Morally Monotonic Choice in Public Good Games

James C. Cox, Vjollca Sadiraj, and Susan Xu Tang

Experimental Economics, 26, 2023, 697 - 725

Morally monotonic choice in public good games

James C. Cox[1], Vjollca Sadiraj[2], and Susan Xu Tang[3]

Abstract. Rational choice theory, including models of social preferences, is challenged by decades of robust data from public good games. Provision of public goods, funded by lump-sum taxation, does *not* crowd out private provision on a one-for-one basis. Provision games elicit more of a public good than *payoff-equivalent* appropriation games. This paper offers a morally monotonic choice theory that incorporates *observable* moral reference points and is consistent with the two empirical findings. The model has idiosyncratic features that motivate a new experimental design. Data from our new experiment and three previous experiments favor moral monotonicity over alternative models including rational choice theory, prominent belief-based models of kindness, and popular reference-dependent models with loss aversion.

Keywords: public goods; experiment; payoff equivalence; non-binding contractions; rational choice; morally monotonic choice; belief-based kindness choice; reference-dependent choice with loss aversion

JEL Classification: C91, D03, H41

We are grateful to the editor and reviewers for comments and suggestions that motivated significant improvements in the paper. The National Science Foundation (grant number SES-1658743) and the Experimental Economics Center at Georgia State University provided research funding. The replication material for the study is archived at https://doi.org/10.5281/zenodo.7425943

1. jccox@gsu.edu. Experimental Economics Center and Department of Economics, Georgia State University, Atlanta, GA

2. vsadiraj@gsu.edu. Experimental Economics Center and Department of Economics, Georgia State University, Atlanta, GA

3. susan.tang@asu.edu. Department of Economics, Arizona State University, Tempe, AZ

1. Introduction

Theory and behavior for voluntary allocations of resources to public goods is a central topic in economics. This paper is motivated by two robust patterns of data from public good experiments that have been anomalous to theory for more than 25 years.

Much attention has focused on two payoff-equivalent game forms.[1] In a "provision game," n agents each start with an endowment e in their private account that they can provide to an equally-shared public account; each unit transferred to the public account is multiplied by $k < n$. In an "appropriation game," the n agents start with an endowment $E = nke$ in the shared public account that they can individually appropriate for their private accounts; each unit transferred to a private account is divided by k. These provision and appropriation games are "payoff equivalent" because any allocation to the public and private accounts that is feasible in one game is also feasible in the other.

Despite extensive research on modeling public good games,[2] two robust empirical findings continue to be anomalous to theory: (a) provision games elicit more of a public good than *payoff-equivalent* appropriation games[3] and (b) provision of a public good, funded by lump-sum taxation, does *not* crowd out voluntary provision on a *one-for-one* basis.[4]

Conventional rational choice theory,[5] including (unconditional) social preferences models, can account for limited free-riding behavior and other stylized facts in public good games (Ledyard 1995; Cox and Sadiraj 2007). But such models also predict: (i) *no* (provision vs. appropriation) game-form effect in payoff-equivalent games; and (ii) *one-for-one* crowding out of voluntary provision by tax-financed provision. Both predictions are inconsistent with robust data.

---

[1] Andreoni (1995) refers to these two game forms as "positively-framed" and "negatively-framed" public good games.
[2] For literature reviews, see the early work by Ledyard (1995) and the selective survey by Chaudhuri (2011). Our Online Appendix I.1 provides another selective survey.
[3] See, for examples: Andreoni (1995); Sonnemans, Schram, and Offerman (1998); Park (2000); Messer, et al. (2007); Fujimoto and Park (2010); Bougherara, Denant-Boemont, and Masclet (2011); Cubitt, Drouvelis, and Gächter (2011); Dufwenberg, Gächter, and Hennig-Schmidt (2011); Cox, et al. (2013); Fosgaard, Hansen, and Wengström (2014); Cox (2015); Cox and Stoddard (2015); Khadjavi and Lange (2015); and van Soest, Stoop, and Vyrastekova (2016).
[4] See, for examples: Abrams and Schmitz (1978, 1984); Clotfelter (1985); Kingma (1989); Andreoni (1993); Bolton and Katok (1998); Khanna and Sandler (2000); and Ribar and Wilhelm (2002). For incomplete crowding out with distortionary taxes see Bernheim (1986).
[5] By "conventional rational choice theory" we mean consequentialist rationality (e.g., Samuelson 1938; Chernoff 1954; Arrow 1959; Sen, 1971, 1986, 1993) and its prominent special cases including conventional preference theory (e.g., Hicks 1946; Samuelson 1947; Debreu 1959; textbooks), revealed preference theory (e.g., Afriat 1967; Varian 1982; textbooks) and (unconditional) social preferences models (e.g., Fehr and Schmidt 1999; Bolton and Ockenfels 2000; Andreoni and Miller, 2002; Cox and Sadiraj 2007).

What about popular alternative theories? A prominent belief-based model of kindness (Dufwenberg and Kirchsteiger 2004) predicts that provision of public goods, funded by lump-sum taxation, will crowd out voluntary provision on a *higher* than one-for-one basis, the opposite of the robust empirical pattern. Implications of reference-dependent models with loss aversion (e.g. Tversky and Kahneman 1991; Kőszegi and Rabin 2006) depend on specification of the reference point. If the reference point is conditional on others' allocations then the prediction is less public good in the provision than the appropriation game, the opposite of the robust empirical pattern.[6]

In this paper we discuss morally monotonic choice theory[7] that extends the fundamental axiom of rationality, the Consistency property (Arrow 1959), by advancing the idea of choice monotonicity with respect to *observable* moral reference points. For empirical validity, we use data from previous experimental studies (Andreoni 1995, Reuben and Riedl 2013, Khadjavi and Lange 2015) and a new experiment that includes payoff-equivalent game forms with provision or appropriation or both, and elicits first-order beliefs in addition to allocations. A central feature of the new experimental design is implementation of endogenous contractions of feasible sets that discriminate between morally monotonic choice and existing alternative theories. The contractions implemented in the new experiment take the simple form of non-binding lower bounds on permissible allocations to the public account.[8]

The paper makes several contributions. First, we offer a choice theory that agrees with conventional rational choice theory when the moral reference point is preserved but can monotonically diverge otherwise. We demonstrate existence of a reference-dependent choice function for this moral monotonicity theory that can be used in applied work. Second, we propose *observable* moral reference points that incorporate two features of the environment: (game) initial endowments and (set-conditional) minimal expectations payoffs (Roth 1977). Third, we show that moral monotonicity theory can explain the two robust features of data: provision vs. appropriation game-form effects; and less than one-for-one crowding-out by minimum required provision. Fourth, we use data from three previous experimental studies to inform on empirical validity of moral monotonicity of choices. Fifth, we report a new experiment with three game forms and non-binding contractions that discriminate between implications of morally monotonic choice and

[6] These and other implications of these models are explained in sections 2 and 3.
[7] See Cox, et al. (2019) for applications to dictator games (non-strategic environments).
[8] We define "non-binding" lower bound as one that is smaller than previously-observed allocations of all group members in the full game as well as their reported beliefs about the others' choices.

alternative theories including conventional rational choice theory, a prominent belief-based model (Dufwenberg and Kirchsteiger 2004), and popular reference-dependent models of loss aversion (Tversky and Kahneman 1991; Kőszegi and Rabin 2006).

2. Theories of Choice and Their Implications for Allocations

Theory of play in *n*-player strategic games depends on models of individual decision making. In this section, we present examples that highlight the divergent testable implications of alternative theoretical models. We illustrate the implications of conventional rationality (Example 1),[9] moral monotonicity (Examples 2.b and 3.b), belief-based model of kindness (Example 2.a), and reference-dependent model of loss aversion (Example 3.a) for allocations in a two-player public good game (as in our new experiment).[10]

It is useful for clarity in discussion to use the word "token" to refer to discrete units of the scarce resource, that can be transferred between private accounts and the public account, because this resource has different dollar values in the two types of accounts. It is also useful to distinguish between the "action space" in which tokens can be transferred and the "payoff space" which reports the (monetary) payoff implications of final token allocations to public and private accounts. Agents choose vectors of payoffs by allocating tokens between private accounts and the public account in the action space.[11]

2.1 Conventional Rational Choice Theory

We use standard terminology from rational choice theory. X denotes the finite set of all alternatives (vectors of payoffs in this paper), $c(\cdot)$ denotes a choice function that maps each set $S \subseteq X$ to a non-empty subset, that is, $c(S) \subseteq S$. The domain of the choice function consists of all nonempty finite subsets of X. A set $c(S)$ containing the chosen elements of S is called a "choice set" (Sen

---

[9] Popular models of (consequentialist) social preferences belong to conventional rational choice theory because they are characterized by utility functions defined over final (monetary) payoffs that provide complete and transitive orderings in the payoff space.

[10] The scarce resource in our experiment consists of 10 "tokens" that can be allocated between private and public accounts, with each token worth \$1 in the private account and \$1.5 in the public account that is equally shared by two players (i.e., the marginal per capita return is 0.75). In both appropriation and provision games, each player's final payoff is the sum of tokens in the private account and 0.75 times the total number of tokens in the public account.

[11] In the provision game, with endowments of 10 tokens in each private account, my *feasible* set, in (own payoff, other's payoff)-space, is $X = \{(10 - x + .75(5 + x), 10 - 5 + .75(5 + x)) \mid x \in \{0, \cdots, 10\}\}$ when the other player contributes (or so I believe) 5. If I contribute 7 to the public account then my choice set is the singleton $X^* = \{(12,14)\}$. Generally, an agent's choice set may not be a singleton, but we work with singleton choice sets in examples so the illustration of ideas is simple.

1971, p. 307). We follow the literature and do not require the choice sets to be singletons (Arrow 1959, p. 122).[12]

The fundamental axiom of rationality is the Consistency property. It is equivalent to existence of a complete and transitive order (Arrow, 1959; Sen 1971, 1986) for finite sets. Let T and S denote two nonempty finite feasible sets in $\mathbb{R}^n$. Let $T^* \subseteq T$ and $S^* \subseteq S$ denote an agent's (non-empty) choice sets. Rational choice theory (Arrow 1959) states that choice sets satisfy[13]

Consistency. For all nonempty finite sets $T \subseteq S$: $S^* \cap T \neq \varnothing \Rightarrow T^* = S^* \cap T$

In words, the Consistency property states that if some chosen element from the larger set, S is available in the smaller set, T then the choice set, T^* contains all elements of S^* that are available in T and no others.

Example 1 provides an illustration. Section 3.1 of the paper provides general implications of the Consistency property (in the payoff space) for behavior in public good games.

Example 1. Consider a two-player provision game, as in our experiment (described in more detail in section 5). Each player is endowed with a resource of 10 tokens in their private account. The value of each token in a player's private account is \$1. The value of each token in the public account is \$0.75 to each player.

Full game. In the full game, permissible provisions are from $\{0,1,\cdots,10\}$. Suppose that player 2 provides 8 (or player 1 believes so) to the public account. In this case, player 1's feasible set, S in the two-player payoff space is

$$S = \left\{\pi(g_1,8) = \left(10 - g_1 + .75g_1 + .75(8), 10 - 8 + .75g_1 + .75(8)\right) \mid g_1 \in \{0,1,\cdots,10\}\right\}.$$

When player 1 also provides 8 to the public account, the vector of payoffs is (14,14). Hence, her (assumed-to-be-singleton) choice set is $S^* = \{(14, 14)\}$.

Contraction game. In the contraction game, suppose the smallest permissible provision is 8; that is, each player must provide an amount to the public account selected from $\{8,9,10\}$. Now, when player 2 provides 8 (or player 1 believes so), player 1's feasible set in the payoff space contracts to

$$T = \left\{\pi(g_1,8) = \left(10 - g_1 + .75(g_1 + 8), 10 - 8 + .75(g_1 + 8)\right) : g_1 \in \{8,9,10\}\right\}$$

[12] Even with a stronger conventional assumption, such as GARP, choice sets are not generally singletons.
[13] See also Samuelson 1938; Chernoff 1954; Sen 1993.

which is a subset of *S* that contains (14,14); hence, $S^* \cap T \neq \emptyset$. What can we say about player 1's choice in *T*? Consistency requires, $T^* = S^* \cap T = \{(14,14)\}$. In this way, if player 2 provides 8 to the public account then player 1 provides 8 in the contraction game, just as in the full game. This example may be transparent; it is stated here to illustrate a direct application of the Consistency property (no need for any utility specification) and to provide a basis of comparison with Examples 2 and 3, where we look at alternative models.

2.2 Moral Monotonicity Theory

As in subsection 2.1, let *T* and *S* be two nonempty finite feasible sets in $\mathbb{R}^n$. Let t^z and s^z denote moral reference points for sets *T* and *S* from the perspective of agent *z*. Reference points may depend on features of the decision environment such as the status quo (Tversky and Kahneman 1991), disagreement point (Nash 1953), or disagreement and minimal expectations payoffs (Roth 1977), or maximal payoffs (Kalai and Smorodinsky 1975), or the equitable payoff (the average of the minimum and maximum payoffs) as in belief-based kindness (Rabin 1993; Dufwenberg and Kirchsteiger 2004). In section 3, we will propose reference points dependent on the initial endowed payoffs (at the beginning of the game) and (set-conditional) minimal expectation payoffs. In this section, we are agnostic about the specific identification of the reference point and allow them to have more dimensions than the number of players.[14] We do postulate how choices of payoffs respond to changes in reference points by proposing two basic properties: M-Consistency and M-Monotonicity.

An intuition about these properties can be conveyed by considering a special case of singleton choice sets, two agents, and two-dimensional reference points. Without loss of generality, consider agent 1. Let t^* and s^*, respectively, denote agent 1's choices from feasible sets T and S in $\mathbb{R}^2$. Suppose *T* is a subset of *S* that contains the choice s^* from *S*. Let t^1 and s^1 be the reference points from agent 1's perspective for sets *T* and *S*. M-Consistency has the same implications as Consistency when feasible sets *T* and *S* have the same reference point, $t^1 = s^1$.[15] Analogous statements hold for the reference points and choices of agent 2.

[14] For an example, see Table 1 in section 3.3.1, which reports (observable) four-dimensional reference points (two dimensions for each player).

[15] More generally, if the moral reference point has four dimensions (i.e., two dimensions for each player), then $t^1, s^1 \in \mathbb{R}^4$ and the expression is $(t^1_{11} - s^1_{11}) + (t^1_{12} - s^1_{12}) = 0 = (t^1_{21} - s^1_{21}) + (t^1_{22} - s^1_{22})$.

The implications of moral monotonicity theory diverge from conventional rational choice theory when the change between reference points favors one of the agents. We continue with a special case example. Let T and S contain the same elements: $T = S$. For the special case, M-Monotonicity states that, compared to the choice from S, agent 1's choice from T will favor agent 1 if the moral reference point in T favors agent 1, $t_1^1 - s_1^1 > t_2^1 - s_2^1$. In this way, choices are assumed to monotonically track reference points.

Writing the formal statements that define the theory will require some additional notation: $N = \{1, \cdots, n\}$ denotes the set of players; (X, x^z) denotes the "feasible problem" of agent z where $X \subset \mathbb{R}^n$ is a nonempty finite set in the payoff space, and the moral reference point is $x^z = \{x_i^z \in \mathbb{R}^m : i \in N\}$[16]; $X^*(x^z) \subseteq X$ denotes the nonempty choice set of agent z; $X_i^*(x^z) \subset \mathbb{R}$ denotes the set of payoffs that agent z's choice set $X^*(x^z) \subset \mathbb{R}^n$ allocates to agent $i \in N$; $\delta_i(t^z, s^z) = \sum_{j=1..m} (t_{ij}^z - s_{ij}^z), i \in N$ denotes agent i's total change between reference points t^z and s^z; $K = \{k \in N : \delta_k(t,s) = \max_{i \in N} \delta_i(t^z, s^z) > 0\}$, is the set of players most favored by t^z over s^z; $\rhd$ denotes a partial order of feasible sets on $\mathbb{R}$ defined as: for all $Z, Y \subset \mathbb{R}$, $Z \rhd Y$ means $\min(Z) \geq \min(Y)$ and $\max(Z) \geq \max(Y)$.

The M-Consistency property is a modification of the conventional Consistency property to incorporate reference points.

M-Consistency. For all feasible problems (T, t^z) and (S, s^z) such that $T \subseteq S$ and $\delta_i(t^z, s^z) = 0$, for all $i \in N$:

$$S^*(s^z) \cap T \neq \varnothing \Rightarrow T^*(t^z) = S^*(s^z) \cap T$$

In words, M-Consistency says that if the total change from reference point t^z to reference point s^z is 0 for every agent $i \in N$, then choice sets satisfy the conventional Consistency property. A special case includes sets having the same reference point, $t^z = s^z$.

The M-Monotonicity property postulates choice monotonic response to changes in reference point when the feasible set does not change.

---

[16] Here, the moral reference point has m dimensions per player. If it has only one dimension per player, as in the special case in the paragraph above, then $m = 1$.

