# Greater Flexibility in Mechanism Design Through Altruism

Ruben Brokkelkamp[1], Sjir Hoeijmakers[1], and Guido Schäfer[1,2]

[1] Centrum Wiskunde & Informatica (CWI), Amsterdam, The Netherlands
[2] ILLC, University of Amsterdam, The Netherlands

**Abstract.** We study the problem of designing truthful mechanisms for players that are (partially) altruistic. Our approach is to extend the standard utility model by encoding other-regarding preferences of the players into the utility functions. By doing so we leave the original domain where VCG mechanisms can be applied directly.

We derive a characterization of the class of truthful mechanisms under the new model, crucially exploiting the specific form of the other-regarding preferences. We also derive sufficient conditions for truthfulness which we then exploit to derive mechanisms for two specific models of altruism and with respect to two natural social welfare objectives. As it turns out, altruistic dispositions lead to the positive effect that the designer needs to extract less payments from the players to ensure truthfulness. Further, we investigate the effect of redistribution mechanisms that can redistribute the payments among the players. Also here it turns out that altruism has a positive effect in the sense that the payments needed to guarantee truthfulness can be further reduced.

Finally, we illustrate our theoretical results by applying them to well-studied mechanism design problems such as the public project problem and the multi-unit auction problem. Among other results, we show that the problem of funding a public project can be resolved by our mechanism even for moderate altruistic dispositions, while this is impossible in the standard utility setting.

## 1 Introduction

Most models in mathematical economics are based on the *self-interest hypothesis* which assumes that human beings take decisions following purely selfish motives. Certainly, this hypothesis applies in many economical settings, oftentimes simplifies analysis and allows us to make strong predictions on the outcome of economic situations. However, especially when they concern behavior of individuals, these predictions often reflect more what the outcomes 'should' be according to this assumption rather than what they actually are. This has become all the clearer in the past decades through the advent of behavioral economics. Various empirical studies have shown that this assumption oftentimes just 'fails' (see, e.g., [19,1,9]).

Mechanism design, in particular, is a branch that heavily relies on this assumption. On a high level, the goal here is to counter the negative effects of self-interested decision making in a group context. When a group of individuals has to take a decision (or someone has to make a choice on behalf of a group), simply asking each member of the

group individually how much they like each alternative might motivate them to over- or understate their actual preferences. A mechanism provides incentives to the involved individuals so that they will reveal their true preferences such that the best decision for the group as a whole can be made. Basically, the self-interest of the participants is used to make them act in the interest of the group after all. Of course, this comes at a cost: often payments need to be made by (or to) the participants. Mechanism design is the study of finding the 'best' mechanisms that incur the least 'damage' to the group (or the person making the decision) according to certain criteria.

If individuals would be fully altruistic in that their preferences are aligned with the group's interests, there would be no need for mechanism design. But what happens if they reside in the large spectrum between 'full altruism' and 'pure selfishness', i.e., when they care about others but not as much as about themselves? It seems a reasonable guess that nearly all human beings fall into this category and this constitutes the main motivation for our investigations in this paper. The main question that we address here is: How does partially altruistic behavior of the involved individuals impact the payments in mechanism design?

*Our contributions.* The main contributions of this paper are as follows:

1. We introduce a general utility model incorporating other-regarding preferences of players. Our approach is to adapt the standard utility model by adding to each player's utility an extra term which represents their dispositions towards the other players.
2. By adding these disposition functions, the utilities of the players become interdependent. As a consequence, the general class of VCG mechanisms cannot straightforwardly be applied to our setting. However, we are able to derive a characterization of truthful mechanisms in our new utility model with other-regarding preferences. The key in deriving our characterization is to exploit the specific form of the disposition functions.
3. Unfortunately, this characterization does not provide us with a "recipe" of how to obtain truthful mechanisms. We therefore establish a sufficient condition for truthfulness. We also derive sufficient and necessary conditions for when the resulting mechanisms satisfy the no positive transfer (NPT) and individual rationality (IR) property. This also serves as a design template for our mechanisms.
4. We then address the question of how the payments can be redistributed among the players (while maintaining truthfulness) such that the overall payments are minimized. In general, we cannot expect that such redistribution mechanism are strongly budget-balanced (i.e., the sum of the payments equals zero). We therefore use a relation of *individual dominance* between mechanisms, introduced by Apt et al. [18], and provide a characterization of such redistribution mechanisms for our new utility model.
5. We then consider two specific models of altruism that are captured by our utility model with other-regarding dispositions in combination with two natural social welfare objectives. We derive truthful mechanisms satisfying NPT and IR for all four settings. As it turns out, the altruistic dispositions of the players provide us with some additional flexibility in choosing the payments. A common property is

that as the degree of altruism of a player increases, the designer needs to pay them less to have them reveal their private valuations.

6. We demonstrate the usefulness of our mechanisms by applying them to some fundamental problems in mechanism design: For the bilateral trade problem, we show that our truthful mechanism can be run without any subsidy (if the involved players are sufficiently altruistic), while this is impossible when using VCG payments. For the public project problem, we show that our mechanism allows us to overcome some pathological deficiencies that are unavoidable in the standard setting. In fact, even a modest degree of altruism turns out to be sufficient to resolve the problem positively. Finally, we show that any mechanism that does not take altruism into account can be converted into a mechanism that does, and at the same time uses lower payments, where the gain is proportional to the altruism levels of the players.

Altogether, our results provide some evidence that altruism can only help in the mechanism design setting considered here. This is in contrast to some previous works (although in a purely strategic setting) showing that altruism may also have a negative impact on equilibrium outcomes[10,6,5].

*Related work.* There are different types of other-regarding preferences. In this work we focus mainly on altruism (and spite), other types are reciprocity [20] and inequity aversion [14]. For more information on the different types of other-regarding preferences, we refer to [13].

Although the role of altruism in algorithmic game theory has sparked some interest in recent years (see for instance [6,12,2,25,10,5], literature on incorporating altruism (or its counterpart spite) in mechanism design has been relatively scarce up to this point. We provide a few references of articles that go into this direction.

Brandt and Weiß [4] show that when spiteful bidders are present, the second-price auction fails in that it loses its favorable property of truthfulness. They do not present a truthful alternative themselves, but provide a motivation to research implications of spite (and altruism) in mechanism design and to search for such a truthful alternative. Tang and Sandholm [26] also consider single-item auctions and spiteful bidders and direct themselves to finding revenue-maximizing mechanisms. They model spite and altruism using a *player-oriented model*, which is a special case of our utility model with other-regarding preferences. In their proposed revenue-optimal mechanism, a player's own valuation for the object to be auctioned may directly influence their payment. Also, bidders may have to pay even if the auctioneer keeps the item, whereas on the other hand losers are sometimes subsidized by the mechanism. Kucuksenel [21] also models altruism according to the player-oriented model but studies its implications in the Bayesian setting. He characterizes a class of mechanisms that are *interim efficient*: they lead to outcomes which cannot be unanimously improved upon by the players utility-wise. Cavallo [8] proposes a regret-based model of altruism. In this model, players are $\alpha$-altruistic if they are willing to sacrifice up to $\alpha$ of their potential utility if that improves the aggregate egoistic utility. He also treats a 'proportional' variant, in which $\alpha$ is not a fixed value but a percentage of the potential utility of a player. In his paper, he uses redistribution to come up with strongly budget-balanced mechanisms (i.e. no net payments are made to or by the mechanism) for the single-item allocation setting when players are at least 'mildly' altruistic.

A simple yet effective way to redistribute a large portion of the surplus on the payments can be done by the Bailey-Cavallo redistribution function [3,7]. Other ways to redistribute payments are studied by Guo & Conitzer [16,17] and Moulin [23]. Once you start redistributing a natural question is what redistribution functions are best in some sense. Apt et al. [18] define partial orders on non-deficit Groves mechanisms and give characterizations of the maximal elements.

## 2 Preliminaries

We are given a finite set $N = \{1, \ldots, n\}$ of $n \geq 1$ players and a finite set $A$ of alternatives to choose from. Each player $i \in N$ has a *private valuation function* $v_i : A \to \mathbb{R}$ which specifies their preferences over the set of alternatives $A$, independently of the other players' preferences. Note that the valuation function $v_i$ is considered to be *private* information, i.e., only known to player $i$ themselves. Given an alternative $a \in A$, we say that $v_i(a)$ is the *valuation* of player $i$ for alternative $a$. We define $V_i$ as the set of all possible valuation functions of player $i$. Unless stated otherwise, we assume that $V_i \subseteq \mathbb{R}^A$ is unrestricted and commonly known. Define $V = V_1 \times \cdots \times V_n$.

We use the following standard notation. Given an $n$-dimensional vector $\boldsymbol{x} = (x_1, \ldots, x_n)$ of objects (i.e., reals, sets, functions) and a player $i \in N$, we define $\boldsymbol{x}_{-i} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$ as the same vector with the $i$-th component removed. We also slightly abuse notation and write $(x_i, \boldsymbol{x}_{-i})$ instead of $\boldsymbol{x}$. Similarly, we use $V_{-i}$ to refer to $V_1 \times \cdots \times V_{i-1} \times V_{i+1} \times \cdots \times V_n$.

Suppose there is a central *designer* (e.g., principal, government) who wants to determine a socially desirable outcome, taking the preferences of the players into account. Each player $i \in N$ expresses their preferences over the available alternatives by reporting a valuation function $b_i \in V_i$ (not necessarily equal to their private valuation function $v_i$). The designer then utilizes a mechanism to decide on an outcome. A *(direct revelation) mechanism* $M = (f, \boldsymbol{p})$ is specified by a *social choice function* $f : V \to A$ and a vector of *payment functions* $\boldsymbol{p} = (p_1, \ldots, p_n)$ with $p_i : V \to \mathbb{R}$ for all $i \in N$. Given the reported valuation functions $\boldsymbol{b} = (b_1, \ldots, b_n)$, the mechanism determines an alternative $f(\boldsymbol{b})$ and for each player $i \in N$ a payment $p_i(\boldsymbol{b})$ to be made to the designer.

