the ESS model. For these individual regressions only, we focus on the 79/101/84 subjects in the Quiz/No-Quiz/Vignettes treatments who made choices for all seven characteristic functions, so that each regression has 14 observations. As we find a few individuals with $\delta > 1$ (which violates individual rationality in CF1) or with $\delta < 0$ (which violates Desirability and Monotonicity), we perform the regression while constraining $\delta \in [0, 1]$. While δ's outside this range could be attributed to noise, they may simply be the result of model misspecification. We precisely intend to account for the possibility that individuals apply different models. To do this, we consider 'three' possible solution concepts for each subject: the nucleolus, the strong solution, and their 'best' linear combination of the Shapley value and equal split solution, as estimated from the normalized, individual-level regression just discussed. We classify a Decision Maker according to which of these solution concepts minimizes the sum (over CF1-CF6) of the squared error between their chosen *normalized* allocation for the three recipients and the predicted normalized allocation.[29]

By its nature, the linear model estimated nests both the equal split solution and the Shapley value, as well as combinations thereof. We consider the single-

[29] Again, we normalize all allocation by the total amount available to prevent conclusions being unduly influenced by characteristic functions with large amounts available. However, the results are nearly identical to those without normalization.

valued nucleolus rather than the core, which is very permissive in some of our characteristic functions. Since the strong-CEA selects two possible allocations for CF1, we consider the minimum error among these two. We do not consider the weak solution in this analysis, as it is undefined for CF2-CF3. Table 11 displays the classification results.

	ESS	**Nucleolus**	**Strong-CEA**
Quiz	93.7%	3.8%	2.5%
No-Quiz	92.1%	5.9%	2.0%
Vignettes	88.1%	8.3%	3.6%

Table 11: Percentage of subjects each model best explains.

Some subjects in each treatment are indeed best explained by the nucleolus or strong-CEA. One might think the continuum of possible values for δ in the ESS model stacks the deck against these other models; however, each subject would be classified under exactly the same solution concept (ESS, nucleolus or strong-CEA) if instead of estimating δ in the ESS model through regression, we were to allow only $\delta \in \{0, 1/3, 2/3, 1\}$, that is, *only two intermediate values of* δ along with the equal split solution and Shapley value. In that simple setting, about 38%/51%/20% of subjects are classified as equal splitters ($\delta = 0$) in the Quiz/No-Quiz/Vignettes treatments, with 56%/42%/67% classified as some $\delta \in \{1/3, 2/3, 1\}$.[30] Consistent with our earlier discussion of D-equal splitters, those individuals are almost entirely classified under the equal split solution, with just a couple of exceptions per treatment; and they constitute 67%/76%/76% of those assigned to that category.

A benefit of estimating δ through individual regressions is confidence intervals. Figure 9 graphs the individual δ's estimated through regression, along with their 95% confidence interval using robust standard errors when the $\delta \in [0, 1]$ constraint is slack. Each individual regression, of course, has only 14 observations per Decision Maker, which are likely to be noisy; hence some

[30]In particular: 27%/16%/21% are classified as $\delta = 1/3$, 24%/15%/26% as $\delta = 2/3$, and 5%/11%/20% as Shapley, for Quiz/No-Quiz/Vignettes, respectively.

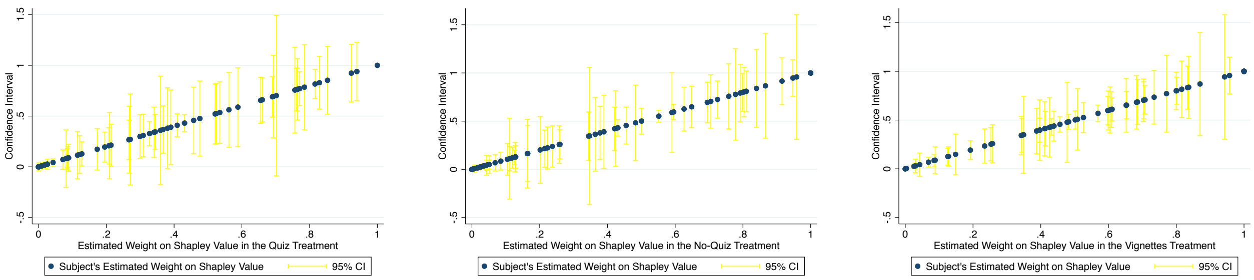


FIGURE 9: Estimated weight on Shapley and its 95% robust CI for each individual best explained by the ESS model, per treatment.

confidence intervals are quite large. The distribution of δ's in the population, seen in Figure 10, may provide a more informative picture of heterogeneity and comparisons across treatments, which we discuss next.

5.4 Comparing Treatments

Expanding upon Result 4, our estimates of the ESS model in each treatment suggest that, at least at the aggregate level, the estimated weight on the Shapley value is similar across the Quiz and No-Quiz treatment, and significantly smaller than in the Vignettes treatment. This is confirmed statistically. Among the population of non D-equal splitters, these differences are less pronounced: while the estimates for the Quiz and Vignettes treatments are statistically significant, the estimate for the No-Quiz treatment (which falls in better) is only marginally different from that of Vignettes (p-value of 0.0824).