M-Monotonicity. For all feasible problems (T,t^z) and (S,s^z) if $T = S$, $K \neq \varnothing$ and $\delta_{i \notin K}(t^z, s^z) = 0$ then there exists a nonempty set of agents $K^* \subseteq K$ such that

$$T_k^*(t^z) \rhd S_k^*(s^z) \text{ for all } k \in K^*;\ z \in K^* \text{ if } K = N$$

This property states that agent *z*'s choice set becomes more favorable for some of the agents who are favored by moral reference point t^z over s^z; meaning agent *z*'s smallest and largest payoffs chosen for such agents are both (weakly) larger in choice set $T^*(t^z)$ than in choice set $S^*(s^z)$. Furthermore, if the moral reference point becomes equally more favorable for all $i \in N$ (i.e., $K = N$) then agent *z*'s choice in *T* must (weakly) favor herself.

M-Consistency postulates what happens to payoff choices when the feasible set contracts while preserving the moral reference point whereas M-Monotonicity postulates what happens to choices when the moral reference point changes but the feasible set remains the same. The first inquiry relates to implications of moral monotonicity for choices when both the feasible set contracts and the moral reference point changes. Our first result (Proposition 1 below) provides an answer. A second immediate question concerns existence of a reference-dependent choice function that satisfies M-Consistency and M-Monotonicity (Proposition 2 below provides an answer.)

Scenario A. Consider an agent *z* who faces two feasible problems, (S,s^z) and (T,t^z) such that: (1) *T* is a subset of *S*, $T \subseteq S$; (2) the choice set, S^* of *S* and the feasible set *T* intersect, i.e., there exists $s^* \in S^*(s^z) \cap T$.

Proposition 1. Refer to Scenario A. If $\delta_{-k}(t^z, s^z) = 0$, and $\delta_k(t^z, s^z) \neq 0$ for some agent *k*, then there exists $t^* \in T^*(t^z)$ such that $(t_k^* - s_k^*)\delta_k(t^z, s^z) \geq 0$.

Proposition 1 (see Online Appendix II.1 for proof) says that the pattern of agent *z's* payoff choices follows the pattern of changes in moral reference point. So, if t^z is more favorable to agent *k* than s^z then agent *z* in problem (T,t^z) leaves agent *k* with larger payoff than *k* gets from s^*. The opposite happens if t^z is less favorable to agent *k*. A straightforward corollary for the special case of Pareto-efficient singleton choice sets and $n = 2$ is:

Corollary 1. Refer to Scenario A and suppose that choice sets are from $\mathbb{R}^2$, singleton and Pareto efficient. If $\delta_{-k}(t^k, s^k) < 0 < \delta_k(t^k, s^k)$ then $t^*_{-k} - s^*_{-k} \leq 0 \leq t^*_k - s^*_k$.

The corollary states that if the change in moral reference point favors *k* (one of the two players) and disfavors the other, then agent *z* will choose a larger payoff for agent *k* (and a smaller one for the other).

Our next result, on existence of a choice function is stated in Proposition 2 (see Online Appendix II.1 for proof).

Proposition 2. There exists a reference-dependent choice function that satisfies M-Consistency and M-Monotonicity.

The proof uses choice function, $U(\pi \mid r)$ defined over payoff vectors, $\pi \in \mathbb{R}^n$ for a given reference point, $r \in (\mathbb{R}^m)^n$, written (without any loss of generality) for agent 1 as

$$U(\pi \mid r) = \sum_{k=1}^{n} w_k(r) u(\pi_k) \text{ with weights } w_k(r) = \theta(\sigma_k \sum_{j=1}^{m} r_{kj}) / \sum_{i=1}^{n} \theta(\sigma_i \sum_{j=1}^{m} r_{ij}), \ \sigma_1 > 1 = \sigma_{k>1},$$

for some increasing concave function, $u(\cdot)$ and increasing function, $\theta(\cdot)$ such that $\theta(y+z) = \theta(y)\theta(z)$. An idiosyncratic feature of the choice function is that the reference points are in the $w_k(\cdot)$ weights rather than the $u(\cdot)$ values.[17]

2.3 Examples Illustrating Differences Between Three Alternatives to Rational Choice Theory

The following two examples offer a preview of general implications of alternative models to conventional rationality by comparing implications of moral monotonicity to implications of a belief-based model of kindness and a model of reference-dependence with loss aversion. In these examples, we apply moral monotonicity using reference points *from the other models* so as to make clear that differences between models do not come solely from different definitions of reference points.[18]

[17] As an illustration for applied research, in Online Appendix I.4, we apply a parametric special case of $U(\cdot)$ − where the weights $w_k(\cdot)$ are normalized natural exponential functions of the moral reference point − to data from two previous experiments (Andreoni 1995; Khadjavi and Lange 2015) and data from a new experiment reported herein.

[18] We will explain our specification of moral reference point in Section 3.

Example 2. *Contraction and two alternative models to conventional rationality*. We refer to the Contraction game in Example 1 and ask: what are the implications of the belief-based model of kindness or moral monotonicity? To answer these questions, we need to know how contraction affects the reference point. We here use equitable payoff as the reference point (Dufwenberg and Kirschteiger 2004) and illustrate that, for contraction, the prediction of the belief-based model of kindness is the opposite of the prediction of moral monotonicity.

a. *Belief-based model of kindness*. In our linear public good game (Example 1), the kindest action a player can adopt is to allocate 10 to the public account and the least kind is to allocate the minimum permissible level, c. The reference point, the average maximum and minimum payoffs, is reached at allocation of 5 in the full game (c=0) and 9 in the contraction game ($c = 8$). If player 1 believes player 2 allocates 8 (first order-belief), then player 2 is perceived to be kind in the full game (8 > 5) but unkind in the contraction game (8 < 9), and therefore player 1 allocates less in the contraction game than in the full game.[19]

b. *Moral Monotonicity*. Use the same definition of reference point and apply the choice function in Proposition 2. At player 2's contribution of 8, the reference point − the average of the maximum and minimum payoffs − in sets S and T (as in Example 1) are: $s = (14.75, 11.75)$ corresponding to player 1's allocation of 5 and $t = (13.75, 14.75)$ corresponding to player 1's allocation of 9. Compared to s, reference point t is more favorable to player 2 (but not to player 1), so player 1 would leave player 2 with a larger payoff in the game with contraction, which player 1 can do by contributing more than 8 (her contribution in the full game),[20] which is the opposite prediction of the belief-based model of kindness.

---

[19] Example 2.a is very similar to one suggested by an anonymous reviewer. Let b_2 and c_1 denote player 1's first- and second-order beliefs. For the Dufwenberg and Kirschteiger (2004) model, player 1's utility in the full game is $\pi_1(g_1, b_2) + Y\gamma^2(g_1 - 5)(b_2 - 5)$, and $\pi_1(g_1, b_2) + Y\gamma^2(g_1 - 9)(b_2 - 9)$ in the contraction game with low bound of 8, where $\gamma = .75$ and g_1 is player 1's contribution. A player 1 with sensitivity parameter, $Y > 0.15$, first-order belief, $b_2 = 8$ (and any second order belief) is a full contributor ($g_1 = 10$) in the full game but a "free rider" ($g_1 = 8$) in the contraction game.

[20] At optimal interior choice, π^s in problem (S, s), $u'(\pi_1^s)/u'(\pi_2^s) = (\theta_2^s/\theta_1^s)/(1/\gamma - 1) < (\theta_2^t/\theta_1^t)/(1/\gamma - 1)$ where the equality follows from optimality of π^s and the inequality follows from t favoring 2 and disfavoring 1 compared to s. Hence, π_2^s is too small to be optimal in problem, (T, t), which implies that player 1's allocation must be larger, $g_1^t(g_2) > g_1^s(g_2)$.

Example 3. *Game form effect and two alternative models to conventional rationality*. What are implications of reference-dependence with loss aversion (Tversky and Kahneman 1991, Koszegi and Rabin 2006) and moral monotonicity for allocations in provision and appropriation games? Again, to answer the question we first need to specify the reference point. In this example, let the reference point be the "status quo" payoffs when no player alters the initial position (that is, contributes nothing in the provision game and appropriates nothing in the appropriation game.) Hence, the reference points are $s^p = (10,10)$ in the provision game and $s^a = (15,15)$ in the appropriation game. Using this definition of reference point in both models, we illustrate how the two models' predictions differ.

a. *Reference dependence with loss aversion*. Assume that player 1's preferences are defined over final payoffs (utility of consumption) and changes from the reference point (gain/loss utility). When player 2 allocates 8 to the public account (i.e., provides 8 in the provision game, or appropriates 2 in the appropriation game), the feasible set of player 1 is *S* (as in Example 1) in both games. In the provision game, all player 1's allocations larger than 2 result in payoffs in the gain domain (i.e., both players' payoffs are larger than 10). Suppose that player 1 allocates 3 in the provision game, resulting in payoffs (15.25,10.25). In the appropriation game, compared to (15,15), allocation of 3 (i.e., appropriation of 7), however, comes with a meager gain of 0.25 in own payoff dimension and a loss 19 times as high (4.75) in other's payoff dimension. Allocation 4 results in payoffs (15,11), and a loss averse player 1 would prefer it over (15.25,10.25); that is, she would prefer to give up 0.25 in gain to reduce the loss in the other's dimension by three times as much (0.75). Hence, when player 2 allocates 8, our player 1's allocation in the appropriation game is larger than 3, the allocation in the provision game.[21]

b. *Moral Monotonicity*. Assume that player 1's choice sets satisfy M-Consistency and M-Monotonicity. Comparing $s^p = (10,10)$ and $s^a = (15,15)$, we see that both players are favored by the reference point in the appropriation game, that is, they both belong to set *K*. So by M-

[21]Assume additively separable preferences, with utility of consumption $v(\pi) = \sum u_i(\pi_i)$, and for each dimension *i*, gain/loss utility $(\pi_i - r_i)$ for gains, and $\lambda(\pi_i - r_i)$ for losses (with $\lambda > 1$). A player 1 that allocates 3 in the provision game, reveals (*) $(v_1+1)(\gamma-1)+(v_2+1)\gamma = 0.25(-v_1+3v_2+2)=0.$ In the appropriation game, her marginal utility at allocation 3, is $(v_1+1)(\gamma-1)+(v_2+\lambda)\gamma = 0.25(-v_1+3v_2+3\lambda-1) = 0.25(3\lambda-3) > 0$ where the second equality follows form (*) and the inequality follows from loss aversion, $\lambda > 1$. Hence, player 1's allocation in the appropriation game must be larger than in the provision game.

Monotonicity, player 1 leaves herself with a larger payoff in the appropriation game, which player 1 can do by allocating less to the public account in the appropriation game.[22] Hence, when player 2 allocates 8, moral monotonicity predicts that the appropriation game will elicit *smaller* allocation from player 1 than the provision game, which is the opposite of the prediction of the reference-dependence model with loss aversion.

3. Play in Provision, Appropriation, and Mixed Games

Here we derive implications for (best response) *allocations* of tokens to the public account in two-player, payoff-equivalent provision, appropriation and mixed games with and without restrictions on chosen allocations. In a linear public good game, each player, i chooses how much, g_i of an amount W of a scarce resource to allocate to a public account shared with others. Let $\gamma \in (1/n, 1)$ denote the marginal per capita rate of return from the public account. When the vector of others' allocations to the public account is g_{-i}, player i's money payoff, π_i is the sum of returns from the private account, $w_i = W - g_i$ and the public account, $\gamma(g_i + G_{-i})$ where $G_{-i} = \sum_{j \neq i} g_j$. The initial *per capita* endowment of tokens, $g^e \in [0, W]$ in the public account uniquely identifies the g^e-game. Special cases include: provision game $(g^e = 0)$, where a public good can be provided; appropriation game $(g^e = W)$, where a public good can be appropriated; and mixed games $(g^e \in (0, W))$, where both provision and appropriation of a public good are feasible.

The discussion will be informal and for two-player games and, without any loss of generality, we focus on player 1.[23] Player 2's allocation of tokens, g_2 to the public account determines the feasible set, $S(g_2)$ of player 1 (in the money payoff space). For a concrete illustration, consider the two-player public good game in our experiment (described in section 5)

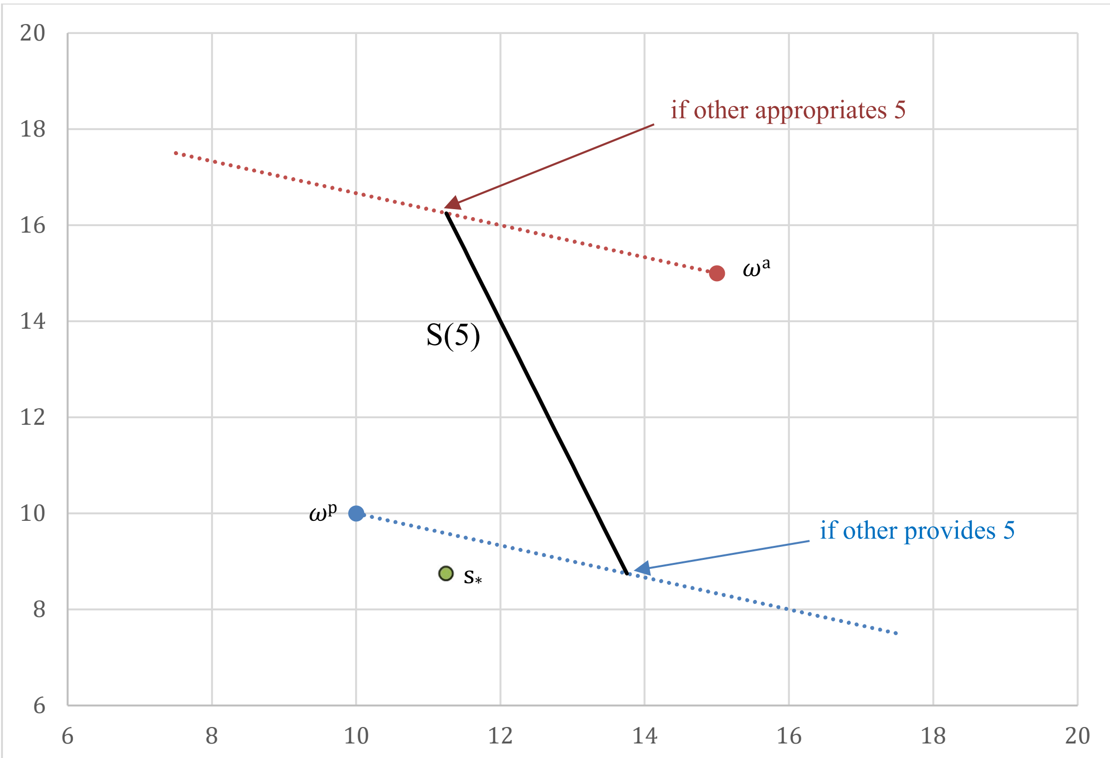
---

[22] Note that (*) $\theta_2^a / \theta_1^a = \theta(15)/\theta(\sigma 15) = \left(\theta_2^p / \theta_1^p\right)\left(\theta(5)/\theta(\sigma 5)\right) < \theta_2^p / \theta_1^p$ for the choice function below Proposition 2. At optimal, π^p in the provision game, $u'(\pi_1^p)/u'(\pi_2^p) = (\theta_2^p / \theta_1^p)/(1/\gamma - 1) > (\theta_2^a / \theta_1^a)/(1/\gamma - 1)$, where the equality follows from optimality of π^p and the inequality follows from (*). So, the appropriation game requires a smaller left hand side ratio, hence, player 1's allocation must be smaller, $g_1^a(g_2) < g_1^p(g_2)$.

[23] Formal arguments for n-player games for conventional rational choice and morally monotonic choice are reported in Online Appendix I.2.

with parameterization $W = 10$ and $\gamma = 0.75$. If player 2 allocates 5 tokens to the public account, player 1's feasible set in payoff space, $S(5)$ consists of discrete points on the solid line in Figure 1.

Figure 1. Player 1's Feasible Set in Payoff Space



Notes: Player 1 and player 2 payoffs are on the horizontal and vertical axes in a two-player public good game with mpcr of 0.75 and per capita token endowment $W = 10$. In the provision game, the token endowment is 0 in the public account and 10 in each private account, so each player's endowed payoff is 10, shown by ω^{p}. In the appropriation game, the token endowment is 20 in the public account and 0 in each private account, so each player's endowed payoff is 15, as shown by ω^{a}. Discrete points on the dotted "lines" correspond to player 2's possible allocations in provision (lower dotted "line") and appropriation (upper dotted "line") games. If player 2 "appropriates 5" in appropriation, or "provides 5" in provision, then player 1's feasible set in the payoff space is (a set of discrete points on) the solid line, $S(5)$.

Note that set $S(5)$ is the same whether the initial per capita allocation, g^{e} is 10 (as in the appropriation game) or 0 (as in the provision game). Suppose player 1's choice in $S(5)$ is (13,11), that is 13 for herself and 11 for the other. Player 1 can implement it by allocating, $g^{*}_{1}(5) = 3$ to the public account, which corresponds to providing 3 in the provision $(g^{e} = 0)$ game or appropriating 7 in the appropriation $(g^{e} = 10)$ game or appropriating 5 in the mixed $(g^{e} = 8)$ game. Thus, if player 1 cares only about monetary payoff for player 2 and herself then she allocates 3 to the public account in any of these g^e-games to realize the desired (13,11) payoffs. This takes us to our first generalized observation in section 3.1.