We assume that each player wants to maximize a given utility function. In the *standard utility model*, each player $i \in N$ has a quasi-linear utility function defined as $u_i^s(\boldsymbol{b}) = v_i(f(\boldsymbol{b})) - p_i(\boldsymbol{b})$. The goal of the designer is to determine an alternative that maximizes a given *design objective* $D : V \times A \to \mathbb{R}$, i.e., $f(\boldsymbol{b}) \in \arg\max_{a \in A} D(\boldsymbol{b}, a)$. A commonly used design objective is to maximize the *social welfare*, i.e., the sum of the valuations of all players; formally, $D^{sw}(\boldsymbol{b}, a) = \sum_{i \in N} b_i(a)$. For any design objective considered in this paper, we assume that we can decompose $D(\boldsymbol{b}, a) = \sum_{i \in N} d_i(b_i, a)$ for functions $d_i : V_i \times A \to \mathbb{R}$. Then, we write $D_{-i}(\boldsymbol{b}, a) = D_{-i}(\boldsymbol{b}_{-i}, a) = \sum_{j \in N \setminus \{i\}} d_j(\boldsymbol{b}_j, a)$.

A mechanism $M = (f, \boldsymbol{p})$ is *truthful* if for every player $i \in N$, for any vector of reported valuations $\boldsymbol{b} \in V$, we have that $u_i(v_i, \boldsymbol{b}_{-i}) \geq u_i(b_i, \boldsymbol{b}_{-i})$. In other words, a truthful mechanism ensures that for each player $i$ it is always at least as good to report their private valuation $v_i$ than any other valuation, independent of what the other players report. Another desirable property of a mechanism is that it never makes payments to

the players. A mechanism $M$ satisfies the *no positive transfers (NPT)* property if for every player $i \in N$ and all $\boldsymbol{b} \in V$ we have that $p_i(\boldsymbol{b}) \geq 0$. Sometimes, we only require that the sum of payments is non-negative, i.e., $\sum_{i \in N} \boldsymbol{p}_i(\boldsymbol{b}) \geq 0$ for all $\boldsymbol{b} \in V$, then we call the mechanism *non-deficit*. Finally, every player should be guaranteed to receive a non-negative utility if they report their valuations truthfully. A mechanism $M$ satisfies the *individual rationality (IR)* property if for every player $i \in N$ and for all reported valuations $\boldsymbol{b}_{-i} \in V_{-i}$ of the other players, $u_i(v_i, \boldsymbol{b}_{-i}) \geq 0$.

**Definition 1.** *A mechanism $M = (f, \boldsymbol{p})$ is called a* Vickrey-Clarke-Groves (VCG) mechanism *if the following two conditions are satisfied:*

1. $f(\boldsymbol{b}) \in \arg\max_{a \in A} \sum_{i \in N} b_i(a)$*;*
2. *for every player $i \in N$ there is a function $h_i : V_{-i} \to \mathbb{R}$ such that*

$$p_i(\boldsymbol{b}) = h_i(\boldsymbol{b}_{-i}) - \sum_{j \neq i} b_j(f(\boldsymbol{b})).$$

VCG mechanisms allow for different instantiations of functions $h_i$ to define the payments of the players. However, if the valuation functions are non-negative and one additionally insists on satisfying both NPT and IR then there remains a unique payment rule due to Clarke (1971): A VCG mechanism $(f, \boldsymbol{p})$ implements the *Clarke pivot rule* if for every player $i \in N$ we have that $h_i(\boldsymbol{b}_{-i}) = \sum_{j \neq i} b_j(a^{-i})$, where $a^{-i} \in \arg\max_{a \in A} \sum_{j \neq i} b_j(a)$ is an alternative that maximizes the social welfare if player $i$ would not be present.

The following is due to [27,11,15].

**Proposition 1.** *Every VCG mechanism is truthful. The VCG mechanism that uses the Clarke pivot rule satisfies NPT. Further, if all valuation functions of the players are non-negative, then it also satisfies IR.*

Due to page limitations some proofs are deferred to the appendix.

## 3 Modeling Other-Regarding Preferences

### 3.1 Utility Model with Other-Regarding Preferences

We propose a general utility model capturing that players may care about other players (both in the positive and negative sense).

**Definition 2.** *Suppose we are given a function $g_i : \mathbb{R}^{n-1} \times \mathbb{R}^n \to \mathbb{R}$ for every player $i \in N$ modeling their* other-regarding preferences*. The utility $u_i^{g_i}$ of player $i \in N$ in the* utility model with other-regarding preferences *is then defined as*

$$u_i^{g_i}(\boldsymbol{b}) = v_i(f(\boldsymbol{b})) - p_i(\boldsymbol{b}) \; + g_i(\boldsymbol{b}_{-i}(f(\boldsymbol{b})), \boldsymbol{p}(\boldsymbol{b}))$$

Observe that the function $g_i$ does not depend on the private valuations of the other players (which would be infeasible). This reflects the intuition that the other-regarding

preferences of a player originate from *beliefs* about the experiences of others rather than from their true experiences.[3]

In the definition the other-regarding preferences are allowed to depend on the payments. This makes it very general, but in proofs this poses an extra challenge and so for some results below we make the following assumption.

**Assumption 1.** *The other-regarding preferences $g_i$ do not depend on the payments and therefore only depend on $\boldsymbol{b}_{-i}(f(\boldsymbol{b}))$ for all $i \in N$.*

We will see in Section 5 that there are natural models where the other-regarding preferences do not depend on the payments.

### 3.2 Characterization of truthful mechanisms

**Theorem 1.** *A mechanism $M = (f, \boldsymbol{p})$ is truthful in the utility model with other-regarding preferences if and only if it satisfies the following two conditions:*

1. *For every player $i \in N$ the difference between the other-regarding preferences $g_i$ and the payment $p_i$ only depends on the chosen alternative $f(\boldsymbol{b})$ and $\boldsymbol{b}_{-i}$ (but not on $b_i$ itself), i.e., there is a function $\mu_i : A \times V_{-i} \to \mathbb{R}$ such that*

$$p_i(\boldsymbol{b}) - g_i(\boldsymbol{b}_{-i}(f(\boldsymbol{b})), \boldsymbol{p}(\boldsymbol{b})) = \mu_i(f(\boldsymbol{b}), \boldsymbol{b}_{-i}).$$

2. *The alternative chosen by $M$ satisfies for every player $i \in N$ that*

$$f(b_i, \boldsymbol{b}_{-i}) \in \arg\max_{a \in A(\boldsymbol{b}_{-i})} (b_i(a) - \mu_i(a, \boldsymbol{b}_{-i})),$$

*where $A(\boldsymbol{b}_{-i}) = \{f(b'_i, \boldsymbol{b}_{-i}) \mid b'_i \in V_i\}$ refers to the image of $f(\cdot, \boldsymbol{b}_{-i})$.*

*Proof (Theorem 1).* We first prove the if part. Consider a player $i \in N$ and fix $\boldsymbol{b}_{-i} \in V_{-i}$ arbitrarily. Define $\bar{a} = f(v_i, \boldsymbol{b}_{-i})$ and $a = f(b_i, \boldsymbol{b}_{-i})$ as the alternatives chosen by $M$ when $i$ reports their private valuation function $v_i$ truthfully and when $i$ reports an arbitrary valuation function $b_i$, respectively.

By the first condition of the statement, we have

$$u_i^{g_i}(v_i, \boldsymbol{b}_{-i}) = v_i(\bar{a}) - \mu_i(\bar{a}, \boldsymbol{b}_{-i}) \quad \text{and} \quad u_i^{g_i}(b_i, \boldsymbol{b}_{-i}) = v_i(a) - \mu_i(a, \boldsymbol{b}_{-i}). \tag{1}$$

By the second condition, the alternative $\bar{a}$ chosen by $M$ for $(v_i, \boldsymbol{b}_{-i})$ satisfies

$$v_i(\bar{a}) - \mu_i(\bar{a}, \boldsymbol{b}_{-i}) \geq v_i(a) - \mu_i(a, \boldsymbol{b}_{-i}). \tag{2}$$

Combining (1) and (2) proves truthfulness.

Now we prove the only-if part of the first condition. Consider a player $i \in N$ and fix an arbitrary $\boldsymbol{b}_{-i} \in V_{-i}$. Given some $b_i \in V_i$, for notational convenience we define $m_i(b_i, \boldsymbol{b}_{-i})$ as a shorthand for

$$m_i(b_i, \boldsymbol{b}_{-i}) = p_i(b_i, \boldsymbol{b}_{-i}) - g_i(\boldsymbol{b}_{-i}(f(b_i, \boldsymbol{b}_{-i})), \boldsymbol{p}(b_i, \boldsymbol{b}_{-i}))$$

[3] Intuitively, the function $g_i$ of player $i$ can be viewed as being dependent on the reported valuation functions $\boldsymbol{b}_{-i}$ of the other players and the payment functions $\boldsymbol{p}$. Formally, however $g_i$ only depends on the respective *values* of these functions under the outcome $(f(\boldsymbol{b}), \boldsymbol{p}(\boldsymbol{b}))$ determined by the mechanism $M = (f, \boldsymbol{p})$ when run on $\boldsymbol{b}$.

The utility of player $i$ can then be written as $u_i^{g_i}(b_i, \boldsymbol{b}_{-i}) = v_i(f(b_i, \boldsymbol{b}_{-i})) - m_i(b_i, \boldsymbol{b}_{-i})$. Suppose there are two valuation functions $b_i, b_i' \in V_i$ of player $i$ such that $f(b_i, \boldsymbol{b}_{-i}) = f(b_i', \boldsymbol{b}_{-i})$ and $m_i(b_i, \boldsymbol{b}_{-i}) < m_i(b_i', \boldsymbol{b}_{-i})$. Then by identifying the private valuation function $v_i$ of $i$ with $b_i'$, we obtain

$$\begin{aligned} u_i^{g_i}(v_i, \boldsymbol{b}_{-i}) &= v_i(f(b_i', \boldsymbol{b}_{-i})) - m_i(b_i', \boldsymbol{b}_{-i}) \\ &< v_i(f(b_i, \boldsymbol{b}_{-i})) - m_i(b_i, \boldsymbol{b}_{-i}) = u_i^{g_i}(b_i, \boldsymbol{b}_{-i}), \end{aligned}$$

which contradicts the truthfulness of $M$. Thus $m_i(b_i, \boldsymbol{b}_{-i}) = m_i(b_i', \boldsymbol{b}_{-i})$ whenever $f(b_i, \boldsymbol{b}_{-i}) = f(b_i', \boldsymbol{b}_{-i})$. This proves the existence of a function $\mu_i$ only depending on $f(\boldsymbol{b})$ and $\boldsymbol{b}_{-i}$ as claimed.