An interesting explanation is suggested by the empirical CDF's of estimated individual δ's in Figure 10. The distributed of estimated δ's in the Vignettes treatment FOSD the distributions of both the Quiz and No-Quiz treatments, and the Kolmogorov-Smirnov test indeed rejects that they are the same (p-values 0.003 and 0.000, respectively). The Kolmogorov-Smirnov test cannot reject that the distributions from the Quiz and No-Quiz treatments are the same (p-value 0.307). In fact, the distribution of δ's from No-Quiz treatment has some features of a mean-preserving spread of the distribution of δ's from the Quiz treatment, placing greater weight on extreme values.

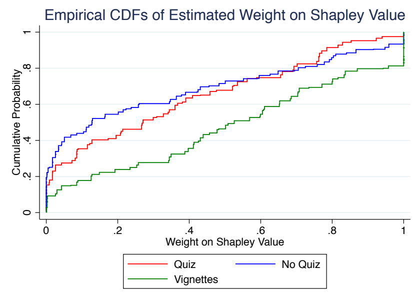
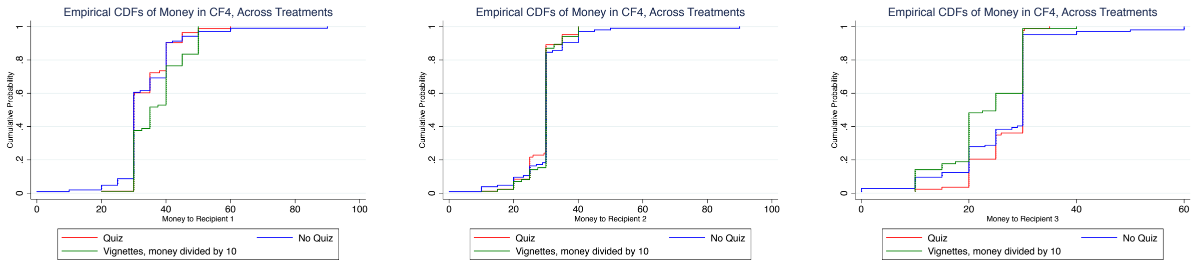


FIGURE 10: CDFs of estimated weights on Shapley, per treatment.

More basically, we may directly compare the distribution of money allocated across treatments, for each characteristic function and each Recipient. With 3 Recipients and 6 characteristic functions where choices vary, there are 18 possible CDFs. For comparison purposes, we divide amounts in the Vignettes Treatment by 10. In principle, these 18 CDFs could differ in arbitrary ways across treatments. It is instructive, however, to first consider what one would expect to see under the ESS model. If different treatments are associated with different δs, then we should observe differences across treatments in the distributions of choices for each characteristic function and each Recipient, *with only one exception*. The CDFs for Recipient 2 in CF4 should be the same *regardless of* δ, because what the Shapley value allocates to R2 in CF4 coincides with what equal split allocates to R2 in CF4.[31] Moreover, if δs tend to be higher in one treatment than another, one would indeed expect to see more (less) allocated to a given Recipient in the treatment with higher δs, whenever the Recipient's Shapley value is above (below) equal split.

We now examine our data from this perspective. Figure 11 superimpose the empirical CDFs of the different treatments in 3 different panels, one for each Recipient in CF4. Figure 16 in the Appendix contains the corresponding panels for all other characteristic functions. Consistent with the theoretical observation above, it appears from Figure 11 that the empirical CDF's of money allocated to R2 in CF4 are nearly identical across treatments. Indeed, for each pair of treatments, the Kolmogorov-Smirnov test cannot reject the null

[31]The nucleolus also prescribes that for R2, increasing the focality of that selection.



Recipient 1 Recipient 2 Recipient 3

Figure 11: Empirical CDFs across treatments, for CF4.

hypothesis that the allocations to R2 in CF4 come from the same distribution.

Beyond this, Figures 11 and 16 reveal that the empirical CDFs of money allocated in the Quiz and No-Quiz Treatments are quite similar in *every* panel, In each case, the Kolmogorov-Smirnov test cannot reject the null that the two distributions are the same. They also reveal that *in every panel except that of R2 in CF4*, the empirical CDFs from the Quiz and No-Quiz Treatments are ranked, or nearly ranked, by FOSD to the empirical CDF from the Vignettes Treatment. The direction of dominance reflects whether the given Recipient's Shapley value is superior or inferior to equal split in that characteristic function. Notice that R1's Shapley value is always above equal split, while R3's Shapley value is always below it. Many of these differences are highly statistically significant according to the Kolmogorov-Smirnov test, with the differences most significant across the No-Quiz and Vignettes Treatments.[32]

[32] When comparing the Quiz and Vignettes treatments, the null hypothesis that the monetary allocations come from the same distribution is rejected at the 5% level for 14 of the 17 panels, at the 1% level for 9 of the 17 panels, and at the 0.1% level for 2 of the 17 panels. When comparing the No-Quiz and Vignettes treatments, the same null hypothesis is rejected at the 5% level for 16 of the 17 panels, at the 1% level for 12 of the 17 panels, and at the 0.1% level for 5 of the 17 panels.

References

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.

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.

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

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, New York.

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.

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.

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

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.

Rogers, W. H., 1993. sg17: Regression Standard Errors in Clustered Samples. *Stata Technical Bulletin* **13**, 1923.

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.

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.

Yaari, Menahem E., and Maya Bar-Hillel, 1984. On Dividing Justly. *Social Choice and Welfare* **1**, 1-24.

Young, H. Peyton, 1985. Monotonic solutions of cooperative games. *International Journal of Game Theory* **14**, 65-72.