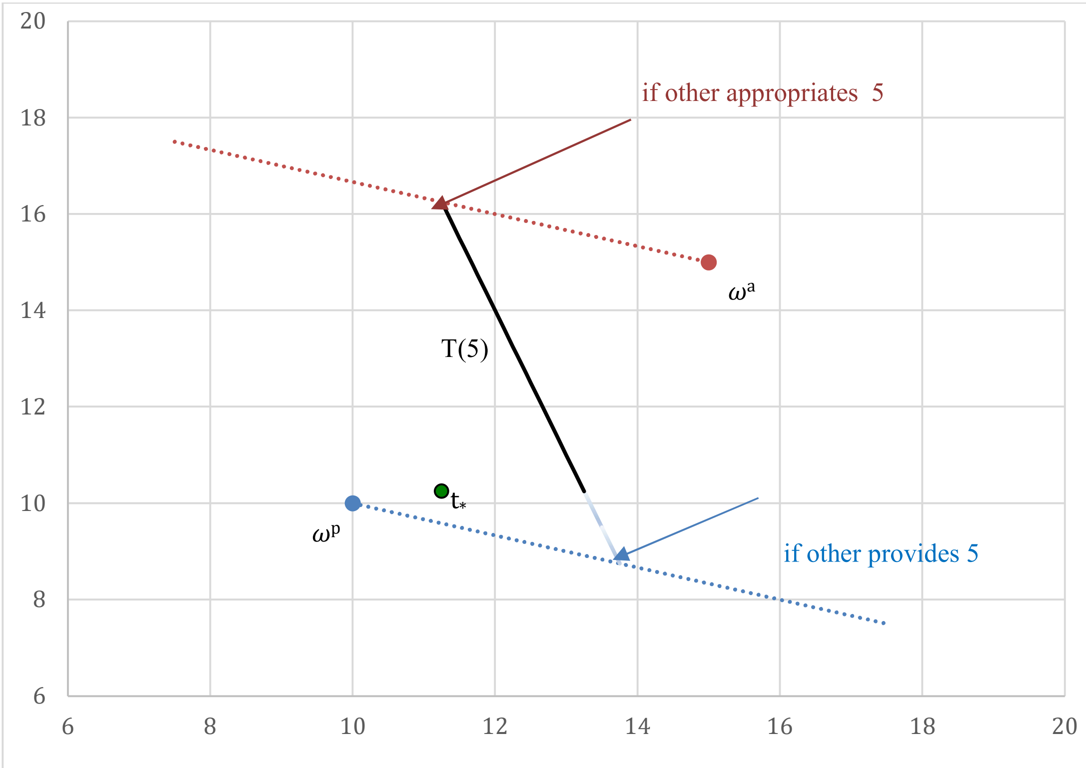
3.1 Implications of Conventional Rational Choice Theory

The first observation is that the Consistency property in the payoff space requires that player *i*'s (best response) allocations are not affected by initial (the endowed per capita) allocation, g^e in the public account (see Online Appendix II.2) because the feasible set in the payoff space, $S(g_{-i})$ remains the same for all g^e. This illustrates why, for any given vector of allocations of others, g_{-i}, player *i*'s set of best response allocations, $W^*(g_{-i} \mid g^e)$ is the same regardless of whether the per capita initial allocation to the public account, g^e is 0 (provision game) or W (appropriation game) or some amount in $(0,W)$ in mixed games.

A second observation (see Online Appendix II.2) is that the Consistency property in the payoff space implies that player *i*'s set of best response allocations, $W^*(g_{-i} \mid g^e)$ remains the same if, instead of $B = \{0,\cdots,W\}$, player *i* faces some subset,[24] C that contains all (best response) allocations of both players in the full game. We call these C subsets "nonbinding contractions." In our example, a low bound of 2 on permissible allocations to the public account is a non-binding contraction C because it contains the other's allocation 5 as well as player 1's allocation 3 (her best response to 5). In the payoff space, the feasible set, $T(5)$ for this (contraction) game is shown in Figure 2. If player 2 allocates 5 to the public account in the full game, player 1's feasible set (in the payoff space) is $S(5)$. In the contracted game, allocations are constrained to C={2,3,…,10}, mapping to $T(5)$ in the payoff space, which is a subset of $S(5)$ that contains the payoffs (13,11). Consistency then requires (13,11) to be in the choice set, $T^*(5)$. The implication in terms of allocations to the public account is $C^*(5 \mid g^e) = W^*(5 \mid g^e)$.

[24] In a provision game, a required minimum contribution, $c > 0$, produces a contraction. Government contribution to a public good financed by lump sum taxation is one way of implementing such a contraction. In an appropriation game, a contraction corresponds to a quota on maximum extraction, $t > 0$. The two types of contractions are payoff equivalent when $c = W - t$.

Figure 2. **Player 1's Contracted Feasible Set in Payoff Space**



Notes: Notation is the same as in the note below Figure 1. When a maximum appropriation of 8 in the appropriation game or a minimum provision of 2 in the provision game is introduced, and the other's allocation is 5, then the feasible set includes discrete points on the solid line *T*(5), which is a subset of *S*(5) in Figure 1.

Implications of the two observations are summarized in the following proposition (see Online Appendix II.2 for a formal derivation).[25]

Proposition 3. If choice sets in the payoff space satisfy the Consistency property then for any given g^e-game and vector of others' allocations g_{-i}:

a. $C^*(g_{-i} \mid g^e) = W^*(g_{-i} \mid g^e)$, for all nonbinding[26] contractions C

---

[25] Note that if the lower bound, *c* is binding then by construction individual allocations are weakly increasing in *c*. For example, *c*=2 (as in our illustration) is binding for player 1 if she chooses to allocate 1 in the public account in the full game but cannot do so in the contraction game.

[26] Here $C = \{c, \cdots, W\}, c > 0$ is non-binding if $0 < c < \min(g_i^b, \min(g_{-i}))$ where g_i^b is the smallest best response allocation of player *i* in the full game (i.e., *c*=0).

b. $W^*(g_{-i} \mid g^e) = W^*(g_{-i} \mid 0)$, for all initial allocations $g^e \in [0, W]$

Proposition 3 says that, for any given vector of allocations to the public account by others, agent *i*'s (best response) allocations to the public account are invariant to: (a) non-binding contractions; and (b) provision, appropriation, or mixed game form.

A straightforward implication, is that if g^* is an equilibrium in the provision game, it is also an equilibrium in all g^e-games as well as for all non-binding contracted games. Two much-studied theoretical properties are among the applications of Proposition 3. A (non-binding) contraction can be implemented by imposition of a lump sum tax in amount $\tau = c$ and use of the tax revenue to finance the public good. Part a implies invariance of the *total* allocation to the public good: voluntary allocations in amount τ are crowded out one-for-one by this public policy. Among interpretations of Part b is another much-studied theoretical property of invariance of allocations to game form (provision or appropriation or mixed) in payoff-equivalent games. Summarizing:

Corollary 2. Conventional rational choice theory implies:

a. One-for-one crowding out of voluntary provision by (nonbinding) lump-sum-tax-financed provision of a public good;

b. Equal allocations to a public good in provision, appropriation, and mixed games that are payoff equivalent.

3.2. Implications of Models of Belief-based Kindness and Reference Dependence with Loss Aversion

Reference dependent models build on the assumption that material payoff is not the only motivator of individual choice. Payoffs exceeding (or falling short) of equitable payoffs matter in the belief-based kindness model whereas gains or losses from a reference point matter in the models of loss aversion. We here look at predictions of these two alternative models for behavior in our two-player linear public good games.

3.2.1 Belief-based Kindness

Player 1's utility (Dufwenberg and Kirchsteiger 2004) at allocation, g_1 and first- and second-order beliefs, $g_2^{\sim}$ and $g_1^{\approx}$, is[27]

[27] We have dropped subscripts on Y, κ and λ to simplify notation.

$$U(g_1, g_2^{\sim}, g_1^{\approx}) = u(\pi_1(g_1, g_2^{\sim})) + Y\kappa(g_1, g_2^{\sim})\lambda(g_2^{\sim}, g_1^{\approx})$$

where: $\pi_1 = W - g_1 + \gamma(g_1 + g_2^{\sim})$; the utility of own material payoff, $u(\cdot)$ is linear; $Y > 0$ is the individual's sensitivity parameter; (un)kindness of player 1 towards player 2 is $\kappa(g_1, g_2^{\sim}) = \pi_2(g_2^{\sim}, g_1) - \pi_2^e(g_2^{\sim})$; and player 1's belief about player 2's (un)kindness towards her is $\lambda(g_2^{\sim}, g_1^{\approx}) = \pi_1(g_1^{\approx}, g_2^{\sim}) - \pi_1^e(g_1^{\approx})$, where $\pi_i^e(\cdot)$ is player *i*'s equitable payoff. It is straightforward to verify the following statements (see Online Appendix II.3).

Observations. For our provision and appropriation games:

1. The equitable payoff, $\pi^e(\cdot)$ corresponds to allocation, $g^c = (W + c)/2$
2. Second-order beliefs are irrelevant for kindness functions as well as for utility
3. For linear $u(\cdot)$, optimal allocations are either the minimum permissible allocation or, depending on the sensitivity parameter, switch to the maximum permissible level at some threshold level of first-order belief larger than $(W + c)/2$. Contraction has a positive effect on the threshold, hence a negative effect on allocations to the public account.
4. For nonlinear $u(\cdot)$ of material payoff, if $g_2^{\sim} > (W + c)/2$ then optimal (interior) allocation to the public account increases in (first-order belief about) other's allocation and decreases in minimum permissible allocation, *c*.

The intuition for the first observation is that, for any beliefs: (a) the largest allocation, W yields the maximum payoff; and (b) the smallest allocation, which is the least permissible level *c*, results in the minimum payoff. Together with linearity of monetary payoffs in allocations, they imply that equitable payoff (while differing across beliefs) is always reached at allocation $(W + c)/2$. So, an allocation larger than $(W + c)/2$ is as kind in the provision game as it is in the appropriation game. Thus, for any given first-order belief, the best-response allocation is invariant across game forms. For non-binding contractions, a low bound on permissible allocations increases the threshold, $(W + c)/2$ for allocations to be considered "kind", which thereby has a predicted negative effect of contractions on allocations to the public account. For example, in the full game (in our experiment) the average maximum and minimum payoffs correspond to allocating 5. In a contraction game, with 4 as a low bound on allocations, the average maximum and minimum payoffs correspond to allocating 7. Player 2 allocating 6 is kind in the full game but it is unkind in the contraction game. So, a player 1 who is motivated by "kindness" of player 2, allocates more than 4 (as contraction is not binding) in the full game but she goes for 4, the minimum permissible allocation, in the contraction game.

3.2.2 Reference Dependence with Loss Aversion Choice

Following the literature (Tversky and Kahneman 1991; Kőszegi and Rabin 2006) we assume additive separability across dimensions and a linear gain-loss utility specification. The implications are derived (see Online Appendix II.4 for details) for two potential reference points of set $S(g_{-i})$: (1) " g_{-i} conditional status quo" − the vector of payoffs conditional on other players' allocations when player *i* contemplates her choice; and (2) "unconditional status quo" − the vector of initial payoffs at the beginning of the game (before any player makes a choice).

Specification 1: g_{-1} conditional reference point. The reference points are $r^p = (W + \gamma g_2, W - g_2 + \gamma g_2)$ and $r^a = (\gamma W + \gamma g_2, W - g_2 + \gamma W + \gamma g_2)$ in the provision and appropriation games, respectively. In Figure 1, the reference point of $S(5)$ in the provision game is the most southeast point, (13.75, 8.75) whereas in the appropriation game it is the most northwest point, (11.25, 16.25). If (13,11) (i.e., allocate 3 to the public account) maximizes utility of consumption, then the choice in the appropriation game will be northwest of (13,11) (i.e., allocate more than 3) because r^a is northwest whereas in the provision game it will be southeast of (13,11) (i.e., allocate less than 3) because r^p is southeast of (13,11). This is also true generally (see Online Appendix II.4). An implication of reference dependent choice with loss aversion is smaller (best-response) allocations to the public account in provision than appropriation game.

In case of non-binding floors, at any given g_2 player 1's reference point in the appropriation game corresponds to allocation *W*, and therefore it is not affected by non-binding contractions; so the prediction is no contraction effect on play in an appropriation game. In the provision game, depending on internalization of the floor, the reference point either remains r^p, (and so there is no effect on allocations) or it moves northwest, towards $r^p + (\gamma - 1, \gamma)c$. For a visualization, in Figure 1 the reference points in the full and contraction provision game are the most southeast points in *S*(5) and *T*(5). For non-binding contractions, the chosen point in the full game is in the loss-gain domain (i.e., first player loss, second player gain) with respect to either reference point. This, together with the gain-loss utility being additively separable, imply that the chosen point in the contraction game is the same as in the full game. The prediction of the reference dependent model with loss aversion is that non-binding lower bounds have no effect on (best-response) allocations.

Specification 2: initial endowed payoffs reference point. Suppose, instead, that the reference points are initial endowed payoffs, (W, W) in the provision game and $(2\gamma W, 2\gamma W)$ in the appropriation game.[28] The non-binding contractions have no effect on either reference point, so a null effect on best-response allocations is predicted. The game-form effect, however, is ambiguous (see Online Appendix II.4).

Summarizing, we have the following implications for reference dependent models with loss aversion (see Online Appendix II.4):

a. Non-binding contractions have no effect on allocations.
b. Depending on the reference point, either the appropriation game elicits larger allocations to the public account than the provision game or the effect is ambiguous.

3.3 Implications of Moral Monotonicity

We first identify moral reference points and, subsequently, use M-Consistency and M-Monotonicity to derive implications of moral monotonicity.

3.3.1 Moral Reference Points in g^e-games

It seems promising for the moral reference point to incorporate two intuitions into theory of choice: my ethical constraints on interacting with others may depend on (a) endowed (or initial) payoffs (a.k.a. "property rights") and (b) the payoffs one can receive when the other's payoff is maximized (a.k.a. "minimal expectation payoffs"). Monotonicity in (a) captures the intuition that larger endowed payoffs entitle one to larger payoffs when interacting with others. Monotonicity in (b) captures the intuition that sense of entitlement is dependent on the environment, not an absolute that is independent of the environment. If my payoff y_F^* associated with you getting your maximum in environment *F* is larger than my payoff y_G^* associated with you getting your maximum in environment *G*, then I feel more entitled to claiming $x \geq y_F^*$ in *F* than in *G*.

[28] In Figure 1, $r^p = (10,10)$ and $r^a = (15,15)$.

We first focus on the moral reference point from the perspective of player 1.[29] Any given allocation, g_2 to the public account by player 2 determines player 1's feasible set (in payoff space). In a g^e-game, the initial distribution of a scarce resource, W is g^e in the public account and $W-g^e$ in the private account, so the initial endowed payoff of each player is

(1) $\omega^e = W - g^e + \gamma(g^e + g^e)$

The minimal expectation payoff, 2_* of player 2 is when player 1 free-rides at the full extent the game allows, and allocates the minimum required amount, c to the public account:

(2) $2_* = W - g_2 + \gamma(g_2 + c)$

So, from the perspective of player 1, the moral reference point with respect to player 2 is the ordered pair from (1) and (2):

(3) $r_2^1 = (\omega^e, 2_*) = (W - g^e + 2\gamma g^e, W - g_2 + \gamma g_2 + \gamma c)$

The minimal expectation payoff, 1_* of player 1 is when player 1 allocates all her W units of resource to the public account:

(4) $1_* = \gamma(W + g_2)$

So, from the perspective of player 1, the moral reference point with respect to oneself is

(5) $r_1^1 = (\omega^e, 1_*) = (W - g^e + 2\gamma g^e, \gamma W + \gamma g_2)$

The moral reference point from the perspective of player 2 follows immediately from symmetry. Table 1 shows moral reference points *from the perspectives of both players* in the special case of a two-player g^e-game with contraction $c \geq 0$. It is important to note that everything in Table 1 is observable from the experimental design.

---

[29] Separate detailed explanations of moral reference points in provision, appropriation, and mixed games with contraction ($c > 0$) or without contraction ($c = 0$) can be found in Online Appendices I.2 and II.5.A.

Table 1. Moral Reference Points in a Two-Player g^e Game with Contraction, $c \geq 0$

	Player 1 Perspective at $T(g_2)$	Player 2 Perspective at $T(g_1)$
Player 1 Dimensions	$(\omega^e, \gamma W + \gamma g_2)$	$(\omega^e, W - g_1 + \gamma g_1 + \gamma c)$
Player 2 Dimensions	$(\omega^e, W - g_2 + \gamma g_2 + \gamma c)$	$(\omega^e, \gamma W + \gamma g_1)$

Notes: W is total amount of resource; γ is the mpcr; g_i is player i's allocation to the public account; $c = 0$ when there is no contraction; $\omega^e = W - g^e + 2\gamma g^e$ is the endowed payoff in the two-player g^e- game for any $g^e \in [0, W]$.