Finally, we prove the only-if part of the second condition. Again, consider a player $i \in N$ and fix an arbitrary $\boldsymbol{b}_{-i} \in V_{-i}$. Suppose there is some $b_i \in V_i$ such that $f(b_i, \boldsymbol{b}_{-i})$ is not a maximizer of the expression. Let $a' \in A(\boldsymbol{b}_{-i})$ be such a maximizer, i.e.,

$$a' \in \arg \max_{a \in A(\boldsymbol{b}_{-i})} (b_i(a) - \mu_i(a, \boldsymbol{b}_{-i})).$$

By the definition of $A(\boldsymbol{b}_{-i})$, we have $a' = f(b_i', \boldsymbol{b}_{-i})$ for some $b_i' \in V_i$. By identifying the private valuation function $v_i$ of $i$ with $b_i$ and defining $\bar{a} = f(b_i, \boldsymbol{b}_{-i})$, we obtain

$$u_i^{g_i}(v_i, \boldsymbol{b}_{-i}) = b_i(\bar{a}) - \mu_i(\bar{a}, \boldsymbol{b}_{-i}) < b_i(a') - \mu_i(a', \boldsymbol{b}_{-i}) = u_i^{g_i}(b_i', \boldsymbol{b}_{-i}),$$

which contradicts the truthfulness of $M$. □

It might be difficult to know what the other-regarding preferences look like for a specific player $i$. When making assumption 1 we actually do not have to worry too much about truthfulness. Even when we design a mechanism assuming different other-regarding preferences it still works out.

**Corollary 1.** *Making Assumption 1. Suppose we have a truthful mechanism with respect to some other-regarding preferences $g_i'$. Then this mechanism is also truthful with respect to $g_i$.*

### 3.3 Design template

Theorem 1 gives a characterization of truthful mechanisms but does not provide us with a "recipe" of how to obtain such mechanisms for a given design objective $D$.

**Theorem 2.** *Fix a design objective $D$. A mechanism $M = (f, \boldsymbol{p})$ is truthful in the utility model with other-regarding preferences if the following two conditions are satisfied:*

1. $f(\boldsymbol{b}) \in \arg\max_{a \in A} D(\boldsymbol{b}, a)$.
2. *For every player $i \in N$ there exist functions $h_i, \gamma_i : V_{-i} \to \mathbb{R}$ such that*

$$p_i(\boldsymbol{b}) = h_i(\boldsymbol{b}_{-i}) + g_i(\boldsymbol{b}_{-i}(f(\boldsymbol{b})), \boldsymbol{p}(\boldsymbol{b})) - \gamma_i(\boldsymbol{b}_{-i}) \cdot D_{-i}(\boldsymbol{b}_{-i}, f(\boldsymbol{b})).$$

As in the standard utility setting, we would like to ensure that our mechanisms satisfy NPT and IR. In light of Theorem 2, it is now easy to specify which choices of the functions $h_i$ are feasible for this.

**Proposition 2.** *Let $M = (f, \boldsymbol{p})$ be a mechanism as defined in Theorem 2. Then $M$ satisfies NPT if and only if for every player $i \in N$, $h_i(\boldsymbol{b}_{-i}) \geq \gamma_i(\boldsymbol{b}_{-i}) \cdot D_{-i}(\boldsymbol{b}_{-i}, f(\boldsymbol{b})) - g_i(\boldsymbol{b}_{-i}(f(\boldsymbol{b})), \boldsymbol{p}(\boldsymbol{b}))$. Further, $M$ satisfies IR if and only if for every player $i \in N$, $h_i(\boldsymbol{b}_{-i}) \leq \gamma_i(\boldsymbol{b}_{-i}) \cdot D_{-i}(\boldsymbol{b}) + v_i(f(\boldsymbol{b}))$.*

The above proposition shows that in principle there is a leeway of choosing $h_i$ of size $v_i(f(\boldsymbol{b})) + g_i(\boldsymbol{b}_{-i}(f(\boldsymbol{b})), \boldsymbol{p}(\boldsymbol{b}))$; however, recall that $h_i$ may only depend on $\boldsymbol{b}_{-i}$ and we might thus be unable to exploit the full range. Further, Proposition 2 shows that if the valuation functions can be negative then we cannot guarantee both NPT and IR. In fact, the same holds if $g_i(\boldsymbol{b}_{-i}(f(\boldsymbol{b})), \boldsymbol{p}(\boldsymbol{b}))$ is allowed to be negative.

What effect does not knowing $g_i$ exactly have on individual rationality? When making Assumption 1, as long as we underestimate it, we do not have to worry about it.

**Proposition 3.** *Making Assumption 1. Let $g_i, g_i'$ be a other-regarding preferences such that $g_i(\boldsymbol{b}_{-i}(f(\boldsymbol{b}))) \geq g_i'(\boldsymbol{b}_{-i}(f(\boldsymbol{b})))$ for all $\boldsymbol{b}$. Let $M = (f, \boldsymbol{p})$ be a mechanism as in Theorem 2 with respect to $g_i'$ for player $i$ that is individually rational. The mechanism with respect to $g_i$ is also truthful and individually rational.*

## 4 Minimizing Payments

As mentioned in the previous section there is a leeway of choosing $h_i$. There are situations in which extracting higher payments is preferred. Often, when a seller sells an item via an auction they want to extract highest payments possible. On the other hand, if a group of housemates have a shared car and they have to decide who gets to use it on Friday they might not want to use any payments at all, and they definitely do not want payments to go to some third party overseeing the decision. Especially in these situations one can argue that higher altruism levels are more plausible as the players are more familiar with each other. In Section 6.3 we will use the results from this section to observe that incorporating altruism can significantly reduce the surplus of the payments of the mechanism.

Holding on to IR and NPT, and wanting to use a mechanism that follows the recipe from Theorem 2, Proposition 2 restricts us to payment payment functions that satisfy

$$h_i(\boldsymbol{b}_{-i}) \geq \gamma_i(\boldsymbol{b}_{-i}) \cdot D_{-i}(\boldsymbol{b}_{-i}, f(\boldsymbol{b})) - g_i(\boldsymbol{b}_{-i}(f(\boldsymbol{b})), \boldsymbol{p}(\boldsymbol{b}))$$

In Theorem 2 the payments depend on the other-regarding preferences which in turn can depend on the payments. To avoid issues with recursive definitions we make Assumption 1 in this section.

Making Assumption 1 we can easily find the payment functions that minimize the sum of payments $\sum_i p_i(b) = \sum_i h_i(\boldsymbol{b}_{-i}) + g_i(\boldsymbol{b}_{-i}(f(\boldsymbol{b}))) - D_{-i}(\boldsymbol{b}_{-i}, f(\boldsymbol{b}))$ while making sure NPT still holds:

$$h_i(\boldsymbol{b}_{-i}) = \sup_{b_i'} \gamma_i(\boldsymbol{b}_{-i}) \cdot D_{-i}(\boldsymbol{b}_{-i}, f(b_i', \boldsymbol{b}_{-i})) - g_i(\boldsymbol{b}_{-i}(f(\boldsymbol{b}))) \tag{3}$$

here we write $\boldsymbol{b}' = (b_i', \boldsymbol{b}_{-i})$. If it is clear what $b_i'$ is, we use this notation throughout this section.

There is no need to hold on to NPT in general. Actually, in some situations, we would rather have that the payments are redistributed over the players, so that as little money as possible is wasted on a trusted third party. Unfortunately, it is impossible to have a *strongly budget-balanced*, i.e., the sum of payments is 0, mechanism in many settings [22,24]. As we cannot aim for 0, we would like to minimize the amount we cannot redistribute.

Letting go of NPT makes us much more flexible. However, we do not want to subsidize the mechanism and thus we keep the requirement that the mechanism should be non-deficit, i.e., $\sum_i \boldsymbol{p}_i(\boldsymbol{b}) \geq 0$ for all $\boldsymbol{b}$.

Apt et al. [18] characterized Groves mechanisms that are undominated in terms of the amount of money flowing from the mechanism to the auctioneer. We will extend this to the mechanisms with other-regarding preferences.

We say that a non-deficit mechanism with payment vector $\boldsymbol{p}$ *collectively dominates* a non-deficit mechanism with payment vector $\boldsymbol{p}'$ if for all $\boldsymbol{b}$ : $\sum_i \boldsymbol{p}_i(\boldsymbol{b}) \leq \sum_i \boldsymbol{p}'_i(\boldsymbol{b})$ and there is at least one $\boldsymbol{b}$ for which this inequality is strict. Getting characterizations for payments that are collectively undominated is difficult. We can however relax the requirement a bit and for the following partial order we are able to find a characterization of the maximal elements.

**Definition 3.** *A non-deficit mechanism with payment vector $\boldsymbol{p}$ is said to individually dominate a non-deficit mechanism with payment vector $\boldsymbol{p}'$ if for all $\boldsymbol{b}$ and $i$*

$$\boldsymbol{p}_i(\boldsymbol{b}) \leq \boldsymbol{p}'_i(\boldsymbol{b})$$

*and there is at least one $\boldsymbol{b}$ and $i$ for which $\boldsymbol{p}_i(\boldsymbol{b}) < \boldsymbol{p}'_i(\boldsymbol{b})$.*

Using payments of the form as in Theorem 2 we closely follow Apt et al. [18] to characterize when a mechanism is non-deficit and when it is individually undominated.

**Lemma 1.** *A mechanism $M = (f, \boldsymbol{p})$ with $\boldsymbol{p}_i = h_i(\boldsymbol{b}_{-i}) + g_i(\boldsymbol{b}_{-i}(f(\boldsymbol{b}))) - \gamma_i(\boldsymbol{b}_{-i}) \cdot D_{-i}(\boldsymbol{b}_{-i}, f(\boldsymbol{b}))$ is non-deficit if and only if for all $i$ and $\boldsymbol{b}_{-i}$*

$$h_i(\boldsymbol{b}_{-i}) \geq \sup_{b'_i} \sum_j (\gamma_j(\boldsymbol{b}'_{-j}) \cdot D_{-j}((\boldsymbol{b}'_{-j}, f(\boldsymbol{b}'))) - g_j(\boldsymbol{b}'_{-j}(f(\boldsymbol{b}')))) - \sum_{j \neq i} h_j(\boldsymbol{b}'_{-j}) \quad (4)$$

**Theorem 3.** *A mechanism $M = (f, \boldsymbol{p})$ with $\boldsymbol{p}_i = h_i(\boldsymbol{b}_{-i}) + g_i(\boldsymbol{b}_{-i}(f(\boldsymbol{b}))) - \gamma_i(\boldsymbol{b}_{-i}) \cdot D_{-i}(\boldsymbol{b}_{-i}, f(\boldsymbol{b}))$ is individually undominated if and only if for all $i$ and $\boldsymbol{b}_{-i}$*

$$h_i(\boldsymbol{b}_{-i}) = \sup_{b'_i} \sum_j \left(\gamma_j(\boldsymbol{b}'_{-j}) \cdot D_{-j}(\boldsymbol{b}'_{-j}(f(\boldsymbol{b}'))) - g_j(\boldsymbol{b}'_{-j}(f(\boldsymbol{b})))\right) - \sum_{j \neq i} h_j(\boldsymbol{b}'_{-j}) \quad (5)$$

## 5 A case study: Altruism

### 5.1 Two altruism models and design objectives

We consider two models of altruism that are instantiations of the utility model with other-regarding preferences. We assume that each player $i \in N$ is equipped with an

*altruism level* $\alpha_i \in [0,1]$ which interpolates between a 'purely selfish' ($\alpha_i = 0$) and a 'fully altruistic' ($\alpha_i = 1$) attitude[4].