Example. Figures 1 and 2 show initial endowed payoffs for provision and appropriation games (labeled, respectively, as ω^p and ω^a) and the locations of minimal expectations payoffs (labeled s_* and t_*). Table 2 reports the moral reference points *from the perspective of player 1* for the $g_2 = 5$, $W = 10$, and $\gamma = 0.75$ example in Figures 1 and 2. The first coordinates are initial endowed payoffs whereas the second coordinates correspond to minimal expectation payoffs. In the left column (provision game), the initial endowed payoff of each player is 10 whereas the minimal expectation payoffs are: 11.25 for player 1 (when she adopts the most generous action of contributing 10), and 8.75 for player 2 (when player 1 adopts the most greedy action of contributing 0).[30] In the middle column (appropriation game), the initial endowed payoff of each player is 15,[31] whereas the minimal expectation payoffs (second coordinates) are 11.25 (player 1) and 8.75 (player 2). The right column shows minimal expectation payoffs for the provision game when the smallest permissible allocation is 2 and the feasible payoff set is $T(5)$, shown in Figure 2. Here, the most selfish allocation is to allocate 2 (rather than 0), so the other's minimal

Table 2. Illustration of Moral Reference Points (from the perspective of player 1)

Game	Provision (Fig.1) $S(g_2 = 5 \mid g^e = 0, c = 0)$	Appropriation (Fig.1) $S(g_2 = 5 \mid g^e = 10, c = 0)$	Prov. with Min. of 2 (Fig.2) $T(g_2 = 5 \mid g^e = 10, c = 2)$
Player 1 Dim.	(10, 11.25)	(15, 11.25)	(10, 11.25)
Player 2 Dim.	(10, 8.75)	(15, 8.75)	(10, 10.25)

[30] 11.25 = (10 -10) + 0.75(5+10) and 8.75 = (10-5)+ 0.75(5+0).
[31] 15 = 0 + 0.75(10+10)

expectation payoff is 10.25 (up from 8.75 absent contraction) but remains 11.25 for player 1 (as the most generous allocation is still 10). Moral monotonicity (Proposition 4 below) requires that choices follow the same pattern: larger payoff for player 2 in contraction than in full provision game, which player 1 can achieve by increasing her allocation to the public account.

3.3.2 Implications of Moral Monotonicity for (Best Response) Allocations

M-Consistency and M-Monotonicity properties hold for choice sets in the *payoff* space. What are the implications of such properties for (best response) allocations of *tokens* to the public account?

Contraction Effect. Refer to the first column in Table 1 to verify that, from the perspective of player 1, the player 1 dimensions of the moral reference point are invariant to a lower bound, c on allocations to the public account. The player 2 dimension (in column 1) depends on a lower bound, c because that affects player 1's most selfish allocation, and therefore player 2's minimal expectation payoff (the second coordinate). Hence, for any given two distinct lower bounds, $c_1 > c_2 \geq 0$, the total change in player 1 dimensions is zero whereas the total change in player 2 dimensions is $\gamma(c_1 - c_2)$.[32] Compared to a no-contraction game ($c_2 = 0$), in games with contraction ($c_1 > 0$) player 2 is in the set, K of "favored" players but player 1 is not, and moral monotonicity (Proposition 1) requires that player 1 leave player 2 with a larger payoff in a (non-binding) contraction game, which she can do by increasing her allocation to the public account. Therefore, contrary to crowding out, moral monotonicity implies that a (nonbinding) floor on allocations to the public account has a positive effect on a player's best-response allocations.

Initial Endowment Effect. Similarly, from Table 1 and the definition of $\omega^e = W - g^e + 2\gamma g^e$, we observe that the dimensions of the moral reference points (for both players) depend on the per-capita initial endowment of the public account, g^e. From the perspective of player 1, the total change in *both* player 1 and player 2 dimensions of the moral reference point when g^e increases from g^t to $g^s > g^t$ is $(2\gamma - 1)(g^s - g^t)$. Therefore, the set, K of most favored players includes both players, and by M-Monotonicity, player 1 aims at a larger final payoff for herself in the game with the larger g^e, so she reduces her allocation to the public account.

---

[32] The definition of total change, δ appears in the notation paragraph just above the definition of M-Consistency.

These findings are summarized in Proposition 4 (see Online Appendix II.5.B for proof).

Proposition 4. If choice sets in the payoff space satisfy M-Consistency and M-Monotonicity properties then for every g^e-game and vector of others' allocations g_{-i}:

a. $C^*(g_{-i} \mid g^e) \triangleright W^*(g_{-i} \mid g^e)$ for all nonbinding[33] contractions C

b. $W^*(g_{-i} \mid g^t) \triangleright W^*(g_{-i} \mid g^s)$ for all $g^s > g^t$ from $[0, W]$

An implication from part a is that a nonbinding increase in minimum contribution to the public account financed with lump sum taxes will increase the total public good level. Part a is consistent with the intuition that allocating c units of the scarce resource in the public account can be a cooperative action in the full game but it is a free-riding action in the contracted game, and a "morally conscious" player would refrain from free-riding. Part b of Proposition 4 says that, compared to mixed games, extreme (best response) allocations to the public account are larger in the provision game and smaller in the appropriation game. It is consistent with the intuition that the larger the payoff at the beginning of the game, the larger is the expected payoff from playing the game, which translates to lower allocations to the public account (as the individual's payoff decrease in own allocation). In this way we have:

Corollary 3. Moral monotonicity implies:

a. Incomplete crowding out of voluntary provision by (nonbinding) lump-sum-tax-imposed provision of a public good;
b. Higher allocation to the public good in a provision game than in a payoff-equivalent appropriation game with mixed game allocation in between.

3.3.3 Extreme Nash Equilibria

Implications of Propositions 3 and 4 for effects of nonbinding contractions and initial (per capita) endowment of the public account on extreme Nash equilibria when allocations are strategic complements are summarized in Proposition 5 (see Online Appendix II.6 for proof).

[33] Recall that $C = \{c, \cdots, W\}, c > 0$ is non-binding if $0 < c < \min(g_i^b, \min(g_{-i}))$ where g_i^b is the smallest best response allocation of player i in the full game (i.e., c=0).

Proposition 5. For increasing best response, extreme (the largest and the smallest) Nash equilibrium allocations:

a. Do not vary with c and g^e for conventional rational choice theory;

b. Increase in c and decrease in g^e for morally monotonic choice theory.

3.4 *Summary of implications of different models*

These theoretical models have the following predicted effects for experimental treatments.

I. Predictions for Non-binding Contractions. The effect of a (non-binding) floor on (best response) allocations in the public account is predicted to be:

a. Negative or Null: by belief-based model of kindness

b. Null: by conventional rational choice theory and reference-dependence with loss aversion

c. Positive: by morally monotonic choice theory

II. Predictions for Game Form. The effect of shifting endowment from the private to the public account on (best response) allocations to the public account is predicted to be:

a. Positive (or ambiguous): by reference-dependence with loss aversion

b. Null: by conventional rational choice theory and belief-based model of kindness

c. Negative: by morally monotonic choice theory

We observe that while conventional rational choice theory predicts no game-form effect and no (nonbinding) contraction effect, the three alternative theoretical models predict non-zero effects but there is no general agreement in their predictions. What does the data tell us about empirical validity of these predictions?

4. Testing Alternative Theoretical Models with Existing Data

We use data from experiments reported by Andreoni (1995), Khadjavi and Lange (2015), and Reuben and Riedl (2013) to test hypotheses I.a – I.c and II.a – II.c for alternative theoretical models. We chose the Andreoni (1995) paper because it is the seminal paper for the large literature on effects of (provision vs. appropriation) game form reviewed in Online Appendix I.1. We use the Khadjavi and Lange (2015) data because (to the best of our knowledge) it is the first to introduce mixed public good games and contractions (albeit exogenous). We use the Reuben and Riedl (2013) data because it has provision games with an *upper* bound on provisions .[34]

The three experiments all include 10 rounds of play, pay subjects their total earnings from all rounds and, after each round, provide subjects with information on their own payoff and the total allocation to the public account. Andreoni (1995) uses a strangers design, with groups of 5 players, whereas partners designs are used in Khadjavi and Lange (2015), with groups of 4 players, and in Reuben and Riedl (2013), with groups of 3 players. Andreoni (1995) uses evocative subject instructions that highlight positive externalities in the provision game and negative externalities in the appropriation game. Khadjavi and Lange (2015) uses neutral wording in subject instructions. Andreoni (1995) includes payoff-equivalent provision and appropriation games, Khadjavi and Lange (2015) adds a payoff-equivalent mixed game in which subjects can make transfers in both directions between the public account and their private accounts. Khadjavi and Lange (2015) also includes a treatment with exogenous contraction in a mixed game that places a lower bound on individual allocations to the public account. Reuben and Riedl's (2013) treatment URE includes an exogenous contraction in the provision game that restricts provision by the high-endowment player *from above*.[35]

Table 3 reports (random-effects) GLS regression for the last five rounds of data on individual allocations in each of these three experiments. Explanatory variables include: $(G_{-i})_{t-1}$, the total allocation by others to the public account in the previous period; g^e, the per capita *endowment* (of tokens) in the public account; and dummies for contractions.

A central theoretical prediction about game form can be tested with data from the Andreoni and Khadjavi and Lange experiments. The value of g^e determines whether the game form is provision (g^e =0) or appropriation (g^e =60), in Andreoni's experiment, and whether the game form is provision (g^e =0) or mixed (g^e =8) or appropriation (g^e=20) in Khadjavi and Lange's experiment. The significantly negative estimates of the coefficient for g^e are consistent with prediction II.c for morally monotonic choice theory.[36] These negative estimates are inconsistent with predictions II.a

---

[34] We thank the editor for suggesting the Reuben and Riedl (2013) paper.

[35] In both URE and UUE treatments, the game is a provision game with mpcr of 0.5, and in each group of three players, one player is endowed with 40 tokens whereas the other two are endowed with 20 tokens. The high-endowment player can contribute all 40 tokens in UUE but only up to 20 in URE. The low-endowment players can contribute all 20 tokens in both UUE and URE treatments.

[36] These estimated effects are for "token" endowments. The implied dollar amounts are economically significant. For example, the -0.11 token coefficient for Andreoni's data corresponds to -6.6 (=0.11*60) tokens when endowment of

and II.b for models of reference dependence with loss aversion, conventional rational choice theory (including consequentialist models of social preferences), and belief-based model of kindness.

Table 3. Individual Allocations to Public Account in Previous Experiments (rounds 6 to 10)

Dep. Var: g_i Allocation	**Andreoni Data** {0...60}	**Khadjavi and Lange Data** {0…20} & C={8…20}		**Reuben and Riedl Data** Low: {0...20} High: {0…40} & C={0…20}		
Group size	5	4		3		
Range of G_{-i}	{0...240}	{0...60}	{24...60}	High {0...40}	Low {0...60}	{0...40}
	(1)	(2)	(3)[a]	(4)	(5)	(6)[b]
$(G_{-i})_{t-1}$	0.05*** (0.019)	0.13*** (0.022)	0.09*** (0.034)	0.37*** (0.080)	0.22*** (0.029)	0.28*** (0.029)
g^e	-0.11** (0.047)	-0.09* (0.054)	-0.24** (0.100)			
(D) Contraction: Contrib. Floor		4.67*** (0.830)	3.71*** (0.951)			
(D) Contraction: Contrib. Ceiling				-7.77*** (2.070)	1.32 (1.031)	0.04 (1.182)
Constant	11.53*** (2.598)	3.39*** (0.984)	6.77*** (2.159)	4.38** (2.057)	1.19 (1.121)	1.17 (1.035)
Punishment Opp.	no	no	no	yes	yes	yes
R-Squared (overall)	0.065	0.368	0.151	0.651	0.635	0.647
Nr of Subjects	80	160	95	35	70	57
Observations	400	800	345	175	350	269

Notes: Random-effects GLS regression. Estimates shown in columns (2) and (4) are based on all data, whereas in columns (3) and (6), we use only data from the full game when [a] $(G_{-i})_{t-1} \geq 24$ and [b] $(G_{-i})_{t-1} \leq 40$. Robust standard errors in parentheses; clustered at group level for data in K&L and R&R. Feasible set of allocations in braces. *** p<0.01, ** p<0.05, * p<0.1

A second central question is how subjects' allocations respond to non-binding contractions. Andreoni's experimental design does not include contractions. The Khadjavi and Lange experiment and the Reuben and Riedl experiment do include contractions. The significantly positive estimates of coefficients for a floor on allocations (in the Khadjavi and Lange experiment) are consistent with prediction I.c for morally monotonic choice theory but inconsistent with

the public account is changed from 0 to 60 tokens. So, with n = 5 and mpcr = 0.5, the payoff from the public account decreases by $16.50.

predictions I.a and I.b for the other theoretical models. The URE treatment in the Reuben and Riedl experiment puts a ceiling on provisions by the high endowment subjects. The significantly negative coefficient in column (4) for high players is consistent with the prediction from morally monotonic choice theory, as are the insignificant coefficients in columns (5) and (6) for low players.[37]

There are several limitations in using data from these studies. First, the findings for effects of these studies' *exogenous* contractions of feasible sets could result, mechanically, from floors (or ceilings) being binding rather than from validated or contradicted predictions from alternative theoretical models. Secondly, the subjects' decisions could be motivated by reciprocity, in particular in Khadjavi and Lange, and Reuben and Riedl experiments, that use a partners design. Thirdly, using previous period total allocation as a proxy for individuals' beliefs, while reasonable, can be arguable. These limitations motivated our new experiment. In the new experiment, subjects' first-order beliefs are elicited. The elicited beliefs, and subjects' allocations in a previous round, are used in imposition of *endogenous*, non-binding contractions. The new experiment limits the number of decision rounds to three, uses a strangers design without feedback between rounds, and pays one randomly-selected round.

5. New Experimental Design with Endogenous Contractions

We design a two-player experiment with provision, appropriation and mixed games.[38] We observe individuals' chosen allocations in the full game (baseline) and elicit subjects' guesses about others' allocations, and use them to inform nonbinding contractions of feasible sets that exclude only alternatives that have not previously been chosen nor believed in being chosen by subjects matched in a subsequent play of a contracted game. This design provides sharp discrimination between implications of alternative models. We cross contractions with provision or appropriation game forms. In addition, we have treatments for mixed games that allow both provision and

[37] For the high-endowment player, the effect of a non-binding upper bound on own contribution in URE is a larger own minimal expected payoff than in UUE. So, for any given contribution of the low-endowment player, moral monotonicity predicts the high type's best response is lower in URE than in UUE. For the low-endowment player, the non-binding upper bound on high's contribution has no effect on low's moral reference point, so moral monotonicity predicts that for any given contribution of the high-endowment player, low's best response is the same in the two treatments.

[38] The experiment was approved by the Institutional Review Board at Georgia State University.

appropriation. In all treatments, the game is between two players and the public account marginal per capita rate of return is 0.75. Table 4 shows parameter configurations.

Table 4. Experimental Design and Treatments

	Contracted Provision	**Provision**	**Mixed Games**			**Approp.**	**Contracted Approp.**
Initial Endowed Payoff	\$10	\$10	\$11	\$12.5	\$14	\$15	\$15
Initial Tokens in Private Account	10	10	8	5	2	0	0
Action Set	[c, 10]	[0, 10]	[-2, 8]	[-5, 5]	[-8, 2]	[-10, 0]	[-t, 0]
Feasible Allocations[a] in Public Account	[c, 10][b]	[0, 10]	[0, 10]	[0, 10]	[0, 10]	[0, 10]	[10-t, 10][c]
Design	Within Subjects		Within Subjects			Within Subjects	
Subjects: Order[d]	40: BCB 40: CBC		72: random order of 8,5,2			40: BCB 40: CBC	
Decisions per Subject	3		3			3	
Nr. of Subjects	80		72			80	
Observations	240		216			240	

Note: [a] Feasible allocations include discrete amounts in the intervals. [b] $c = \min_i \{g_i^* - 1, guess(g_{-i})\}$.
[c] $t = \max_i \{t_i^* + 1, guess(t_{-i})\}$. [d] B = $\{0, \cdots, 10\}$. C is $\{c, \cdots, 10\}$ in provision and $\{-t, \cdots, 0\}$ in appropriation.

The decision task consists of allocating W=10 tokens between the private and public accounts. Different subjects participated in the provision game, mixed game and appropriation game treatments. Subjects who participated in the mixed games faced tasks in g^e = 2, 5, and 8 games in random order. Each subject made three decisions without feedback on others' allocations and was randomly and anonymously paired with a different other subject in each of the three decision tasks. After making each decision, each subject was also asked to report own expectation ("guess") about the other's decision; correct guesses were paid \$2 but incorrect guesses were not paid. One of the three decisions was randomly selected for payment at the end of each experiment session, which yielded average subject salient payoff of \$15.71. After all allocations and guesses had been entered, subjects were asked to complete a questionnaire (included in Online Appendix II.8). In addition to demographic questions, it contained questions about a subject's altruistic activities and about their opinions of the altruism vs. selfishness of others. Sessions lasted about one and one-half hours including time for reading instructions, making decisions, answering the questionnaire, and receiving payment. There were 36 or 40 subjects in each session.