**Definition 4.** *Given an altruism level* $\alpha_i \in [0,1]$ *for every player* $i \in N$*, in the* welfare-oriented model *the utility* $u_i^w : V \to \mathbb{R}$ *of player* $i \in N$ *is defined as:*

$$u_i^w(\boldsymbol{v}) = v_i(f(\boldsymbol{b})) - p_i(\boldsymbol{b}) \ + \alpha_i \sum_{j \neq i} b_j(f(\boldsymbol{b})).$$

In the welfare-oriented model each player $i$ receives a fraction of $\alpha_i$ of the reported valuations of all other players. Note that $i$ fully cares about their own payment. Altruism here corresponds to a willingness to contribute to the creation of value in the form of valuations of alternatives.

**Definition 5.** *Given an altruism level* $\alpha_i \in [0,1]$ *for every player* $i \in N$*, in the* omnistic model *the utility* $u_i^o : V \to \mathbb{R}$ *of a player* $i \in N$ *is given by:*

$$u_i^o(\boldsymbol{b}) = v_i(f(\boldsymbol{b})) - p_i(\boldsymbol{b}) \ + \alpha_i \left( p_i(\boldsymbol{b}) + \sum_{j \neq i} b_j(f(\boldsymbol{b})) \right).$$

In the omnistic model each player $i$ cares about every other player the same way as in the welfare-oriented model. The difference however is that player $i$ perceives their payment $p_i$ to the designer as being discounted by a fraction of $(1 - \alpha_i)$ (although they pay $p_i$ eventually). Put differently, $i$ enjoys a fraction of $\alpha_i$ of the payment $p_i$ that the designer receives from them. This is also the reason why we call this the 'omnistic' model (omnes = all/everybody).

We derive mechanisms for these models with respect to the following two design objectives:

$$D^{sw}(\boldsymbol{b}, a) = \sum_{i \in N} b_i(a) \quad \text{and} \quad D^{dw}(\boldsymbol{b}, a) = \sum_{i \in N} \left(1 + \sum_{k \neq i} \alpha_k\right) b_i(a)$$

The design objective $D^{sw}$ is the classical social welfare objective. In our context, it captures situations where the designer only cares about the sum of the individual valuations of the players, disregarding the positive perceptions that they receive from other players. Intuitively, here the utility functions serve merely as a means to model the positive attitudes of players towards others.

The design objective $D^{dw}$ models situations in which the designer takes both the individual valuations of the players and their positive other-regarding preferences towards others into account. Note that this objective is equal to the sum of all valuations that the players receive (directly or indirectly). We refer to it as the *dispositional welfare* objective.

[4] Note that although our focus here is on altruism levels in the range $[0,1]$ many results given below can be extended in a straight-forward way to other cases such as spiteful players ($\alpha_i < 0$) or players that care about others more than about themselves ($\alpha_i > 1$).

| $(m, D)$ | payment function and altruism-adjusted Clarke pivot rule |
|---|---|
| $(w, D^{sw})$ | $p_i(\boldsymbol{b}) = h_i(\boldsymbol{b}_{-i}) - (1-\alpha_i)\sum_{j\neq i} b_j(f(\boldsymbol{b}))$<br>$h_i(\boldsymbol{b}_{-i}) = (1-\alpha_i+c_i)\sum_{j\neq i} b_j(a^{-i})$,<br>$a^{-i} \in \arg\max_{a\in A} \sum_{j\neq i} b_j(a)$ |
| $(o, D^{sw})$ | $p_i(\boldsymbol{b}) = \frac{1}{1-\alpha_i} h_i(\boldsymbol{b}_{-i}) - \sum_{j\neq i} b_j(f(\boldsymbol{b}))$<br>$h_i(\boldsymbol{b}_{-i}) = (1-\alpha_i+c_i)\sum_{j\neq i} b_j(a^{-i})$,<br>$a^{-i} \in \arg\max_{a\in A} \sum_{j\neq i} b_j(a)$ |
| $(w, D^{dw})$ | $p_i(\boldsymbol{b}) = h_i(\boldsymbol{b}_{-i}) - \sum_{j\neq i} \left(\frac{1+\sum_{k\neq j}\alpha_k}{1+\sum_{k\neq i}\alpha_k} - \alpha_i\right) b_j(f(\boldsymbol{b}))$<br>$h_i(\boldsymbol{b}_{-i}) = \sum_{j\neq i} \left(\frac{1+\sum_{k\neq j}\alpha_k}{1+\sum_{k\neq i}\alpha_k} - \alpha_i + c_i\right) b_j(a^{-i})$,<br>$a^{-i} \in \arg\max_{a\in A} \sum_{j\neq i} \left(1+\sum_{k\neq j,i}\alpha_k - \alpha_i \sum_{k\neq i}\alpha_k\right) b_j(a)$ |
| $(o, D^{dw})$ | $p_i(\boldsymbol{b}) = h_i(\boldsymbol{b}_{-i}) - \frac{1}{1-\alpha_i}\sum_{j\neq i} \left(\frac{1+\sum_{k\neq j}\alpha_k}{1+\sum_{k\neq i}\alpha_k} - \alpha_i\right) b_j(f(\boldsymbol{b}))$<br>$h_i(\boldsymbol{b}_{-i}) = \frac{1}{1-\alpha_i}\sum_{j\neq i} \left(\frac{1+\sum_{k\neq j}\alpha_k}{1+\sum_{k\neq i}\alpha_k} - \alpha_i + c_i\right) b_j(a^{-i})$<br>$a^{-i} \in \arg\max_{a\in A} \sum_{j\neq i} \left(1+\sum_{k\neq j,i}\alpha_k - \alpha_i \sum_{k\neq i}\alpha_k\right) b_j(a)$ |

**Table 1.** Definition of the payment function of the AAVCG mechanisms and its altruism-adjusted Clarke pivot rule, depending on the altruism model and design objective. The parameter $c_i$ can be fixed arbitrarily in the range $[0, \alpha_i]$

### 5.2 Mechanisms for altruistic players

We derive truthful mechanisms for the welfare-oriented and omnistic model (referred to as $w$ and $o$ for short) with respect to both the social and dispositional welfare objective. In order to keep the presentation concise, we introduce the following generic definition of adjusted VCG mechanisms. The respective payment functions $\boldsymbol{p}$ are stated in Table 1.

**Definition 6.** *Let $m \in \{w, o\}$ refer to an altruism model as defined above and let $D \in \{D^{sw}, D^{dw}\}$ be a design objective. A mechanism $M^{m,D} = (f, \boldsymbol{p})$ is called an* altruism-adjusted VCG mechanism (AAVCG) with respect to altruism model $m$ and design objective $D$ *if the following two conditions are satisfied:*

1. $f(\boldsymbol{b}) \in \arg\max_{a\in A} D(\boldsymbol{b}, a)$.
2. *For every player $i \in N$ and some function $h_i : V_{-i} \to \mathbb{R}$, the payment function $p_i(\boldsymbol{b})$ is defined as in Table 1.*

Similarly, we give a generic definition of an altruism-adjusted Clarke pivot rule for these mechanism. The respective definitions of the functions $h_i$ are stated in Table 1.

**Definition 7.** *We say that an AAVCG mechanism $M^{m,D} = (f, \boldsymbol{p})$ with respect to altruism model $m$ and design objective $D$ implements the* altruism-adjusted Clarke pivot rule *if for every player $i \in N$ there is some $c_i \in [0, \alpha_i]$ such that the function $h_i$ is as defined in Table 1.*

With the help of Theorem 2 and Proposition 2, we can show that the four AAVCG mechanisms as specified in Table 1 are truthful and satisfy NPT and IR.

**Theorem 4.** *Every AAVCG mechanisms $M^{m,D} = (f, \boldsymbol{p})$ with respect to altruism model $m$ and design objective $D$ is truthful. Further, $M^{m,D}$ satisfies NPT and IR if it implements the altruism-adjusted Clarke pivot rule.*

### 5.3 Discussion

We discuss a few main properties of the mechanisms introduced above.

First note that the two AAVCG mechanisms for the social welfare objective reduce to the standard VCG mechanism (Definition 1) if the players are entirely selfish, i.e., $\alpha_i = 0$ for all $i$. This is to be expected because in this case both the welfare-oriented model and the omnistic model reduce to the standard utility model.

Further, these mechanism nicely capture the intuition that altruism counters the negative effect of egoistic predispositions: As the altruism level of a player $i$ increases, the designer needs to pay them less to have them want to reveal the truth about their valuation functions. And in fact, as we would expect the players require no extra incentive at all when they are fully altruistic ($\alpha_i = 1$ for all $i$).

Observe that the altruism-adjusted Clarke pivot gives rise to a family of mechanisms (parametrized by $c_i \in [0, \alpha_i]$ for all $i$). The size of this set grows with the altruism levels $\alpha_i$ of the players. This flexibility can be exploited to extract smaller payments from the players.

The altruism-adjusted Clarke pivot rule has a particularly nice representation in the omnistic model with respect to the social welfare objective, i.e., for every $c_i \in [0, \alpha_i]$

$$p_i(\boldsymbol{b}) = \left(1 + \frac{c_i}{1 - \alpha_i}\right) \sum_{j \neq i} b_j(a^{-i}) - \sum_{j \neq i} b_j(f(\boldsymbol{b})), \tag{6}$$

where $a^{-i}$ is as defined in Table 1.

Note that by choosing $c_i = 0$ for every $i$, the resulting AAVCG mechanism reduces to the standard VCG mechanism with Clarke pivot rule. In particular, this means that for this setting the standard VCG mechanism (not taking care of any other-regarding preferences) is truthful.

## 6 Impact of altruism

### 6.1 Bilateral Trade

A buyer is interested in some object and values it at $v_b$, while some seller has the object and values it at $v_s$. How to trade?

In a mechanism design context this is defined as: (i) The set of alternatives is $A = \{\textit{trade}, \textit{no-trade}\}$ (ii) The buyer values trading at $v_b(\textit{trade}) = v_b$, and not trading at $v_b(\textit{no-trade}) = 0$ (iii) The seller values trading at $v_s(\textit{trade}) = -v_s$ and not trading at $v_b(\textit{no-trade}) = 0$

If we require that there are no payments when there is no trade VCG gives us that if the trade happens, i.e., $v_b \geq v_s$, then we charge the buyer $v_s$ and the seller gets $v_b$. But $v_b > v_s$ implies that the mechanism needs to be subsidized.

In contrast, suppose the buyer has an altruism level of $-\alpha_b$ and the seller of $\alpha_s$, then using the welfare-oriented model we should charge the buyer $(1 + \alpha_b)v_s$ and the seller gets $(1 - \alpha_s)v_b$. Hence, if

$$(1 + \alpha_b)v_s \geq (1 - \alpha_s)v_b$$

we can run this mechanism without subsidizing it.

### 6.2 Funding a public project

In the *public project problem* a contractor (e.g., government) considers to undertake a public project (e.g., building a bridge) at a commonly known cost $C$. Each player $i \in N$ (e.g., citizen) reports a value $b_i$ that the realization of the project is worth to them (not necessarily equal to their private value $v_i$). Given the reports $(b)_{i \in N}$, the contractor determines whether the project is realized and what the contribution $p_i$ of every player $i \in N$ is. Here the project is realized if and only if it can be *funded* by the players, i.e., $\sum_{i \in N} p_i \geq C$.

This models a very realistic situation. It would be desirable that the theory of mechanism design provides us with a mechanism that ensures that the project is undertaken when it should be undertaken, i.e., when the actual value created by the realization of the project is at least $C$; formally, $\sum_{i \in N} v_i \geq C$. This is precisely what a truthful mechanism maximizing social welfare would achieve. Unfortunately, the only instances of the public project problem that can be solved in the standard mechanism design setting are trivial ones.

Formally, the *public project problem* in a mechanism design context can be defined as follows (after Clarke (1971)): (i) The set of alternatives is $A = \{\text{yes}, \text{no}\}$. (ii) The set of players consists of $N = \{1, \ldots, n\}$ and a special player 0, representing the contractor. (iii) Player $i = 0$ has a singleton valuation set $V_i = \{v_i\}$ with $v_i(\text{yes}) = -C$ for some $C \in \mathbb{R}^+$ and $v_i(\text{no}) = 0$. (iv) Every player $i \in N$ has a valuation set $V_i$ such that for every $v_i \in V_i$ we have $v_i(\text{yes}) = w_i$ for some $w_i \in \mathbb{R}^+$ and $v_i(\text{no}) = 0$. (v) The design objective is the social welfare $D^{sw}$.

Note that player $i = 0$ is essentially a dummy player as they do not have any choice other than reporting their valuation function truthfully. Also their valuation is $-C$ (reflecting that the realization of the project incurs a cost to them). In particular, they cannot be asked to contribute anything to the project. We say that the project is *funded* if $\sum_{i \in N} p_i(\boldsymbol{b}) \geq C$.

*Why standard mechanism design fails.* The VCG mechanism with the Clarke pivot rule is known to be the only truthful mechanism that satisfies individual rationality (given the social welfare design objective); see, e.g., Nisan et al. (2007, Chapter 9). In order to understand how this mechanism determines the payments in the public project problem, we need the following concept: Given the reports $(b_i)_{i \in N}$, a player $i \in N$ is called *pivotal* if $\sum_j b_j \geq C$ but $\sum_{j \neq i} b_j < C$; otherwise, $i$ is *non-pivotal*. In other words, a pivotal player is essential to make the project fundable.

The following proposition characterizes when a project is funded:

**Proposition 4.** *Using the VCG mechanism with Clarke pivot rule, the public project in the standard utility model is funded if and only if*

1. $b_i \geq C$ *for some* $i \in N$ *and* $b_j = 0$ *for all* $j \in N$, $j \neq i$, *or*
2. $\sum_{i \in N} b_i = C$.

As a consequence, a public project is only funded if there is exactly one player who benefits from it, or if there is no benefit at all but just a break even between the value created and the investment costs incurred.

We next show how altruism helps to escape the above dilemma. More specifically, we consider the omnistic model and use the VCG mechanism with the altruism-adjusted Clarke pivot rule (6) with $c_i = \alpha_i$ for all $i$.

**Proposition 5.** *Let $N_p$ be the set of pivotal players, and $N_n$ the set of non-pivotal players. Using the VCG mechanism with the altruism-adjusted Clarke pivot rule (choosing $c_i = \alpha_i$ for all $i$), the public project in the omnistic model is funded if and only if*

$$\sum_{i \in N_p} \left( C - \sum_{j \neq i} b_j \right) + \sum_{i \in N_n} \frac{\alpha_i}{1 - \alpha_i} \left( \sum_{j \neq i} b_j - C \right) \geq C$$

What does Proposition 5 tell us? First of all, we see that the project is more likely to be fully funded when it is more profitable for the group to undertake it. For the pivotal players, the higher the contribution of the other players , the higher $C - \sum_{j \neq i} b_j$ and for the non-pivotal players it is clear that the more profitable the project is the higher $\sum_{j \neq i} b_j - C$. This relation is rather satisfying, especially if one compares it with the results for the standard utility model (where, paradoxically, the larger the net benefits of the project are, the less likely it is that it will be funded).

Secondly, we observe that altruism of non-pivotal players *always* has a positive effect on the likelihood of funding the project and this effect is amplified when the altruism levels of the (non-pivotal) players with small valuation is large. This effect becomes even more apparent if one considers uniform altruism levels ($\alpha_i = \alpha$ for all $i$) and only non-pivotal players. The condition of Proposition 5 simplifies to

$$\alpha \left( \sum_{j \in N} b_j - C \right) \geq \frac{C}{n-1}.$$

Also note that the number of players $n$ is positively related to the likelihood of funding the project.

### 6.3 Minimizing payments

We make Assumption 1. Given a mechanism where we do not take altruism into account but where players are actually altruistic we will transform it into a mechanism with smaller total payment. Consider welfare-oriented other-regarding preferences. Let every

player $i$ have an altruism level of $\alpha_i$. Define $\alpha = \min_i \alpha_i$. Suppose we have some non-deficit mechanism satisfying the requirements of Theorem 2 with $\gamma = 1$ without taking the other-regarding preferences into account, i.e., $g_i = 0$. One can think of standard VCG. Hence, we have $h_i(\boldsymbol{b}_{-i})$ such that for $p_i(v) = h_i(\boldsymbol{b}_{-i}) - D_{-i}(\boldsymbol{b}_{-i}, f(\boldsymbol{b}))$ the mechanism is non-deficit. Lemma 1 implies that

$$h_i(\boldsymbol{b}_{-i}) \geq \sup_{b'_i} \sum_j D_{-j}(\boldsymbol{b}'_{-j}, f(\boldsymbol{b}')) - \sum_{j \neq i} h_j(\boldsymbol{b}'_{-i}) \tag{7}$$

Let $g_i(\boldsymbol{b}_{-i}(f(\boldsymbol{b}))) = \alpha_i \sum_{j \neq i} b_j(f(\boldsymbol{b}))$. And let $g'_i(\boldsymbol{b}_{-i}(f(\boldsymbol{b}))) = \alpha \sum_{j \neq i} b_j(f(\boldsymbol{b}))$. First note that $g'_i(\boldsymbol{b}_{-i}(f(\boldsymbol{b}))) = \alpha \cdot D_{-i}(\boldsymbol{b}_{-i}, f(\boldsymbol{b}))$ and also observe that by choice of $\alpha$ it holds that $g'_i(\boldsymbol{b}_{-i}(f(\boldsymbol{b}))) \leq g_i(\boldsymbol{b}_{-i}(f(\boldsymbol{b})))$. This will allow us to invoke Proposition 3 below.

We manipulate equation (7) by multiplying both sides by $(1 - \alpha)$

$$\begin{aligned}(1-\alpha)h_i(\boldsymbol{b}_{-i}) &\geq \sup_{b'_i} \sum_j (1-\alpha) D_{-j}(\boldsymbol{b}'_{-j}, f(\boldsymbol{b}')) - \sum_{j \neq i} (1-\alpha) h_j(\boldsymbol{b}'_{-i}) \\ &= \sup_{b'_i} \sum_j D_{-j}(\boldsymbol{b}'_{-j}, f(\boldsymbol{b}')) - g'_j(\boldsymbol{b}'_{-j}(f(\boldsymbol{b}'))) - \sum_{j \neq i} (1-\alpha) h_j(\boldsymbol{b}'_{-i})\end{aligned}$$

Thus, substituting $(1 - \alpha)h_i$ by $h'_i$, shows that $h'_i$ satisfies Lemma 1 and so the mechanism with respect to $h'_i$ is non-deficit.

Because $g'_i \leq g_i$ Proposition 3 tells us that our mechanism is also IR and truthfulness follows from Theorem 2.

We take a look at the payments. First observe that

$$\begin{aligned}p'_i(\boldsymbol{b}) &= h'_i(\boldsymbol{b}_{-i}) + g'_j(\boldsymbol{b}_{-i}(f(\boldsymbol{b}))) - D_{-j}(\boldsymbol{b}_{-i}, f(\boldsymbol{b})) \\ &= (1-\alpha)h_i(\boldsymbol{b}_{-i}) - (1-\alpha)D_{-j}(\boldsymbol{b}_{-i}, f(\boldsymbol{b})) = (1-\alpha)p_i(\boldsymbol{b})\end{aligned}$$

And hence if $\alpha > 0$ every player pays or receives a factor $(1 - \alpha)$ less, and also the total payment $\sum_i p'_i(\boldsymbol{b}) = (1 - \alpha) \sum_{-i} p_i(\boldsymbol{b})$ is reduced by a factor $(1 - \alpha)$.

*Example 1 (Multi-Unit Auction).* Consider a multi-unit auction with $k$ identical goods. Each player can win at most 1 item.