Provision or appropriation games are implemented (within-subjects) with and without contractions. In a baseline (B) game, the set for tokens that can be allocated to the public account includes integers in [0,10]. In a contraction (C) game, the set of tokens that can be allocated to the public account includes integers in $[c,10]$ for some $c \geq 0$, chosen to be "nonbinding," as explained below. To control for order effects, half of the subjects participated in the BCB treatment order and the other half in the CBC order. For each pair of subjects who faced the contraction set $[c,10]$ in treatment C after the larger set [0,10] in treatment B, the contraction set contained the observed allocations and guesses of both players in the previous baseline treatment.[39] To control for "corner set" effects and/or one-sided errors, the minimum allocation, c was 1 less than the smallest allocation within a pair of subjects.[40] For example, if the allocations of a pair of subjects in the provision game were 3 and 5 and the reported guesses were 4 and 3 then the set of allocations for the pair in the provision game with contraction was $\{2,\ldots,10\}$.

The construction of contractions in the appropriation game treatment was guided by the same logic. As an illustration, for a pair of subjects with appropriations 2 and 6 in the appropriation game and the reported guesses 4 and 3, the contracted set for transfers from the

the public account to the private account would be $\{0,1,\cdots,7\}$.[41]

6. Empirical Play in the New Experiment

As reported in Table 4, seventy-two subjects participated in the mixed-game treatment with each subject making three decisions.[42] In addition, we have data from eighty other subjects who made three decisions in provision games, with and without contraction, and another eighty subjects who

---

[39] In a CBC session, the contraction sets used in the first C task are the same as in a preceding BCB session.

[40] Exceptions to the "$1 less" criterion are when observed allocations in the preceding task are at a corner amount of 0 or close to 10. In a BCB session, if either subject guessed 0 or allocated 0 to the public account in the first B task then the set in treatment C would be integers from [0,10]. If application of the "$1 less" criterion would have resulted in a set with fewer than three options (i.e., lower bound 8 or 9) the set of allocations for task C was $\{5, 6,\ldots,10\}$.

[41] In terms of the number of tokens allowed to be allocated to the public account this set is $\{3,4,\cdots,10\}$.

[42] One decision in each of the 2-game, 5-game, and 8-game; the order of the tasks was randomized across subjects.

made three decisions in appropriation games with and without contraction.[43] Table 5 reports results from regression analysis of data from our experiment using a model specification similar to the one used to analyze data from previous experiments (reported in Table 3). The estimated coefficients for initial endowed tokens in the public account (g^e) and non-binding restrictions on minimum allocations to the public account (the low bound c) provide tests of the predictions for alternative theoretical models summarized statements I.a – I.c and II.a – II.c in section 3.4. The negative estimates of the coefficient for g^e are consistent with prediction II.c for morally monotonic choice theory but inconsistent with predictions II.b and II.a for conventional rational choice theory (including consequentialist models of social preferences), belief-based model of kindness, and models of reference dependence with loss aversion. Similarly, the positive estimates of the coefficient for c are consistent with prediction I.c for morally monotonic choice theory but inconsistent with predictions I.a and I.b for the other theoretical models.

Table 5. Individual Allocations to Public Account in Our Experiment

Dep. Variable: g_i Allocation	**(1)**	**(2)**
Guessed Other's allocation	0.62***	0.61***
	(0.047)	(0.045)
g^e	-0.05*	-0.07**
	(0.030)	(0.031)
c	0.36***	0.37***
	(0.054)	(0.051)
Constant	1.37***	1.56***
	(0.245)	(0.366)
Demographics	no	yes
R-Squared (overall)	0.435	0.458
Subjects	232	232
Observations	696	696

Notes: Linear estimators with standard errors clustered at subject level. Demographics include dummies for Female, Black, Self Image (give to a stranger, give to charity, help others with homework, share secrets) and Other's Image (disabled car assistance, selfish, dislike helping others). Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

[43] The experiment was not pre-registered. All of the data from the experiment we conducted are used in the regression reported in Table 5; we collected no other unreported data.

7. Conclusion

This paper was motivated by two robust patterns of data from public good experiments that were anomalous to existing theory for more than 25 years. The patterns are: (a) games with provision elicit larger amounts of a public good than do *payoff-equivalent* games with appropriation; and (b) games with *non-binding* floors elicit larger amounts of a public good than do games without contractions. Robust pattern (a) is exhibited by experiments reported in Andreoni (1995) and other papers cited in footnote 3 and included in the selective literature survey in Online Appendix I.1. Robust pattern (b) is reported in papers included in footnote 4.

Since the robust data patterns were anomalies, a research priority was development of a theoretical model. A second priority was testing predictions of the new model, and existing alternative models, with data from a selection of experiments in the literature. A third research priority was challenging the new theoretical model with experimental tests of its idiosyncratic predictions. This paper pursued all three research priorities.

We extend the Consistency property that characterizes rational choice theory (Arrow 1959) to incorporate reference points and postulate choice monotonicity to reference-points with the M-Consistency and M-Monotonicity properties. These two properties could be applied with various specifications of reference points. Indeed, in section 2 we use definitions of reference points from a belief-based model of kindness (Dufwenberg and Kirchsteiger 2004) and reference dependence with loss aversion (Tversky and Kahneman 1991; Kőszegi and Rabin 2006), along with M-Consistency and M-Monotonicity, to make clear that the different predictions from the models cannot solely be attributed to different specifications of reference points.

Our application to public good games in section 3 uses a specification of "moral reference point" based on two *observable* features of the environment: endowments and minimal expectations payoffs. We derive moral monotonicity implications for effects of game form (provision, appropriation, or mixed) and nonbinding contractions on best response allocations in public good games, and for efficiency of (Nash) equilibrium play when allocations are strategic complements. Also in section 3, we derive predicted effects for game form and nonbinding contractions of alternative theoretical models including a prominent belief-based model of kindness (Dufwenberg and Kirchsteiger 2004), prominent models of reference points with loss aversion (Tversky and Kahneman 1991; Kőszegi and Rabin 2006), and conventional rational choice theory (including consequentialist social preferences models).

A summary statement of the testable implications of alternative theoretical models for experiments on game form and contractions is reported in statements I.a – I.c and II.a – II.c in section 3.4. Estimated coefficients reported in Tables 2 and 4 with data from experiments reported in three previous papers and data from our experiment are consistent with the implications of morally monotonic choice theory but mostly inconsistent with alternative theoretical models. Other tests reported in Online Appendices I.3 and I.4 support similar conclusions.

References

Abrams, B.A. and Schmitz, M.A., 1978. The Crowding Out Effect on Government Transfers on Private Charitable Contributions. Public Choice 33, 29-39.

Abrams, B.A. and Schmitz, M.A., 1984. The Crowding Out Effect on Government Transfers on Private Charitable Contributions: Cross Sectional Evidence. National Tax Journal 37, 563-568.

Afriat, S.N., 1967. The Construction of Utility Functions from Expenditure Data. International Economic Review 8, 66–77. https://doi.org/10.2307/2525382

Andreoni, J., 1990. Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving. Economic Journal 100, 464-477.

Andreoni, J., 1993. An Experimental Test of the Public-goods Crowding-out Hypothesis. American Economic Review 83, 1317-1327.

Andreoni, J., 1995. Warm-Glow versus Cold-Prickle: The Effects of Positive and Negative Framing on Cooperation in Experiments. Quarterly Journal of Economics 110, 1–21. https://doi.org/10.2307/2118508

Andreoni, J. and Miller, J.H., 2002. Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism. *Econometrica* 70(2), 737-753.

Arrow, K.J., 1959. Rational Choice Functions and Orderings. Economica 26 (102), 121-127.

Bernheim, B.D., 1986. On the Voluntary and Involuntary Provision of Public Goods. American Economic Review 76, 789-93.

Bolton, G.E. and Katok, E., 1998. An Experimental Test of the Crowding out Hypothesis: The Nature of Beneficial Behavior. Journal of Economic Behavior and Organization 37, 315-331.

Bolton, G.E., Ockenfels, A., 2000. ERC: A Theory of Equity, Reciprocity, and Competition. American Economic Review 90, 166–193. https://doi.org/10.1257/aer.90.1.166

Bougherara, D., Denant-Boemont, L., and Masclet, D., 2011. Cooperation and Framing Effects in Provision Point Mechanisms: Experimental Evidence. Ecological Economics 70, 1200–1210. https://doi.org/10.1016/ J. Ecol. Econ. 2011.01.023

Chaudhuri, A, 2011. Sustaining Cooperation in Laboratory Public Goods Experiments: A Selective Survey of the Literature. Experimental Economics 14, 47 – 83. https://doi.org/10.1007/s10683-010-9257-1

Chernoff, H.,1954. Rational Selection of Decision Functions. Econometrica 22 (4), 422-443.

Clotfelter, C.T., 1985. Federal Tax Policy and Charitable Giving. Chicago: University of Chicago Press.

Cox, C.A., 2015. Decomposing the Effects of Negative Framing in Linear Public Goods Games. Economic Letters 126, 63–65. https://doi.org/10.1016 J. Econ. Let. 2014.11.015

Cox, C.A., Stoddard, B., 2015. Framing and Feedback in Social Dilemmas with Partners and Strangers. Games 6, 394–412. https://doi.org/10.3390/g6040394

Cox, J.C., List, J.A., Price, M., Sadiraj, V., Samek, A., 2019. Moral Costs and Rational Choice: Theory and Experimental Evidence. Experimental Economics Center Working Paper 2019-02.

Cox, J.C., Ostrom, E., Sadiraj, V., and Walker, J.M., 2013. Provision versus Appropriation in Symmetric and Asymmetric Social Dilemmas. Southern Economic Journal 79, 496–512. https://doi.org/10.4284/0038-4038-2012.186

Cox, J.C., and Sadiraj, V., 2007. On Modeling Voluntary Contributions to Public Goods. Public Finance Review 35, 311–332.

Cubitt, R.P., Drouvelis, M., and Gächter, S., 2011. Framing and Free Riding: Emotional Responses and Punishment in Social Dilemma Games. Experimental Economics 14, 254–272. https://doi.org/10.1007/s10683-010-9266-0

Debreu, G., 1959. Theory of Value; An Axiomatic Analysis of Economic Equilibrium. Yale University Press.

Dufwenberg, M., and Kirchsteiger, G., 2004. A Theory of Sequential Reciprocity. Games and Economic Behavior 47, 268-298.

Fehr, E., and Schmidt, K.M., 1999. A Theory of Fairness, Competition, and Cooperation. Quarterly Journal of Economics 114, 817–868. https://doi.org/10.1162/003355399556151

Fosgaard, T.R., Hansen, L.G., and Wengström, E., 2014. Understanding the Nature of Cooperation Variability. Journal of Public Economics 120, 134–143. https://doi.org/10.1016/J.JPUBECO.2014.09.004

Fujimoto, H., and Park, E.-S., 2010. Framing Effects and Gender Differences in Voluntary Public Goods Provision Experiments. Journal of Socio-Economics 39, 455–457. https://doi.org/10.1016/J.SOCEC.2010.03.002

Hicks, J., 1946. Value and Capital; An Inquiry into Some Fundamental Principles of Economic Theory. Clarendon Press.

Kalai, E. and Smorodinsky, M. (1975). Other Solutions to Nash's Bargaining Problem. Econometrica. 43 (3): 513–518. doi:10.2307/1914280

Khadjavi, M., and Lange, A., 2015. Doing Good or Doing Harm: Experimental Evidence on Giving and Taking in Public Good Games. Experimental Economics 18, 432–441. https://doi.org/10.1007/s10683-014-9411-2

Khanna, L, and Sandler, T., 2000. Partners in Giving: The Crowding-in Effects of UK Government Grants. European Economic Review 44, 1543-1556.

Kingma, B.R., 1989. An Accurate Measurement of the Crowd-out Effect, Income Effect, and Price Effect for Charitable Contributions. Journal of Political Economy 97: 1197-1207

Kőszegi, B, and Rabin, M., 2006. A Model of Reference-Dependent Preferences. Quarterly Journal of Economics 121 (4), 1133–1165. https://www.jstor.org/stable/25098823

Ledyard, J.O., 1995. Public Goods: A Survey of Experimental Research. Ch. 2 in *The Handbook of Experimental Economics*, J.H. Kagel & A.E. Roth (eds.). Princeton University Press, Princeton, N.J.

Messer, K.D., Zarghamee, H., Kaiser, H.M., and Schulze, W.D., 2007. New Hope for the Voluntary Contributions Mechanism: The Effects of Context. Journal of Public Economics 91, 1783–1799. https://doi.org/10.1016/J.JPUBECO.2007.08.001

Nash, J., 1953. Two-Person Cooperative Games. Econometrica 21, 128-140. doi:10.2307/1906951.

Park, E.-S., 2000. Warm-Glow versus Cold-Prickle: A Further Experimental Study of Framing Effects on Free-Riding. Journal of Economic Behavior and Organization 43, 405–421. https://doi.org/10.1016/S0167-2681(00)00128-1

Rabin, M., 1993. Incorporating Fairness into Game Theory and Economics. American Economic Review 83, 1281-1302.

Reuben, E., and Riedl A., 2013. Enforcement of Contribution Norms in Public Goods Games with Heterogeneous Populations. Games and Economic Behavior 77, 122-137.

Ribar, D.C. and Wilhelm, M.O., 2002. Altruistic and Joy-of-giving Motivations in Charitable Behavior. Journal of Political Economy 110, 425-457.

Roth, A.E., 1977. Independence of Irrelevant Alternatives, and Solutions to Nash's Bargaining Problem. Journal of Economic Theory l6, 247-251.

Samuelson, P.A., 1947. Foundations of Economic Analysis. Harvard University Press.

Sen, A.K., 1971. Choice Functions and Revealed Preference. Review of Economic Studies 38, 307–317. https://doi.org/10.2307/2296384

Sen, A.K., 1986. Social Choice Theory. *Handbook of Mathematical Economics*, vol. III, ed. by K. J. Arrow and M. Intriligator. Amsterdam: North-Holland.

Sen, A.K., 1993. Internal Consistency of Choice. Econometrica 61, 495–521. https://doi.org/10.2307/2951715

Sonnemans, J., Schram, A., and Offerman, T. (1998). Public good provision and public good prevention: the effect of framing. Journal of Economic Behavior and Organization 34, 143-161.

Tarski, A., 1955. A Lattice-Theoretical Fixpoint Theorem and Its Applications. Pacific Journal of Mathematics 5, 285-309.

Tversky, A. and Kahneman, D., 1991. Loss aversion in Riskless Choice: A Reference-Dependent Model. Quarterly Journal of Economics 106, 1039-1061.

van Soest, D., Stoop, J., Vyrastekova, J., 2016. Toward a Delineation of the Circumstances in Which Cooperation can be Sustained in Environmental and Resource Problems. Journal of Environmental Economics and Management. 77, 1–13. https://doi.org/10.1016/J.JEEM.2015.12.004

Varian, H.R., 1982. The Nonparametric Approach to Demand Analysis. Econometrica 50, 945–973. https://doi.org/10.2307/1912771

ONLINE APPENDICES I

Appendix I.1: Related Literature on Payoff-Equivalent Public Good Games

To our best knowledge, Andreoni (1995) is the first study to look at behavior in positively-framed and negatively-framed voluntary contributions public good games. His between-subjects experiment co-varied game form (provision or appropriation) with wording of subject instructions that made highly salient the positive externality from contributions in a provision game or the negative externality from extractions in an appropriation game. Subsequent literature explored both empirical effects of variations in evocative wording of subject instructions and effects of changing game form (from provision to appropriation) with neutral wording in the subject instructions. We here summarize findings on effects of game form and various framings on contributions, extractions, and beliefs.

Subjects' Characteristics

Some studies look at interaction between subjects' attributes (social-value orientation, gender, attitudes towards gains and losses) and game framing (positive or negative). The main findings include: (1) play of individualistic subjects but not social-value oriented subjects is sensitive to the framing of the game (Park 2000); (2) more cooperative behavior by women than men in the negatively-framed game but not in the positively-framed game; (3) for both genders, positive framing elicits higher cooperation than negative framing (Fujimoto and Park 2010); and (4) lower cooperation in taking than in giving scenarios with gain framing but the effect appears to be driven entirely by behavior of male subjects (Cox 2015). With loss framing, no clear effect is detected (Cox 2015).[44] Cox and Stoddard (2015) explore effects of interaction of partners vs. strangers pairing with individual vs. aggregate feedback in payoff equivalent provision (give) and appropriation (take) games and find that the take frame together with individual feedback induces bimodal behavior by increasing both complete free riding and full cooperation.

Beliefs and Emotions

While give vs. take frames are found to affect contributions, this effect appears to be less strong than the effect on beliefs (Dufwenberg, Gächter, and Hennig-Schmidt 2011; Fosgaard, Hansen,

[44] In the Loss-Giving setting, subjects *contribute* to prevent loss whereas in the Loss-Taking setting, subjects *take* to generate a loss (Cox 2015).

and Wengström 2014). A close look at triggered emotions in positively-framed and negatively-framed public good games is offered by Cubitt, Drouvelis, and Gächter (2011) who find no significant effects of punishments or reported emotions.[45] This is one of few studies that find no game form effect on contributions.