Standard VCG without other-regarding preferences charges the highest losing bid to all players. We can capture this in the welfare-oriented model with the classical social welfare objective by setting $\alpha_i = 0$ for all $i \in N$ and $h_i(\boldsymbol{b}_{-i}) = \sum_{j=1}^{k} [\boldsymbol{b}_{-i}]_j$. Let $\alpha = \min_{i \in N} \alpha_i$ where $\alpha_i$ is the altruism level of player $i$. If we require NPT we can take $c_i = 0$ for all $i$ to minimize payments and obtain

$$h_i(\boldsymbol{b}_{-i}) = (1 - \alpha) \sum_{j=1}^{k} [\boldsymbol{b}_{-i}]_j$$

As also the second part of the payment (see Table 1) is multiplied by $(1 - \alpha)$ the total payments are a factor $(1 - \alpha)$ lower.

Letting go of NPT (but still requiring the mechanism to be non-deficit) allows us to do better. Already in the standard VCG setting we can apply the Bailey-Cavallo redistribution function [3,7]. This corresponds to $h_i(\boldsymbol{b}_{-i}) = \left(\sum_{j=1}^{k}[\boldsymbol{b}_{-i}]_j - \frac{k}{n}[\boldsymbol{b}_{-i}]_{k+1}\right)$, significantly reducing the sum of payments to $\sum_i \boldsymbol{p}_i(\boldsymbol{b}) = \sum_{j=1}^{k}[\boldsymbol{b}]_j - k \cdot [\boldsymbol{b}]_{k+1}$. Applying the theory from this subsection we know that we can take

$$h_i(\boldsymbol{b}_{-i}) = (1-\alpha)\left(\sum_{j=1}^{k}[\boldsymbol{b}_{-i}]_j - \frac{k}{n}[\boldsymbol{b}_{-i}]_{k+1}\right)$$

and still have a non-deficit, individually rational and truthful mechanism saving a factor $(1-\alpha)$ on the payments:

$$\sum_{i\in N}\boldsymbol{p}_i(\boldsymbol{b}) = (1-\alpha)\left(\sum_{j=1}^{k}[\boldsymbol{b}]_j - k \cdot [\boldsymbol{b}]_{k+1}\right)$$

It is even true that the last mechanism is individually undominated. Taking $b'_i = [\boldsymbol{b}_{-i}]_{k+1}$ in the supremum of equation (9) will result in equality. For example for $k=1$ and $i=1$ we have $\sum_j D_{-j}(\boldsymbol{b}_{-1}, [\boldsymbol{b}_{-1}]_2) = (1-\alpha)(n-1)[\boldsymbol{b}_{-1}]_1$ and $\sum_{j\neq 1} h_j(\boldsymbol{b}'_{-j}) = (1-\alpha)\left((n-2)[\boldsymbol{b}'_{-j}]_1 + \frac{[\boldsymbol{b}'_{-j}]_2}{n}\right)$. Note that $[\boldsymbol{b}_{-j}]_1 = [\boldsymbol{b}_{-1}]_1$ and $[\boldsymbol{b}'_{-j}]_2 = [\boldsymbol{b}_{-1}]_2$. Taking the difference is equal to $h_1(\boldsymbol{b}_{-1})$. This equality can be verified for all $k$ and $i$.

## References


1. J. Andreoni and J. Miller. Giving according to garp: An experimental test of the consistency of preferences for altruism. *Econometrica*, 70(2):737–753, 2002.
2. K. R. Apt and G. Schäfer. Selfishness level of strategic games. *Journal of Artificial Intelligence Research*, 49:207–240, 2014.
3. M. J. Bailey. The demand revealing process: to distribute the surplus. *Public Choice*, 91(2):107–126, 1997.
4. F. Brandt and G. Weiß. Antisocial agents and vickrey auctions. In *International Workshop on Agent Theories, Architectures, and Languages*, pages 335–347. Springer, 2001.
5. R. Buehler, Z. Goldman, D. Liben-Nowell, Y. Pei, J. Quadri, A. Sharp, S. Taggart, T. Wexler, and K. Woods. The price of civil society. In *International Workshop on Internet and Network Economics*, pages 375–382. Springer, 2011.
6. I. Caragiannis, C. Kaklamanis, P. Kanellopoulos, M. Kyropoulou, and E. Papaioannou. The impact of altruism on the efficiency of atomic congestion games. In *International Symposium on Trustworthy Global Computing*, pages 172–188. Springer, 2010.
7. R. Cavallo. Optimal decision-making with minimal waste: Strategyproof redistribution of vcg payments. In *Proceedings of the fifth international joint conference on Autonomous agents and multiagent systems*, pages 882–889, 2006.
8. R. Cavallo. Efficient auctions with altruism. *Published online at http://www. eecs. harvard. edu/cavallo/papers/cavallo-altru.pdf*, 2012.
9. G. Charness and M. Rabin. Understanding social preferences with simple tests. *The quarterly journal of economics*, 117(3):817–869, 2002.

10. P.-A. Chen, B. d. Keijzer, D. Kempe, and G. Schäfer. The robust price of anarchy of altruistic games. In *International Workshop on Internet and Network Economics*, pages 383–390. Springer, 2011.
11. E. H. Clarke. Multipart pricing of public goods. *Public choice*, pages 17–33, 1971.
12. G. De Marco and J. Morgan. Altruistic behavior and correlated equilibrium selection. *International Game Theory Review*, 13(04):363–381, 2011.
13. E. Fehr and U. Fischbacher. Why social preferences matter–the impact of non-selfish motives on competition, cooperation and incentives. *The economic journal*, 112(478):C1–C33, 2002.
14. E. Fehr and K. M. Schmidt. A theory of fairness, competition, and cooperation. *The quarterly journal of economics*, 114(3):817–868, 1999.
15. T. Groves. Incentives in teams. *Econometrica: Journal of the Econometric Society*, pages 617–631, 1973.
16. M. Guo and V. Conitzer. Worst-case optimal redistribution of vcg payments in multi-unit auctions. *Games and Economic Behavior*, 67(1):69–98, 2009.
17. M. Guo and V. Conitzer. Optimal-in-expectation redistribution mechanisms. *Artificial Intelligence*, 174(5-6):363–381, 2010.
18. M. Guo, V. Markakis, K. Apt, and V. Conitzer. Undominated groves mechanisms. *Journal of Artificial Intelligence Research*, 46:129–163, Jan. 2013.
19. D. Kahneman. *Thinking, fast and slow*. Macmillan, 2011.
20. M. Kozlovskaya and A. Nicolò. Public good provision mechanisms and reciprocity. *Journal of Economic Behavior & Organization*, 167:235–244, 2019.
21. S. Kucuksenel. Behavioral mechanism design. *Journal of Public Economic Theory*, 14(5):767–789, 2012.
22. A. Mas-Colell, M. D. Whinston, J. R. Green, et al. *Microeconomic theory*, volume 1. Oxford university press New York, 1995.
23. H. Moulin. Almost budget-balanced vcg mechanisms to assign multiple objects. *Journal of Economic theory*, 144(1):96–119, 2009.
24. R. B. Myerson and M. A. Satterthwaite. Efficient mechanisms for bilateral trading. *Journal of economic theory*, 29(2):265–281, 1983.
25. M. Rahn and G. Schäfer. Bounding the inefficiency of altruism through social contribution games. In *International Conference on Web and Internet Economics*, pages 391–404. Springer, 2013.
26. P. Tang and T. Sandholm. Optimal auctions for spiteful bidders. In *Twenty-Sixth AAAI Conference on Artificial Intelligence*, 2012.
27. W. Vickrey. Counterspeculation, auctions, and competitive sealed tenders. *The Journal of finance*, 16(1):8–37, 1961.

## A Proofs for Section 3 (Modeling Other-Regarding Preferences)

**Corollary 1.** *Making Assumption 1. Suppose we have a truthful mechanism with respect to some other-regarding preferences $g_i'$. Then this mechanism is also truthful with respect to $g_i$.*

*Proof.* For a truthful mechanism with respect to $g_i'$, Theorem 1 implies that there exists a $\mu_i'(f(\boldsymbol{b}), \boldsymbol{b}_{-i})$ such that

$$p_i(\boldsymbol{b}) - g_i'(\boldsymbol{b}_{-i}) = \mu_i'(f(\boldsymbol{b}), \boldsymbol{b}_{-i})$$

Let $\mu_i(f\boldsymbol{b}), \boldsymbol{b}_{-i}) = \mu_i'(f\boldsymbol{b}), \boldsymbol{b}_{-i}) - g_i(\boldsymbol{b}_{-i}) + g_i'(\boldsymbol{b}_{-i})$, then also

$$p_i(\boldsymbol{b}) - g_i(\boldsymbol{b}_{-i}) = \mu_i(f(\boldsymbol{b}), \boldsymbol{b}_{-i})$$

showing that we still have a truthful mechanism. □

**Theorem 2.** *Fix a design objective $D$. A mechanism $M = (f, \boldsymbol{p})$ is truthful in the utility model with other-regarding preferences if the following two conditions are satisfied:*

1. *$f(\boldsymbol{b}) \in \arg\max_{a \in A} D(\boldsymbol{b}, a)$.*
2. *For every player $i \in N$ there exist functions $h_i, \gamma_i : V_{-i} \to \mathbb{R}$ such that*

$$p_i(\boldsymbol{b}) = h_i(\boldsymbol{b}_{-i}) + g_i(\boldsymbol{b}_{-i}(f(\boldsymbol{b})), \boldsymbol{p}(\boldsymbol{b})) - \gamma_i(\boldsymbol{b}_{-i}) \cdot D_{-i}(\boldsymbol{b}_{-i}, f(\boldsymbol{b})).$$

*Proof.* If $p_i(\boldsymbol{b}) = h_i(\boldsymbol{b}_{-i}) + g_i(\boldsymbol{b}_{-i}(f(\boldsymbol{b})), \boldsymbol{p}(\boldsymbol{b})) - \gamma_i(\boldsymbol{b}_{-i}) \cdot D_{-i}(\boldsymbol{b}_{-i}, f(\boldsymbol{b}))$ then

$$p_i(\boldsymbol{b}) - g_i(\boldsymbol{b}_{-i}(f(\boldsymbol{b})), \boldsymbol{p}(\boldsymbol{b})) = h_i(\boldsymbol{b}_{-i}) - \gamma_i(\boldsymbol{b}_{-i}) \cdot D_{-i}(\boldsymbol{b}_{-i}, f(\boldsymbol{b}))$$