Environment

Studies in this category focus on effects of features of the environment (such as status quo, communication, power asymmetry) on play across take or give public good games. Messer, et al. (2007) report an experimental design that interacts status quo (giving or not giving) in a public good game with presence or absence of cheap talk or voting. They find that changing the status quo from "not giving" to "giving" increases average contributions in the last 10 rounds from 18% (no cheap talk, no voting) up to an astonishing 94% (with cheap talk and voting). Cox, et al. (2013) report an experiment involving three pairs of payoff-equivalent provision and appropriation games. Some game pairs are symmetric while others involve asymmetric power relationships. They find that play of symmetric provision and appropriation, simultaneous-move games produces comparable efficiency whereas power asymmetry leads to significantly lower efficiency in sequential appropriation games than in sequential provision games. Cox, et al. (2013) conclude that reciprocity, but not unconditional other-regarding preferences, can explain their data. A framing effect on behavior is observed in public good games with provision points (Bougherara, Denant-Boemont, and Masclet 2011, Sonnemans, Schram and Offerman, 1998). In their experiment, van Soest, Stoop and Vyrastekova (2016) compare outcomes in a provision (public good) game with outcomes in a claim game in which subjects can appropriate the contributions of others before the public good is produced. They report non-positive production of the public good in the claim game even in early rounds of the experiment.

The experiment in the literature that is most closely related to ours is reported by Khadjavi and Lange (2015). They report on play in a mixed game with a between-subjects design that includes opportunities for both provision (give) and appropriation (take) with the initial (exogenously-specified) endowments between those in give or take scenarios. They find that (1) the appropriation game induces less cooperative behavior than the provision game (replicating the

[45] Cubitt et al. (2011) use two measures of emotional response including self-reports and punishment.

central result in Andreoni 1995) and that (2) their mixed frame data does not differ significantly from data for their provision game.

One notable difference of our experimental approach from previous literature is inclusion of a within-subjects design for eliciting provision and appropriation responses in three different mixed games that span the design space between the pure provision and appropriation games. A more fundamental departure from previous experimental literature is our inclusion of *endogenous* contractions of feasible sets, in a within-subjects design, that is motivated by the Consistency property of rational choice theory (Arrow 1959, Sen 1971, 1986). While the Khadjavi and Lange (2015) design allows for *exogenous* contraction in the mixed game our design introduces endogenous contractions known to include previous allocations in (provision or appropriation) games in addition to elicited beliefs about others' allocations. Such endogenous contractions are essential to ascertaining whether behavior in provision, appropriation, and mixed games exhibits monotonicity in moral reference points.

Appendix I.2: Play in Provision, Appropriation, and Mixed Games

A general description of provision, appropriation, and mixed games with public goods is as follows. Each player, $i \in N = \{1, \cdots, n\}$ chooses an allocation (w_i, g_i) of a scarce resource, W tokens, between two accounts: w_i to player *i*'s private account and g_i to the public account shared with *n*-1 other players. When the total of others' allocations to the public account is G_{-i}, player *i*'s money payoff is the sum of returns from the private and public accounts:

(I.2.1) $$\pi_i = w_i + \gamma(g_i + G_{-i})$$

where $w_i = W - g_i$ and $\gamma \in (1/n, 1)$ denotes the mpcr from the public account.

The initial *per capita* endowed tokens, g^e in the public account uniquely identifies the g^e-game with total endowment $ng^e \in [0, nW]$ to the public account and endowment $W - g^e$ to the private account of each of the *n* players. Special cases include: provision game $(g^e = 0)$, where a public good can be provided; appropriation game $(g^e = W)$, where a public good can be appropriated; and mixed games $(g^e \in (0, W))$, where both provision and appropriation of a public good are feasible.

Let $g_{-i} \in \{0,\cdots,W\}^{n-1}$ be a vector of allocations to the public account by players other than player *i*. Let $\pi \in \Re^n$ denote the *n*-vector of payoffs to all players including *i*. In the g^e-game the feasible set of player *i* (in the money payoff space) is

(I.2.2) $S(g_{-i}) = \{\pi(x, g_{-i}) \mid (x, g_{-i}) \in \{0,\cdots,W\}^n\}$

If we let $g_i^* = br(g_{-i})$ denote a best-response allocation by agent *i* when others' *n*-vector of allocations to the public account is g_{-i} then the *n*-vector of payoffs, $\pi(g_i^b, g_{-i})$ belongs to the choice set, $S^*(g_{-i})$, that is

(I.2.3) $\pi(g_i^*, g_{-i}) \in S^*(g_{-i}) \subseteq S(g_{-i})$

Implications of Conventional Rational Theory for Choice in Provision, Appropriation and Mixed Games

The first observation (see Online Appendix II.2) is that Consistency implies that player *i*'s allocation set, $W^*(g_{-i} \mid g^e)$ remains the same if, instead of $\{0,\cdots,W\}$, player *i* is asked to choose from some (non-binding) contracted subset, $C = \{c,\cdots,W\}, c > 0$ that contains all allocations in vector g_{-i} as well as *i*'s smallest best response allocation, g_i^b for which $\pi(g_i^b, g_{-i}) \in S^*(g_{-i})$. For any given *c* such that (*) $0 < c < \min(g_i^b, \min(g_{-i}))$ the feasible payoff set is

(I.2.4) $T(g_{-i}) = \{\pi(x, g_{-i}) \mid (x, g_{-i}) \in C^n\} \subseteq S(g_{-i})$

where the inclusion follows from the minimum compulsory allocation (*) and payoff function (I.2.1). By the Consistency property $S^*(g_{-i}) \cap T(g_{-i}) = T^*(g_{-i})$, and by construction, $S^*(g_{-i}) \subset T(g_{-i})$, so $S^*(g_{-i}) = T^*(g_{-i})$ Hence *i*'s allocations in the (nonbinding) contraction game remains the same, $C^*(g_{-i} \mid g^e) = W^*(g_{-i} \mid g^e)$ for all *c* that satisfy (*).

The second observation is that the Consistency property requires that player *i*'s (best response) chosen allocations are not affected by initial (the endowed per capita) allocation, g^e in the public account (see Online Appendix II.2) because the feasible set in the payoff space, $S(g_{-i})$ remains the same for all $g^e \in [0, W]$.

Implications of these two observations are summarized in Proposition 3 in the text.

Moral Reference Point in g^e-games

Without any loss of generality, we focus on moral reference point from the perspective of player 1.[46] The initial endowed payoff of each player $i \in N$ in a g^e-game is

(I.2.5) $\omega_i^e = W - g^e + \gamma(ng^e) = W + (\gamma n - 1)g^e$

For any given vector, g_{-1} of others' allocation in the g^e-game with feasible allocations from $C = \{c, \cdots, W\}$, player 1's feasible set (in payoff space) is $T(g_{-1}) = \{\pi(x, g_{-1}) \mid (x, g_{-1}) \in C^n\}$. The minimal expectation payoff of a player $k \in N \setminus \{1\}$, as a consequence of player 1's allocation, is when player 1 allocates the minimum required amount and leaves player k with payoff

(I.2.6) $k_* = W - g_k + \gamma(G_{-1} + c)$

where G_{-1} is the total of voluntary allocations, g_{-1} in the public account by other players. So, from the perspective of player 1, the moral reference point with respect to player $k \neq 1$ is the ordered pair,

(I.2.7) $r_k^1 = (\omega_k^e, k_*) = (W - g^e + \gamma ng^e, W - g_k + \gamma(G_{-1} + c))$

The minimal expectation payoff of player 1, as a consequence of player 1's allocation, is when player 1 allocates all his W tokens in the public account. Hence, player 1's minimal expectation payoff is

(I.2.8) $1_* = \gamma(W + G_{-1})$

So, from the perspective of player 1, the moral reference point with respect to oneself is the ordered pair,

(I.2.9) $r_1^1 = (\omega_1^e, 1_*) = (W - g^e + \gamma ng^e, \gamma(W + G_{-1}))$

Replace "1" with "*i*" in statements (I.2.7-9) to get the moral reference point, $r^i \in \mathbb{R}^{2n}$ from the perspective of player *i* at feasible set $T(g_{-i})$:

---

[46] Separate detailed explanations of moral reference points in provision, appropriation, and mixed games with contraction ($c > 0$) or without contraction ($c = 0$) can be found in Online Appendix II.5.A.

$$\text{(I.2.10)} \qquad \begin{aligned} r_k^i &= (\omega_k^e, k_*) = (W - g^e + \gamma n g^e, W - g_k + \gamma(G_{-i} + c)), \quad & k \in N \setminus \{i\} \\ r_i^i &= (\omega_i^e, i_*) = (W - g^e + \gamma n g^e, \gamma(G_{-i} + W)), & k = i \end{aligned}$$

Implications of Moral Monotonicity for Best Response Allocations across g^e-games

Contraction Effect. Let the g^e-game and vector of others' allocations, g_{-i} be given. For any two constraints $c_1 > c_2 \geq 0$ on minimum permissible allocations, let $r^{ic_1}, r^{ic_2} \in \mathbb{R}^{2n}$ denote the respective moral reference points as in statement (I.2.10):

$$\begin{aligned} r_k^{ix} &= (\omega^e, k_*^x) = (W - g^e + \gamma n g^e, W - g_k + \gamma(G_{-i} + x)), \;\; k \neq i \\ r_i^{ix} &= (\omega^e, i_*^x) = (W - g^e + \gamma n g^e, \gamma(G_{-i} + W)), \qquad k = i \end{aligned}$$

where $x \in \{c_1, c_2\}$. Verify that player k's total change (defined in the Notation paragraph in the text) between the two reference points is

$$\begin{aligned} \delta_k(r^{ic_1}, r^{ic_2}) &= (\omega_k^e - \omega_k^e) + (k_*^{c_1} - k_*^{c_2}) = 0 + \gamma(c_1 - c_2), \quad & k \in N \setminus \{i\} \\ \delta_i(r^{ic_1}, r^{ic_2}) &= (\omega_i^e - \omega_i^e) + (i_*^{c_1} - i_*^{c_2}) = 0 + 0, & k = i \end{aligned}$$

For games with contraction, $c_1 = c > 0$ compared to no contraction, $c_2 = 0$ the set $K = \{k \in N : \; \delta_k = \gamma c > 0\} = N \setminus \{i\}$. Moral monotonicity requires that player *i* leaves some other player with larger extreme payoffs in the g^e-game with contraction (than in the game without contraction), which player *i* can do by increasing his (best response extreme) allocations to the public account

Initial Endowment Effect. By statement (I.2.10) for any two g^e-games with initial (per capita) allocations $g^s > g^t$ in the public account, $\delta_k = (\omega_k^s - \omega_k^t) + (k_* - k_*) = (\gamma n - 1)(g^s - g^t) + 0$ for all $k \in N$. It follows from $\gamma \in (0, 1/n)$ that $K = \{k \in N : \delta_k = \max_{i \in N} \delta_i > 0\} = N$ and by M-Monotonicity, player *i* aims for larger (extreme) final payoff in the game with the larger per capita endowed tokens, g^e in the public account. These findings are summarized in Proposition 4 in the text (see Online Appendix II.5.B for formal proofs).

Appendix I.3: Robustness Tests

Table I.3.1 Individual Allocations to Public Account in Our Experiment (Linear Reg.)

Dep. Variable: g_i Allocation	Exclude data from contractions where the rule at least "-$1" does not apply				
	Exclude data from	1st C in CBC		C=B or 1st C in CBC	
Guessed Other's allocation		0.61***	0.60***	0.64***	0.63***
		(0.049)	(0.047)	(0.047)	(0.044)
g^e [-]		-0.05*	-0.07**	-0.05*	-0.07**
		(0.031)	(0.032)	(0.029)	(0.030)
c [+]		0.43***	0.44***	0.40***	0.42***
		(0.064)	(0.061)	(0.065)	(0.059)
Demographics		no	yes	no	yes
Observations		657	657	571	571
R-Squared		0.421	0.446	0.442	0.472

Notes: Total number of subjects is 232. Robust standard errors (clustered at subject level) in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table I.3.1a. Individual Allocations to Public Account in Our Experiment (Tobit Reg.)

Dep. Variable: g_i Allocation	All Data		exclude 1st C in CBC		xclude C=B and 1st C in CBC	
Guessed Other's allocation	1.10***	1.08***	1.06***	1.04***	1.07***	1.06***
	(0.091)	(0.087)	(0.093)	(0.088)	(0.089)	(0.083)
g^e [-]	-0.18***	-0.20***	-0.16***	-0.19***	-0.14**	-0.16***
	(0.059)	(0.060)	(0.059)	(0.060)	(0.054)	(0.055)
c [+]	0.23**	0.24**	0.46***	0.48***	0.43***	0.45***
	(0.109)	(0.103)	(0.128)	(0.120)	(0.126)	(0.115)
Demographics	no	yes	no	yes	no	yes
Observations	696	696	657	657	571	571
(left-, un-, right-) censored obs	(242, 352, 102)		(217, 340, 100)		(177, 305, 89)	

Notes: Total number of subjects is 232. Predicted signs for moral monotonicity in square brackets. Demographics include dummies for Female, Black, Self Image (give to a stranger, give to charity, help others with homework, share secrets) and Other's Image (disabled car assistance, selfish, dislike helping others). Robust standard errors (clustered at subject level) in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Further Data Analysis We looked at game form effect utilizing non-parametric tests for statistical inferences and conducted a within-subject analysis focusing only on allocations of subjects whose beliefs did *not* change.

<u>*Allocations in Provision, Appropriation, and Mixed Games (Between-Subjects Analysis)*</u>

Figure I.3.1 shows histograms across games of subjects' allocations in the full games, that is, allocations are from {0,…,10}. Extensive margin effect is visible: free-riding behavior (allocating nothing in the public account) is lowest in the provision game (21%), highest in the appropriation game (48%), with the mixed games in between (39%).[47] Average token allocations in the public account exhibit a decreasing pattern: 4.01 (provision), 3.64 (mixed) and 3.09 (appropriation).[48] For statistical inferences we use the Kolmogorov-Smirnov test for distributions

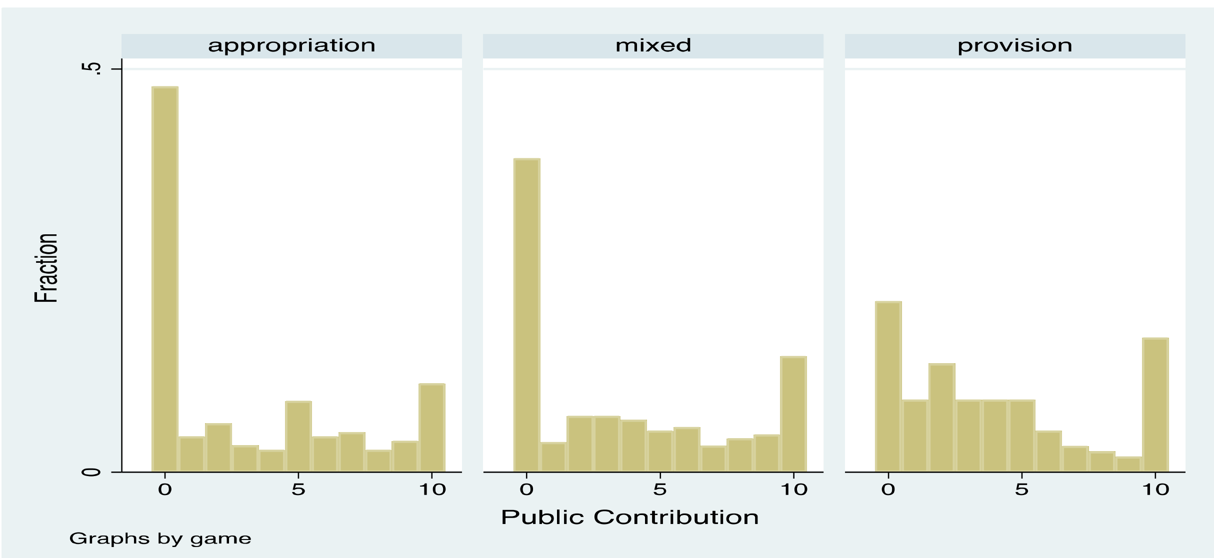


Figure I.3.1. Histograms of subject's *g* allocation from the full set, {0,…,10}

of *g* allocations in the public account and the Pearson chi-square test for free-riding behavior, and find that public good allocations of subjects in our experiment are characterized by:

(i) Larger public account allocations (p-value=0.022) and less free-riding (p-value=0.003) in provision than appropriation game data;

[47] If we allow for one token error, classifying 0 or 1 token allocations as free-riding, we get similar figures: 30.13% in provision game, 42.59% in mixed games and 52.2% in appropriation game. The odds of free-riding in provision game is less than half (0.42, *p*-value=0.01) in mixed games but in appropriation game it is 1.44 (*p*-value=0.18).

[48] The 95% Confidence Intervals are: [3.46, 4.57] in provision game, [3.13, 4.15] in mixed game and [2.55, 3.63] in appropriation game.

(ii) Similar public account allocations (p-value=0.497) and less free-riding (p-value=0.247) in provision than mixed game data;

(iii) Similar public account allocations (p-values=0.384) but less free-riding (p-value=0.075) in mixed and appropriation game data.