Let

$$\mu_i(f(\boldsymbol{b}), \boldsymbol{b}_{-i}) = h_i(\boldsymbol{b}_{-i}) - \gamma_i(\boldsymbol{b}_{-i}) \cdot D_{-i}(\boldsymbol{b}_{-i}, f(\boldsymbol{b}))$$

and we see that the conditions of Theorem 1 are satisfied. □

**Proposition 2.** *Let $M = (f, \boldsymbol{p})$ be a mechanism as defined in Theorem 2. Then $M$ satisfies NPT if and only if for every player $i \in N$, $h_i(\boldsymbol{b}_{-i}) \geq \gamma_i(\boldsymbol{b}_{-i}) \cdot D_{-i}(\boldsymbol{b}_{-i}, f(\boldsymbol{b})) - g_i(\boldsymbol{b}_{-i}(f(\boldsymbol{b})), \boldsymbol{p}(\boldsymbol{b}))$. Further, $M$ satisfies IR if and only if for every player $i \in N$, $h_i(\boldsymbol{b}_{-i}) \leq \gamma_i(\boldsymbol{b}_{-i}) \cdot D_{-i}(\boldsymbol{b}) + v_i(f(\boldsymbol{b}))$.*

*Proof.* Immediate from the definitions of IR ($u_i(\boldsymbol{b}) \geq v_i$) and NPT ($p_i(\boldsymbol{b}) \geq 0$). □

**Proposition 3.** *Making Assumption 1. Let $g_i, g_i'$ be a other-regarding preferences such that $g_i(\boldsymbol{b}_{-i}(f(\boldsymbol{b}))) \geq g_i'(\boldsymbol{b}_{-i}(f(\boldsymbol{b})))$ for all $\boldsymbol{b}$. Let $M = (f, \boldsymbol{p})$ be a mechanism as in Theorem 2 with respect to $g_i'$ for player $i$ that is individually rational. The mechanism with respect to $g_i$ is also truthful and individually rational.*

*Proof.* We have

$$u_i(\boldsymbol{b}) = v_i(f(\boldsymbol{b})) - p_i(\boldsymbol{b}) + g_i(\boldsymbol{b}_{-i}(f(\boldsymbol{b}))$$

By assumption $v_i(f(\boldsymbol{b})) - p_i(\boldsymbol{b}) + g_i'(\boldsymbol{b}_{-i}(f(\boldsymbol{b})) \geq 0$. Observe that

$$u_i(\boldsymbol{b}) = v_i(f(\boldsymbol{b})) - p_i(\boldsymbol{b}) + g_i(\boldsymbol{b}_{-i}(f(\boldsymbol{b})) \geq v_i(f(\boldsymbol{b})) - p_i(\boldsymbol{b}) + g_i'(\boldsymbol{b}_{-i}(f(\boldsymbol{b})) \geq 0$$

□

## B Proofs for Section 4 (Minimizing Payments)

**Lemma 1.** *A mechanism $M = (f, \boldsymbol{p})$ with $\boldsymbol{p}_i = h_i(\boldsymbol{b}_{-i}) + g_i(\boldsymbol{b}_{-i}(f(\boldsymbol{b}))) - \gamma_i(\boldsymbol{b}_{-i}) \cdot D_{-i}(\boldsymbol{b}_{-i}, f(\boldsymbol{b}))$ is non-deficit if and only if for all $i$ and $\boldsymbol{b}_{-i}$*

$$h_i(\boldsymbol{b}_{-i}) \geq \sup_{b'_i} \sum_j (\gamma_j(\boldsymbol{b}'_{-j}) \cdot D_{-j}((\boldsymbol{b}'_{-j}, f(\boldsymbol{b}'))) - g_j(\boldsymbol{b}'_{-j}(f(\boldsymbol{b}')))) - \sum_{j \neq i} h_j(\boldsymbol{b}'_{-j}) \quad (8)$$

*Proof.* For a mechanism to be non-deficit we need $\sum_i \boldsymbol{p}_i(\boldsymbol{b}) \geq 0$ for any $\boldsymbol{b}$. Fix some $\boldsymbol{b}$ and an $i$. Suppose a mechanism satisfies equation 8 then we know that

$$\begin{aligned} h_i(\boldsymbol{b}_{-i}) &\geq \sup_{b'_i} \sum_j \left(\gamma_j(\boldsymbol{b}'_{-j}) \cdot D_{-j}((\boldsymbol{b}'_{-j}, f(\boldsymbol{b}'))) - g_j(\boldsymbol{b}'_{-j}(f(\boldsymbol{b}')))\right) - \sum_{j \neq i} h_j(\boldsymbol{b}'_{-j}) \\ &\geq \sum_j \left(\gamma_j(\boldsymbol{b}_{-j}) \cdot D_{-j}(\boldsymbol{b}_{-j}, f(\boldsymbol{b})) - g_j(\boldsymbol{b}_{-j}(f(\boldsymbol{b})))\right) - \sum_{j \neq i} h_j(\boldsymbol{b}_{-j}) \end{aligned}$$

Rewriting yields

$$\sum_j h_j(\boldsymbol{b}_{-j}) + g_j(\boldsymbol{b}_{-j}(f(\boldsymbol{b}))) - \gamma_j(\boldsymbol{b}_{-j}) \cdot D_{-j}(\boldsymbol{b}_{-j}, f(\boldsymbol{b})) \geq 0$$

As the left hand side is exactly the sum of payments we have proved the if direction.

Suppose a mechanism is non-deficit, then $\sum_i \boldsymbol{p}_i(\boldsymbol{b}) \geq 0$ for all $\boldsymbol{b}$, thus also

$$\sum_j h_j(\boldsymbol{b}_{-j}) + g_j(\boldsymbol{b}_{-j}(f(\boldsymbol{b}))) - \gamma_j(\boldsymbol{b}_{-j}) \cdot D_{-j}(\boldsymbol{b}, f(\boldsymbol{b})) \geq 0$$

Fix some arbitrary $i$, we can rewrite

$$h_i(\boldsymbol{b}_{-i}) \geq \sum_j \gamma_j(\boldsymbol{b}_{-j}) \cdot D_{-j}(\boldsymbol{b}, f(\boldsymbol{b})) - g_j(\boldsymbol{b}_{-j}(f(\boldsymbol{b}))) - \sum_{j \neq i} h_j(\boldsymbol{b}_{-j})$$

as this holds for any $b_i$ this also holds for the the supremum. □

**Theorem 3.** *A mechanism $M = (f, \boldsymbol{p})$ with $\boldsymbol{p}_i = h_i(\boldsymbol{b}_{-i}) + g_i(\boldsymbol{b}_{-i}(f(\boldsymbol{b}))) - \gamma_i(\boldsymbol{b}_{-i}) \cdot D_{-i}(\boldsymbol{b}_{-i}, f(\boldsymbol{b}))$ is individually undominated if and only if for all $i$ and $\boldsymbol{b}_{-i}$*

$$h_i(\boldsymbol{b}_{-i}) = \sup_{b'_i} \sum_j \left(\gamma_j(\boldsymbol{b}'_{-j}) \cdot D_{-j}(\boldsymbol{b}'_{-j}(f(\boldsymbol{b}'))) - g_j(\boldsymbol{b}'_{-j}(f(\boldsymbol{b})))\right) - \sum_{j \neq i} h_j(\boldsymbol{b}'_{-j}) \quad (9)$$

*Proof.* For both implications we will prove the contrapositive. Since an individually undominated mechanism must be non-deficit we know by the previous lemma that the equality must be greater or equal. Suppose it is a strict inequality for some $i, \tilde{\boldsymbol{b}}_{-i}$, i.e., there exists a $\delta > 0$ such that

$$\begin{aligned} h_i(\tilde{\boldsymbol{b}}_{-i}) - \Big( \sup_{b'_i} \sum_j \left(\gamma_j(\tilde{\boldsymbol{b}}'_{-j}) \cdot D_{-j}(\tilde{\boldsymbol{b}}'_{-j}, f(\tilde{\boldsymbol{b}}')) - g_j(\tilde{\boldsymbol{b}}'_{-j}(f(\tilde{\boldsymbol{b}}')))\right) \\ - \sum_{j \neq i} h_j(\tilde{\boldsymbol{b}}'_{-j}) \Big) = \delta > 0 \end{aligned} \quad (10)$$

Define

$$h'_j(\boldsymbol{b}_{-j}) = \begin{cases} h_i(\tilde{\boldsymbol{b}}_{-i}) - \delta & \text{if } i = j \text{ and } \boldsymbol{b}_{-j} = \tilde{\boldsymbol{b}}_{-i} \\ h_j(\boldsymbol{b}_{-j}) & \text{otherwise} \end{cases} \tag{11}$$

and let

$$\boldsymbol{p}'_i(\boldsymbol{b}) = h'_i(\boldsymbol{b}_{-i}) + g_i(\boldsymbol{b}_{-i}(f(\boldsymbol{b}))) - \gamma_i(\boldsymbol{b}_{-i}) \cdot D_{-i}(\boldsymbol{b}_{-i}, f(\boldsymbol{b}))$$

If we can verify that $\boldsymbol{p}'$ is non-deficit then it individually dominates $\boldsymbol{p}$. This is clear because we have shifted $h'_i(\tilde{\boldsymbol{b}})$ exactly by the amount of slack there was, namely $\delta$, and the rest remained unchanged. As the mechanism with respect to $h$ was non-deficit, it is also with respect to $h'$.

For the other implication. Suppose $\boldsymbol{p}$ is individually dominated by $\boldsymbol{p}'$, then there exist $i$ and $\boldsymbol{b}$ such that $\boldsymbol{p}_i(\boldsymbol{b}) > \boldsymbol{p}'_i(\boldsymbol{b})$. In particular,

$$\begin{aligned} \boldsymbol{p}_i(\boldsymbol{b}) &= h_i(\boldsymbol{b}_{-i}) + g_i(\boldsymbol{b}_{-i}(f(\boldsymbol{b}))) - \gamma_i(\boldsymbol{b}_{-i}) \cdot D_{-i}(\boldsymbol{b}_{-i}, f(\boldsymbol{b})) \\ &> h'_i(\boldsymbol{b}_{-i}) + g_i(\boldsymbol{b}_{-i}(f(\boldsymbol{b}))) - \gamma_i(\boldsymbol{b}_{-i}) \cdot D_{-i}(\boldsymbol{b}_{-i}, f(\boldsymbol{b})) \\ &= \boldsymbol{p}'_i(\boldsymbol{b}) \end{aligned}$$

As $g_i$ and $\gamma_i(\boldsymbol{b}_{-i}) \cdot D_{-i}$ are the same on both sides of the inequality this actually implies $h_i(\boldsymbol{b}_{-i}) > h'_i(\boldsymbol{b}_{-i})$.

Write

$$\sigma = \sup_{b'_i} \sum_j \left(\gamma_j(\boldsymbol{b}_{-j}) \cdot D_{-j}((b'_i, \boldsymbol{b}_{-i}), f(b'_i, \boldsymbol{b}_{-i})) - g_j(\boldsymbol{b}_{-j}(f(b'_i, \boldsymbol{b}_{-i})))\right) - \sum_{j \neq i} h_j(\boldsymbol{b}'_{-j})$$

But then

$$h'_i(\boldsymbol{b}_{-i}) - \sigma > h'_i(\boldsymbol{b}_{-i}) - \sigma \geq 0$$

showing that (9) is not satisfied for $\boldsymbol{p}$.

□

# C Proofs for Section 5 (A case study: Altruism)

**Theorem 4.** *Every AAVCG mechanisms $M^{m,D} = (f, \boldsymbol{p})$ with respect to altruism model $m$ and design objective $D$ is truthful. Further, $M^{m,D}$ satisfies NPT and IR if it implements the altruism-adjusted Clarke pivot rule.*

*Proof.* We will check for every combination that it fits the design template from Theorem 2

$(w, sw)$: We verify

$$\begin{aligned} p_i(\boldsymbol{b}) &= h_i(\boldsymbol{b}_{-i}) - (1-\alpha_i)\sum_{j\neq i} b_j(f(\boldsymbol{b})) \\ &= h_i(\boldsymbol{b}_{-i}) + \alpha_i \sum_{j\neq i} b_j(f(\boldsymbol{b})) - \sum_{j\neq i} b_j(f(\boldsymbol{b})) \\ &= h_i(\boldsymbol{b}_{-i}) + g_i(\boldsymbol{b}_{-i}(f(\boldsymbol{b}))) - 1 \cdot D_{-i}(\boldsymbol{b}_{-i}, f(\boldsymbol{b})) \end{aligned}$$

$(o, sw)$: Here $g_i(\boldsymbol{b}_{-i}(f(\boldsymbol{b})), p_i(\boldsymbol{b})) = \alpha_i p_i(\boldsymbol{b}) + \alpha_i \sum_{j\neq i} b_j(f\boldsymbol{b})$. Hence for

$$\begin{aligned} p_i(\boldsymbol{b}) &= \frac{1}{1-\alpha_i} h_i(\boldsymbol{b}_{-i}) - \sum_{j\neq i} b_j(f(\boldsymbol{b})) \\ (1-\alpha_i) p_i(\boldsymbol{b}) &= h_i(\boldsymbol{b}_{-i}) - (1-\alpha_i)\sum_{j\neq i} b_j(f(\boldsymbol{b})) \\ p_i(\boldsymbol{b}) &= h_i(\boldsymbol{b}_{-i}) + \alpha_i p_i(\boldsymbol{b}) + \alpha_i \sum_{j\neq i} b_j(f(\boldsymbol{b})) - \sum_{j\neq i} b_j(f(\boldsymbol{b})) \\ p_i(\boldsymbol{b}) &= h_i(\boldsymbol{b}_{-i}) + g_i(\boldsymbol{b}_{-i}(f(\boldsymbol{b}))) - 1 \cdot \sum_{j\neq i} b_j(f(\boldsymbol{b})) \end{aligned}$$

it fits the criteria of Theorem 2.

$(w, dw)$: Take $\gamma_i(\boldsymbol{b}_{-i}) = \frac{1}{1+\sum_{k\neq i}\alpha_i}$. Then

$$\begin{aligned} p_i(\boldsymbol{b}) &= h_i(\boldsymbol{b}_{-i}) - \sum_{j\neq i} \left( \frac{1+\sum_{k\neq j}\alpha_k}{1+\sum_{k\neq i}\alpha_k} - \alpha_i \right) b_j(f(\boldsymbol{b})) \\ &= h_i(\boldsymbol{b}_{-i}) + \alpha_i \sum_{j\neq i} b_j(f(\boldsymbol{b})) - \sum_{j\neq i} \frac{1+\sum_{k\neq j}\alpha_k}{1+\sum_{k\neq i}\alpha_k} b_j(f(\boldsymbol{b})) \\ &= h_i(\boldsymbol{b}_{-i}) + \alpha_i \sum_{j\neq i} b_j(f(\boldsymbol{b})) - \frac{1}{1+\sum_{k\neq i}\alpha_k} \sum_{j\neq i} \left( 1 + \sum_{k\neq j}\alpha_k \right) b_j(f(\boldsymbol{b})) \\ &= h_i(\boldsymbol{b}_{-i}) + g_i(\boldsymbol{b}_{-i}(f(\boldsymbol{b}))) - \gamma_i(\boldsymbol{b}_{-i}) D_{-i}(\boldsymbol{b}_{-i}, f(\boldsymbol{b})) \end{aligned}$$

$(o, dw)$: For $\gamma_i(\boldsymbol{b}_{-i}) = \frac{1}{1+\sum_{k\neq i}\alpha_i}$

$$p_i(\boldsymbol{b}) = \frac{1}{1-\alpha_i} h_i(\boldsymbol{b}_{-i}) - \sum_{j\neq i} \left( \frac{1+\sum_{k\neq j}\alpha_k}{1+\sum_{k\neq i}\alpha_k} - \alpha_i \right) b_j(f(\boldsymbol{b}))$$

$$(1-\alpha_i)p_i(\boldsymbol{b}) = h_i(\boldsymbol{b}_{-i}) - (1-\alpha)\sum_{j\neq i} \left( \frac{1+\sum_{k\neq j}\alpha_k}{1+\sum_{k\neq i}\alpha_k} - \alpha_i \right) b_j(f(\boldsymbol{b}))$$

$$p_i(\boldsymbol{b}) = h_i(\boldsymbol{b}_{-i}) + \alpha_i p_i(\boldsymbol{b}) + \alpha_i \sum_{j\neq i} b_j(f(\boldsymbol{b})) - \sum_{j\neq i} \frac{1+\sum_{k\neq j}\alpha_k}{1+\sum_{k\neq i}\alpha_k} b_j(f(\boldsymbol{b}))$$

$$p_i(\boldsymbol{b}) = h_i(\boldsymbol{b}_{-i}) + \alpha_i \left( \sum_{j\neq i} b_j(f(\boldsymbol{b})) + p_i(\boldsymbol{b}) \right) - \frac{1}{1+\sum_{k\neq i}\alpha_k} \sum_{j\neq i} \left( 1 + \sum_{k\neq j} \alpha_k \right) b_j(f(\boldsymbol{b}))$$

$$p_i(\boldsymbol{b}) = h_i(\boldsymbol{b}_{-i}) + g_i(\boldsymbol{b}_{-i}(f(\boldsymbol{b}))) - \gamma_i(\boldsymbol{b}_{-i}) D_{-i}(\boldsymbol{b}_{-i}, f(\boldsymbol{b}))$$

□

# D Proofs for Section 6 (Impact of altruism)

**Proposition 4.** *Using the VCG mechanism with Clarke pivot rule, the public project in the standard utility model is funded if and only if*

1. $b_i \geq C$ *for some* $i \in N$ *and* $b_j = 0$ *for all* $j \in N$, $j \neq i$, *or*
2. $\sum_{i\in N} b_i = C$.

*Proof.* The payment of each player $i$ is $C - \sum_{j\neq i} b_j$ if $i$ is pivotal and 0 otherwise.

The if part for the first condition follows because in this case every player $i$ with non-zero $b_i$ is pivotal and thus

$$\sum_{i\in N} p_i(\boldsymbol{b}) = \sum_{i\in N} \left( C - \sum_{j\neq i} b_j \right) = nC - (n-1)\sum_{i\in N} b_i = C.$$

The if part of the second condition follows because the pivotal player $i$ trivially pays $C$.

We turn to the only-if part. Suppose that both conditions do not hold. Take any pivotal player $i$; note that without them the payments are $0 < C$ so for the project to be funded there must exist such a player. Player $i$ pays $C - \sum_{j\neq i} b_j$. In order to have the project funded, all other players must pay at least $\sum_{j\neq i} b_j$, which is larger than zero because the first condition does not hold. Thus, on average every player has to pay their own value. But only a pivotal player pays and their payment does not exceed their valuation because $C - \sum_{j\neq i} b_j = C - \sum_j b_j + b_i \leq b_i$. The only possibility to fund

the project is when all players with non-zero valuation pay exactly their valuation and are thus pivotal. By the above this happens only when $C - \sum_j b_j = 0$. But then the second condition holds, which is a contradiction. □

**Proposition 5.** *Let $N_p$ be the set of pivotal players, and $N_n$ the set of non-pivotal players. Using the VCG mechanism with the altruism-adjusted Clarke pivot rule (choosing $c_i = \alpha_i$ for all $i$), the public project in the omnistic model is funded if and only if*

$$\sum_{i \in N_p} \left( C - \sum_{j \neq i} b_j \right) + \sum_{i \in N_n} \frac{\alpha_i}{1 - \alpha_i} \left( \sum_{j \neq i} b_j - C \right) \geq C$$

*Proof.* The payment is given by

$$p_i(\boldsymbol{b}) = \frac{1}{1 - \alpha_i} \left( \sum_{j \neq i} v_j(a^{-i}) + v_0(a^{-i}) \right) - \sum_{j \neq i} v_j(f(\boldsymbol{b})) - v_0(f(\boldsymbol{b}))$$

If $i$ is a pivotal player and $f(\boldsymbol{b}) = yes$ then

$$p_i(\boldsymbol{b}) = 0 - \sum_{j \neq i} b_j + C = C - \sum_{j \neq i} b_j$$

while if $i$ is non-pivotal then

$$\begin{aligned} p_i(\boldsymbol{b}) &= \frac{1}{1 - \alpha_i} \left( \sum_{j \neq i} b_j - C \right) - \sum_{j \neq i} b_j + C \\ &= \frac{\alpha_i}{1 - \alpha_i} \left( \sum_{j \neq i} b_j - C \right) \end{aligned}$$

Hence the sum of the payments is

$$\sum_i p_i(\boldsymbol{b}) = \sum_{i \in N_p} \left( C - \sum_{j \neq i} b_j \right) + \sum_{i \in N_n} \frac{\alpha_i}{1 - \alpha_i} \left( \sum_{j \neq i} b_j - C \right)$$

And this has to be at least $C$ to have the project funded. □