Based on these findings we conclude:

Result 1. *The provision game elicits higher average allocation to the public account than the appropriation game and the appropriation game elicits more free riding (public account allocations of 0 or 1).*

<u>*Within-Subjects Data Analysis Controlling for Beliefs*</u>

g^e *-Effect*. In mixed game treatments, excluding *selfish* subjects (who allocated 0 in all three tasks) we have 35 observations with unchanged beliefs.[49] For each subject, we constructed $\Delta g = g^i - g^j$, when the subject's guessed allocation of others in games g^i and g^j was the same, where superscripts $i<j$ denote the initial per capita endowed tokens, g^e from {2, 5, 8}. The null hypothesis from conventional rational choice theory is the mean of the distribution of Δg is not statistically different from 0 (Proposition 3, part b and Corollary 2, part b) whereas the alternative hypothesis that follows from moral monotonicity is mean (Δg) > 0 (Proposition 4, part b and Corollary 3, part b). The mean of Δg is 1.23 (95% C.I.=[-0.08, 2.53]) and the (conventional theory) null hypothesis is rejected by the *t*-test (t-statistic=1.91; p-value=0.064) in favor of moral monotonicity.[50] Our next result is:

Result 2. *Allocation to the public account in mixed games decreases as the initial endowment of the public account increases, controlling for belief about other's allocation*.

Contraction Effect. For any given allocation by the other player, conventional theory requires that (best-response) g allocations in the provision game or appropriation game be invariant to *nonbinding* contractions whereas moral monotonicity predicts that (best-response) allocations increase in c for nonbinding contractions. We constructed a new variable, Δg_i^{cb} that takes its

---

[49] If we include 0 allocations of *selfish* subjects with unchanged beliefs, the number of observations increases from 35 to 55.

[50] If we include 0 from *selfish* subjects, the mean of Δg decreases to 0.78 (the 95% C.I. is [-0.05, 1.61]); t-statistic=1.89; p-value=0.064).

values according to the difference between the subject's observed g allocation in the public account from the contracted set, $C=\{c,\ldots,10\}$ and the subject's allocation chosen from the full set, $B=\{0,\ldots,10\}$. The null hypothesis from conventional theory is that the mean of Δg_i^{cb} is not statistically different from 0, provided that the guess of other's contribution did not change. We have 45 observations for Δg_i^{cb} (24 and 21, resp., in the appropriation and provision treatments) observations with unchanged guesses and proper contractions (c>0). The mean of Δg_i^{cb} is significantly larger than 0 in the provision game (0.95, p-value=0.042) but not in the appropriation game (0.54, p-value=0.313). We also looked at the subset of these 45 observations with non-binding contractions; this leaves us with 28 observations for Δg_i^{cb} (17 and 11 in the provision and appropriation game) with unchanged guesses and proper contractions (i.e., c > 0). The mean of Δg_i^{cb} is 0.88 and statistically significantly larger than 0 in the provision game (p-value=0.056) but not in the appropriation game (mean=-0.36, p-value=0.476).

As a further check that the preceding tests are picking up (full vs. contracted game) treatment effects rather than decision-order effects, we also looked at Δg_i^{bb}, the within-subject difference in allocations in tasks in which subjects faced the full set, $B=\{0,\cdots,10\}$ more than once (e.g. in the BCB sessions) and their guesses did not change. There are 73 observations for Δg_i^{bb} with unchanged reported guesses. Both conventional rational choice theory and moral monotonicity require the mean of the distribution of Δg_i^{bb} to be 0. Data fail to reject this null hypothesis (mean of Δg_i^{bb} is 0.05, *t*-statistic=0.35, p-value=0.73). Our third result is:

Result 3. *Nonbinding lower bounds on public account allocations induce higher average allocations to the public account in the provision game, controlling for beliefs about other's allocation.*

I.4: Maximization Approach to Testing Conventional and Moral Monotonicity Theory

As an example for tractable applications, we apply moral monotonicity theory using a parametric choice function and comparative statics analysis for interior solutions.

Special Case Morally Monotonic Best Response Allocations

Without loss of generality, consider allocations by player 1. Assume a parametric form of the choice function, $U(\pi \mid r)$ discussed in the text:[51]

(I.4.1) $$u(\pi) = (1 - e^{-\alpha\pi}), \alpha > 0,\ \theta(r_1) = e^{\sigma(r_{11}+r_{12})}, \sigma > 1 \text{ and for all } k > 1,\ \theta(r_k) = e^{r_{k1}+r_{k2}}.$$

Optimal (interior) solution is determined by:

(I.4.2) $$\sum_{k>1} \mathrm{e}^{(r_{k1}+r_{k2})-\sigma(r_{11}+r_{12})}\,\mathrm{e}^{-\alpha(\pi_k-\pi_1)} = (1-\gamma)/\gamma$$

Let G denote the total allocations to the public account. Verify that for all k

$\pi_k - \pi_1 = (\omega - g_k + \gamma G) - (\omega - g_1 + \gamma G) = g_1 - g_k$, substitute it in (I.4.2) and solve for g_1 to get

(I.4.3) $$g_1^*(g_{-1}, r) = br(g_{-1} \mid r) = \frac{1}{\alpha}\left(\ln\left(\frac{\gamma}{1-\gamma}\right) - \sigma(r_{11}+r_{12}) + \ln\left(\sum_{k>1} e^{r_{k1}+r_{k2}+\alpha g_k}\right) \right)$$

Details of the derivation of (I.4.3) are reported in Online Appendix II.7. It is straightforward (see that appendix for details) to show that, consistent with the general-case Proposition 4, $g_1^*(\cdot)$ increases in c and decreases in g^e.

Structural Analysis of Experimental Data. Estimating equations applied to data come from the best response function in statement (I.4.3). We estimate parameters for α and σ using data from Andreoni (1995), Khadjavi and Lange (2015), and the experiment reported herein.

In our experiment, we have a two-player game and the belief about other's allocation is elicited, so the estimating equation can be written as

(I.4.4) $$g_{1t}^* = \frac{1}{\alpha}\left(\ln\left(\frac{\gamma}{1-\gamma}\right) - \sigma R_{1t} + R_{2t} + \alpha g_{2t} \right)$$

where g_{1t}^* is the individual's allocation at round t, g_{2t} is the elicited belief at round t, and $R_{kt} = (r_{k1})_t + (r_{k2})_t$, where the moral reference point is as reported in Table 1 in the main text.

In the Andreoni experiment and the Khadjavi and Lange experiment, at the end of each round subjects are informed of the total allocation, G_t in the public account and of course they

[51] Superscript "1" on the reference point variables is dropped to simplify notation.

know their own allocation, so they know total allocation by others in the public account, G_{-1t}. Therefore, for empirical estimation we assume that, at the beginning of each period t, the elicited belief is that every other player's allocation is the known average from the preceding period, $G_{-1(t-1)}/(n-1)$. Hence, statement (I.4.3) becomes

(I.4.5) $$g_1^* = br(g_{-1} \mid r) = \frac{1}{\alpha}\left(\ln(\frac{\gamma}{1-\gamma}) - \sigma R_1 + \ln\left((n\text{-}1)\, e^{R_k + \alpha g_k}\right)\right)$$

Parameter estimation with data from the Andreoni experiment and the Khadjavi and Lange experiment thus uses the estimating equation

(I.4.6) $$g_{1t}^* = \frac{1}{\alpha}\left(\ln(\frac{\gamma}{1-\gamma}) - \sigma R_{1t} + \ln(n-1) + R_{-1t} + \alpha g_{-1t}\right)$$

where g_{1t}^* is individual's allocation observed in round t, $g_{-1t} = G_{-1(t-1)}/(n-1)$ reported after round t-1, and the moral reference point specification is as in the main text. An estimate of parameter σ (weakly) smaller than 1 would be inconsistent with moral monotonicity. Table I.4.1 reports nonlinear least square estimates of α and σ for all data, as well as separately for

Table I.4.1. Non-linear Least Squares Estimates for Parametric Choice Function

	Andreoni (1995)[a]	**K&L (2015)**[a]		**New Experiment**	
Parameters	All Data	All Data	No Contraction	All Data	No Contraction
$\sigma\,[>1]$	1.11***	1.17***	1.20***	1.02***	1.03***
	(0.012)	(0.024)	(0.026)	(0.016)	(0.019)
	[1.09, 1.14]	[1.13, 1.22]	[1.15, 1.26]	[0.99, 1.05]	[0.99, 1.07]
α	1.70***	3.38***	3.04***	2.69***	2.74***
	(0.233)	(0.483)	0.453))	(0.340)	(0.380)
	[1.24, 2.17]	[2.43, 4.33]	[2.15, 3.94]	[2.02, 3.36]	[2.00, 3.49]
Observation	720	1440	1080	696	554
R-squared	0.41	0.69	0.53	0.75	0.67
Clusters	80	160	120	232	232

Notes: [a]Round 1 data are not included for Andreoni and K&L data because there is no information on others' contributions. Required value for consistency with moral monotonicity in square brackets. Robust standard errors in parentheses. 95% Confidence Intervals in square brackets.

games without contraction because in Khadjavi and Lange's experiment contraction is exogenous (and therefore can be binding for some subjects). The estimated parameter for α is significantly greater than 0, revealing increasing $u(\cdot)$. The estimated parameter for σ is significantly greater than 1 with data from each of the experiments, which is consistent with moral monotonicity.[52]

ONLINE APPENDICES II

Notation. $N = \{1, \cdots, n\}$ denotes the set of players, superscript * will be used for choices, subscript $_{-i}$ has the conventional meaning (i.e., all players other than *i*).

Appendix II.1 Proofs of Propositions 1 and 2

Proof of Proposition 1 To simplify writing we drop the superscript z (as agent *z* is given and fixed). Take any two feasible problems: (S,s) and (T,t) where *S* and *T* are nonempty finite sets in $\Re^n$ and $s,t \in (\Re^m)^n$ are the reference points, respectively. Suppose $T \subseteq S$, $s^* \in S^*(s) \cap T$, $\delta_k(t,s) \neq 0$ for some agent *k* and $\delta_{-k}(t,s) = 0$.

Consider some other problem, where the feasible set is *T* and the reference point, *x* is as favorable as *s* to every agent *i*, that is (T,x) with reference point $x = \{x_i \in (\Re^m)^n : i \in N\}$ such that $\delta_i(x,s) = 0, \forall i \in N$. By M-Consistency, $T^*(x) = S^*(s) \cap T$, so $s^* \in T^*(x)$. It follows that: (i) $s_k^* \geq \min T_k^*(x)$ and (ii) $\max T_k^*(x) \geq s_k^*$. Next, for problems (T,x) and (T,t), note that for all $i \in N$, $\delta_i(t,x) = \delta_i(t,s) + \delta_i(s,x) = \delta_i(t,s)$.

Case 1. $\delta_k(t,s) > 0$. By $\delta_k(t,x) = \delta_k(t,s) > 0$ and M-Monotonicity, $T_k^*(t) \triangleright T_k^*(x)$, hence $\max T_k^*(t) \geq \max T_k^*(x)$ which together with (ii) imply $\max T_k^*(t) \geq s_k^*$. If $t^* \in T^*(t)$ is some point where $\max T_k^*(t)$ is reached then $(t_k^* - s_k^*)\delta_k(t,s) \geq 0$.

Case 2. $\delta_k(t,s) < 0$. By $\delta_k(x,t) = -\delta_k(t,s) > 0$ and M-Monotonicity, $T_k^*(x) \triangleright T_k^*(t)$, hence $\min T_k^*(x) \geq \min T_k^*(t)$, which together with (i) completes the proof as for $t_* \in T^*(t)$ where

---

[52] Using these estimates, the Nash (symmetric) equilibrium allocations (as a percentage of *W*) in the provision game are: 29% (Andreoni 1995), 33% (K&L 2015) and 32% (our experiment), and lower in the appropriation game: 16% (Andreoni 1995), 24% (K&L 2015) and 31% (our experiment). These figures suggest empirical support for Proposition 5.

$\min T_k^*(t)$ is reached, one has $s_k^* \geq t_{*k}$, that is $(t_{*k} - s_k^*)\delta_k(t,s) \geq 0$.

Proof of Corollary 1. Without any loss of generality, let agent 2 be the one favored by the reference point *t* compared to reference point *s*, and agent 1 be the dis-favored one. Let $\{s^*\}$ and $\{t^*\}$ be agent 1's choice sets in problems (S,s) and (T,t). It suffices to show that $s_2^* \leq t_2^*$, as by Pareto efficiency, $t_1^* \geq s_1^*$. Consider feasible problem, (S,x) for some $x \in (\Re^m)^2$ such that $\delta_1(x,s) = 0, \delta_2(x,s) = \delta_2(t,s) > 0$ and $\delta_1(x,t) = -\delta_1(t,s) > 0$, $\delta_2(x,t) = \delta_2(s,t) = 0$.

Compared to *s*, reference point *x* favors agent 2 but neither favors nor dis-favors agent 1, which together with $s^* \in T$ satisfy Scenario A for feasible problems (S,s) and (T,x). By Proposition 1, $x_2^* \geq s_2^*$ (*).

On the other hand, compared to *t*, reference point *x* favors agent 1 but neither favors nor disfavors agent 2. By M-Monotonicity, applied to feasible problems (T,x) and (T,t), $x_1^* \geq t_1^*$, and by Pareto efficiency, $x_2^* \leq t_2^*$ which together with (*) imply $s_2^* \leq t_2^*$.

Proof of Proposition 2. Let agent 1's choice set, $X^*(r)$ for feasible (finite nonempty) set $X \subset \Re^n$ in the payoff space and moral reference point $r \in (\Re^m)^n$, be determined as follows:

$$X^*(r) = \{\pi^* \in X : U(\pi^* \mid r) \geq U(\pi \mid r), \forall \pi \in X\} \qquad \text{(II.1.1)}$$

where $U(\pi \mid r) = \sum_{k \in N} w_k(r) u(\pi_k)$ for some strictly increasing $u(\cdot)$, and weights $w_k(r) = \theta_k^r / \sum_{i \in N} \theta_i^r$, with $\theta_k^r = \theta(\sigma_k \sum_{j=1}^m r_{kj})$, $\sigma_1 > 1 = \sigma_{k>1}$, for some strictly increasing function $\theta(\cdot)$ such that $\theta(y+z) = \theta(y)\theta(z)$. Note that for any given reference points, *t* and *s* from $(R^m)^n$ and all $i \in N$

$$\theta_i^t = \theta(\sigma_i \sum_{j=1}^m t_{ij}) = \theta(\sigma_i (\sum_{j=1}^m s_{ij} + \delta_i(t,s))) = \theta(\sigma_i \sum_{j=1}^m s_{ij})\theta(\sigma_i \delta_i(t,s)) = \theta_i^s \theta(\sigma_i \delta_i(t,s)), \qquad \text{(II.1.2)}$$

<u>M-Consistency</u> is clearly satisfied; if $\delta_i(t,s) = 0$, for all $i \in N$ then by (II.1.2) and[53] $\theta(0) = 1$, the weights satisfy $w_i(t) = w_i(s), \forall i \in N$, hence $U(\pi \mid s) = U(\pi \mid t)$ for all $\pi \in \Re^n$.

---

[53] Note that: $\theta(0)=1$ as $\theta(1)=\theta(1+0)=\theta(1)\theta(0)$

For M-Monotonicity, write $U(\pi \mid r) = w(r) \cdot u(\pi)$ and note that for any $T = S$ and any choices $t^* \in T^*(t)$ and $s^* \in S^*(s)$

$$\begin{aligned}&(1) \quad w(t) \cdot \left(u(t^*) - u(s^*)\right) \geq 0 \\ &(2) \quad w(s) \cdot \left(u(s^*) - u(t^*)\right) \geq 0\end{aligned} \tag{II.1.3}$$

Multiply (1) and (2) by the respective denominators and add the two expressions to get

$$\left(\theta(t) - \theta(s)\right) \cdot \left(u(t^*) - u(s^*)\right) \geq 0 \tag{II.1.4}$$

M-Monotonicity Let $K = \{k \in N : \ \delta_k = \max_{i \in N} \delta_i(t,s) > 0\}$. We show that if $K \neq \emptyset$ and $\delta_i = 0, \forall i \notin K$ then there exists some nonempty set of players, $K^* \subseteq K$ such that $T_k^*(t) \triangleright S_k^*(s)$ for all $k \in K^*$; $1 \in K^*$ if $K = N$.

It follows from $\theta(\sigma_i \delta_i) = \theta(0) = 1$, $i \notin K$ and (II.1.2) that statement (II.1.4) is

$$\sum_{k \in K} \theta_k^s \left(\theta(\sigma_k \delta_k) - 1\right)\left(u(t_k^*) - u(s_k^*)\right) \geq 0 \tag{II.1.5}$$

For each k, the first term, θ_k^s is positive, the second term is also positive as for all individuals from K, $\theta(\sigma_k \delta_k) > \theta(0) = 1$. Hence, $\left(u(t_k^*) - u(s_k^*)\right) \geq 0$ for some $k \in K$, and by monotonicity of $u(\cdot)$, $t_k^* \geq s_k^*$ for such k.

Next, we show that if $K = N$ then $1 \in K^*$. Divide by $\theta(\delta_1) - 1$ in (II.1.5) to get

$$\frac{\theta(\sigma\delta_1) - 1}{\theta(\delta_1) - 1} \theta_1^s \left(u(t_1^*) - u(s_1^*)\right) + \sum_{k > 1} \theta_k^s \left(u(t_k^*) - u(s_k^*)\right) \geq 0$$

Replace the second term with $\theta(s) \cdot (u(t^*) - u(s^*)) - \theta_1^s (u(t_1^*) - u(s_1^*))$ and rearrange terms

$$\left(\frac{\theta(\sigma\delta_1) - 1}{\theta(\delta_1) - 1} - 1\right) \theta_1^s \left(u(t_1^*) - u(s_1^*)\right) \geq \theta(s) \cdot \left(u(s^*) - u(t^*)\right) \geq 0 \tag{II.1.6}$$

where the last inequality follows from (2) in (II.1.3). The first and the second terms on the left hand side of (II.1.6) are positive, hence $u(t_1^*) - u(s_1^*) \geq 0$, and by monotonicity of $u(\cdot)$, $t_1^* \geq s_1^*$.

Appendix II.2: Payoff Equivalence of g^e-Games and Conventional Theory

Recall that in the g^e-game the initial allocation in the public account is ng^e and $W - g^e$, is in the private account of each player $i \in N$. We first show that each g^e-game is payoff equivalent to the provision game, denoted as g^0-game. Then we use this result and consistency properties to prove statements in Proposition 3.

Provision Game. Let $g \in [0, W]^n$ be a vector of allocations to the public account. Player *i*'s payoff in the provision game is

$$\pi_i^0(g) = W - g_i + \gamma \sum_{k \in N} g_k \tag{II.2.1}$$

We call contribution g_i in the provision game, player *i*'s allocation to the public account.

g^e-Game. Transfers, $x_i \in [-g^e, W - g^e]$ can be made between the two accounts. A negative transfer means moving resource from the public account to a player's private account, whereas a positive transfer means moving resource from own private account to the public account. The consequence of a transfer x_i in g^e-game is a "contribution" of $g^e + x_i$ in the public account, which we call g_i allocation to the public account. The one to one mapping

$$\psi : [-g^e, W - g^e] \to [0, W] \quad \text{s.t.} \quad \psi(x) = g^e + x \tag{II.2.2}$$

between transfers, *x* and *g* allocations will be used to establish payoff equivalence across games. Indeed, for any vector of transfers, $x \in [-g^e, W - g^e]^n$ in a g^e-game, player *i*'s payoff is

$$\pi_i^e(x) = (W - g^e) - x_i + \gamma \left(ng^e + \sum_{k \in N} x_k \right)$$

Use (II.2.2) mapping of *x* transfer vector to *g* allocation vector, $g_i = \psi(x_i) = g^e + x_i$ for all $i \in N$, and verify that *i*'s payoff is exactly the same as the payoff in (II.2.1) in the provision game with contribution vector *g*,

$$\pi_i^e(x) = W - (g^e + x_i) + \gamma \sum_{k=1}^{n} (g^e + x_k) = W - g_i + \gamma \sum_{k=1}^{n} g_k = \pi_i^0(g) \tag{II.2.3}$$

The payoffs in any g^e-game can be written in terms of *g* allocations to the public account, so the *e*-superscripts will be dropped.

Proof of Proposition 3 Let $C = \{c, \cdots, W\}$ and $T(g_{-i}) = \{\pi(g) : g_i \in C\}$ be player *i*'s feasible set in the payoff space when the vector of others' allocations is g_{-i}. If $C^*(g_{-i} \mid g^e)$ denotes the set of player *i*'s best response allocations to the public account then the choice set in the payoff space is $T^*(g_{-i} \mid g^e) = \{\pi(g) : g_i \in C^*(g_{-i} \mid g^e)\}$. When $c = 0$, the feasible set in the payoff space is $S(g_{-i}) = \{\pi(g) : g_i \in \{0, \cdots, W\}\}$ and the choice set is $S^*(g_{-i} \mid g^e) = \{\pi(g) : g_i \in W^*(g_{-i} \mid g^e)\}$. Let $g_i^b(g_{-i} \mid g^e)$ denote *i*'s smallest allocation in the best response set, $W^*(g_{-i} \mid g^e)$.

Part (a). Let $C = \{c, \cdots, W\}$ be a nonbinding contraction, that is $0 < c < \min(g_i^b, \min(g_{-i}))$. By $\pi(g_i^b, g_{-i}) \in T(g_{-i}) \cap S^*(g_{-i} \mid g^e)$ and Consistency, $T^*(g_{-i} \mid g^e) = T(g_{-i}) \cap S^*(g_{-i} \mid g^e)$. Hence, $C^*(g_{-i} \mid g^e) \subseteq W^*(g_{-i} \mid g^e)$, and since every $g_i^* \in W^*(g_{-i} \mid g^e)$ is also from C (as $g_i^* \geq g_i^b > c$), $W^*(g_{-i} \mid g^e) \subseteq C^*(g_{-i} \mid g^e)$.

Part (b). By payoff equivalence, $S^e(x_{-i}) = S^0(g^e + x_{-i}) = S(g_{-i})$. Thus, for any g^e-game, in the payoff space the (g_{-i}- conditional) feasible set is the same as in the provision game, $S(g_{-i})$ and, by Consistency Property, $S^*(g_{-i} \mid 0) = S^*(g_{-i} \mid g^e)$. Hence, the best response allocation set is $W^*(g_{-i} \mid g^e) = W^*(g_{-i} \mid 0)$.

Appendix II.3. Belief-based kindness g-Allocations

<u>Belief-based model of kindness</u>

Player 1's utility (Dufwenberg and Kirchsteiger 2004) at allocation, g_1 and first- and second-order beliefs, $g_{\tilde{2}}$ and $g_{\tilde{\tilde{1}}}$, is

$$U(g_1, g_{\tilde{2}}, g_{\tilde{\tilde{1}}}) = \pi_1(g_1, g_{\tilde{2}}) + Y_1 \kappa(g_1, g_{\tilde{2}}) \lambda(g_{\tilde{2}}, g_{\tilde{\tilde{1}}})$$

where the first term is her material payoff, $\pi_1 = W - g_1 + \gamma(g_1 + g_{\tilde{2}})$, $Y > 0$ is her reciprocity sensitivity parameter, (un) kindness function of player 1 towards player 2 is $\kappa(g_1, g_{\tilde{2}}) = \pi_2(g_{\tilde{2}}, g_1) - \pi_2^e(g_{\tilde{2}})$ and player 1's belief about player 2's (un)kindness is $\lambda(g_{\tilde{2}}, g_{\tilde{\tilde{1}}}) = \pi_1(g_{\tilde{\tilde{1}}}, g_{\tilde{2}}) - \pi_1^e(g_{\tilde{\tilde{1}}})$ where $\pi_i^e(\cdot)$ is player *i*'s equitable payoff.

Equitable payoff as the reference point

Observation 1. In a linear public good game, the equitable payoff corresponds to allocating $g^c = 0.5(W+c)$ to the public account and is the same for all beliefs.

Proof. Let player 1's first-order belief be some $g_2^{\sim}$. Player 2's highest payoff is when player 1 allocates the maximum feasible amount, W

$$\pi_2^h(g_2^{\sim}) = W - g_2^{\sim} + \gamma(g_2^{\sim} + W)$$

whereas the lowest payoff is when player 1 allocates the minimum permissible allocation, c

$$\pi_2^l(g_2^{\sim}) = W - g_2^{\sim} + \gamma(g_2^{\sim} + c).$$

The equitable payoff, the average of the highest and lowest payoffs, is

$$\pi_2^e(g_2^{\sim}) = W - g_2^{\sim} + \gamma(g_2^{\sim} + \frac{W+c}{2})$$

which is the same as player 2's payoff when player 1 contributes $g^c = (W+c)/2$. Similarly, for player 1, $\pi_1^e(g_1^{\approx}) = W - g_1^{\approx} + \gamma(g_1^{\approx} + g^c)$. So, while the level of equitable payoff depends on the first and second order beliefs, it always corresponds to allocating the average of the highest and lowest permissible allocations, $(W+c)/2$ in a linear public good game.

Observation 2. In a linear public good game, second-order beliefs are irrelevant for allocations to the public account.

Proof. Verify that

$$\kappa(g_1, g_2^{\sim}) = \pi_2(g_2^{\sim}, g_1) - \pi_2^e(g_2^{\sim}) = [W - g_2^{\sim} + \gamma(g_2^{\sim} + g_1)] - [W - g_2^{\sim} + \gamma(g_2^{\sim} + g^c)] = \gamma(g_1 - g^c)$$

and $\lambda(g_2^{\sim}, g_1^{\approx}) = \pi_1(g_1^{\approx}, g_2^{\sim}) - \pi_1^e(g_1^{\approx}) = \gamma(g_2^{\sim} - g^c)$. Thus, player 1's utility is

$$U(g_1, g_2^{\sim}, g_1^{\approx}) = W - g_1 + \gamma(g_1 + g_2^{\sim}) + Y_1\gamma^2(g_2^{\sim} - g^c)\cdot(g_1 - g^c)$$

It should be noted that the second order belief, $g_1^{\approx}$ does not appear in player 1's utility.

Observation 3 Contraction has a negative effect on allocations to the public account.

Proof. Differentiating $U(\cdot)$ w.r.t. g_1, $U_{g_1} = (-1+\gamma) + Y_1\gamma^2(g_2^{\sim} - g^c)$ which is positive if and only if $g_2^{\sim} > 0.5(W+c) + (1-\gamma)/Y_1\gamma^2$. Note that the right-hand-side, the smallest level of first-order belief at which player 1 switches from free-riding to full contribution, increases in c.

Specification with a non-linear (concave increasing) u(·) for material payoffs

Player 1's utility is

$$U(g_1, g_2^{\sim}, g_1^{\approx}) = u(W - g_1 + \gamma g_1 + \gamma g_2^{\sim}) + Y_1\gamma^2(g_2^{\sim} - g^c)\cdot(g_1 - g^c)$$

Observation 4. With a nonlinear function of material payoffs:

a. If $g_2^{\sim} < 0.5(W+c)$ then player 1 fully free-rides (allocates c)

b. If $g_2^{\sim} > 0.5(W+c)$ then optimal (interior) allocation increases in other's allocation, $g_2^{\sim}$ and decreases in minimum permissible contribution, c

Proof. Differentiating $U(\cdot)$ w.r.t. g_1, $U_{g_1} = (-1+\gamma)u'(\cdot) + Y_1\gamma^2(g_2^{\sim} - g^c)$ which is negative if $g_2^{\sim} < g^c (= 0.5(W+c))$, hence part a. For part b, note that at an interior solution, $U_{g_1} = 0$, and by the implicit function theorem

$$sign(\frac{\partial g_1}{\partial g_2^{\sim}}) = sign(U_{g_1 g_2^{\sim}}) = sign((\gamma - 1)\gamma u''(\cdot) + Y_1\gamma^2)$$

and

$$sign(\frac{\partial g_1}{\partial c}) = sign(U_{g_1 c}) = sign(-0.5Y_1\gamma^2)$$

Hence, player 1's allocation increases in $g_2^{\sim}$ and decreases in c.

Appendix II.4. Reference Dependence with Loss Aversion

II.4.1 Conventional Loss Aversion Model of Tversky and Kahneman (1991).

For any given other's allocation, g_2, player 1's feasible (payoff) set is $S(g_2)$. We consider two alternative reference points for set $S(g_2)$: (1) the "other's choice conditional" reference point, i.e., the vector of payoffs in $S(g_2)$ before player 1 makes a choice; or (2) the "initial endowment" reference point, i.e., the vector of initial payoffs at the beginning of the game before any player makes a choice.

Reference Point Alternative 1. The reference point is the initial vector of payoffs in set $S(g_2)$ before player 1 makes a choice.

Game Form Effect. In the appropriation game, the payoff vector before player 1 makes her choice is $r^a = (\gamma(W+g_2), W-g_2+\gamma(W+g_2))$. Compared to r^a, any point from $S(g_2)$ is a gain for player 1 as her payoff increases (when she appropriates anything) but a loss for player 2 as other's payoff decreases (compared to player 1 appropriating nothing). Using the TK additive specification (page 1051), when reference point is r^a

$$U^a(\pi_1,\pi_2) = u(\pi_1) - u(r_1^a) + \lambda_2(v(\pi_2) - v(r_2^a))$$

for some concave increasing $u(\cdot)$ and $v(\cdot)$ and some loss aversion parameter $\lambda_2 > 1$. The optimal interior allocation, g_1^a for the appropriation game satisfies the f.o.c.,

$$\lambda_2 v'(\pi_2^a)\gamma = (1-\gamma)u'(\pi_1^a) \tag{II.4.1}$$

In the provision game, the payoff vector before player 1 makes her choice is $r^p = (W+\gamma g_2, W-g_2+\gamma g_2)$. Compared to r^p, any point from $S(g_2)$ is a loss for player 1 (by free-riding condition, $\gamma - 1 < 0$) but a gain for player 2, so

$$U^p(\pi_1,\pi_2) = \lambda_1(u(\pi_1) - u(r_1^p)) + (v(\pi_2) - v(r_2^p))$$

for some loss aversion parameter, $\lambda_1 > 1$. Differentiating w.r.t. g_1, we get

$$U_{g_1}^p = \lambda_1 u'(\pi_1)(\gamma - 1) + \gamma v'(\pi_2) \tag{II.4.2}$$

Evaluate this expression at the optimal allocation, g_1^a in the appropriation game and use (II.4.1)

$$U_{g_1}^p\Big|_{g_1=g_1^a} = \lambda_1 u'(\pi_1^a)(\gamma-1) + \frac{(1-\gamma)}{\lambda_2}u'(\pi_1^a) = \left(\lambda_1 - \frac{1}{\lambda_2}\right)(\gamma-1)u'(\pi_1^a) < 0$$

where the inequality follows from loss aversion, $\lambda_1 > 1 > 1/\lambda_2$, free-riding condition and increasing $u(\cdot)$. Hence, g_1^a is too large to be optimal in provision game, so $g_1^p < g_1^a$, which is at odds with provision games eliciting larger allocations to the public account than payoff-equivalent appropriation games.

Contraction Effect. The low bound on allocations has no effect on the reference point, r^a in the appropriation game, so there is no contraction effect. In the provision game with contraction, $r^{pc} = (W - c + \gamma g_2 + \gamma c, W - g_2 + \gamma g_2 + \gamma c) = r^p + (\gamma - 1, \gamma)c$. If player 1 allocates more than *c*, compared to r^{pc} her payoff decreases (a loss) and player 2's payoff increases (a gain). So, the TK utility at all points from $T(g_2) \subset S(g_2)$ is

$$U^{pc}(\pi_1, \pi_2) = \lambda_1(u(\pi_1) - u(r_1^{pc})) + (v(\pi_2) - v(r_2^{pc}))$$

Recall that contraction is non-binding, so the optimal allocation in the full provision game is also available in the contraction game, and since the f.o.c. is the same as in (II.4.2), the optimal allocation in the full game remains optimal in the contraction game.

Reference Point Alternative 2. The reference point is the payoff vector before any player makes a choice: $r^p = (W, W)$ in the provision game and $r^a = 2\gamma(W, W)$ in the appropriation game.

Contraction Effect. In either game, for any g_2, the reference point remains the same for all $T(g_2 \mid c)$, so the implication is that *non-binding contractions* have no effect on best response allocations.

Game Form Effect. The effect is ambiguous. It suffices to show two cases with opposite predictions. Consider the scenario when, at some other's allocation g_2, player 1's best response (interior) allocation $g_1^*(g_2)$ in the provision game results in payoff vector, P in the gain-gain domain, that is $P = (W - g_1^* + \gamma g_2 + \gamma g_1^*, W - g_2 + \gamma g_2 + \gamma g_1^*) > (W, W)$ and $u'(P_1)(1-\gamma) = \gamma v'(P_2)$ by f.o.c. If this payoff vector, P in the appropriation game is from:

(i) the loss-gain domain then $U_{g_1}^a(P) = \lambda_1 u'(P_1)(\gamma - 1) + \gamma v'(P_2) = u'(P_1)(\gamma - 1)(\lambda_1 - 1) < 0$
(ii) the gain-loss domain then $U_{g_1}^a(P) = u'(P_1)(\gamma - 1) + \lambda_2 \gamma v'(P_2) = u'(P_1)(\gamma - 1)(1 - \lambda_2) > 0$

II.4.2 Reference Dependent Model of Kőszegi and Rabin (2006)

Kőszegi and Rabin (2006, p.1138) define reference dependent utility, $u(\cdot)$ as the sum of the consumption utility, $f(\cdot)$ and gain-loss utility, $h(\cdot \mid r)$, that is

$$V(\pi \mid r) = f(\pi) + h(\pi \mid r) \qquad \text{(II.4.3)}$$

where $f(\cdot)$ and $h(\cdot)$ are additively separable across dimensions. The decision problem is

$$\max_{\pi \in S(G_{-1})} V(\pi \mid r) = \max_{\pi \in S(G_{-1})} \{u(\pi_1) + v(\pi_2) + \mu(u(\pi_1) - u(r_1)) + \mu(v(\pi_2) - v(r_2))\}$$

for some increasing concave $u(\cdot)$, $v(\cdot)$ and $\mu(\cdot)$, a universal gain-loss function: $\mu(0) = 0$ and

$$\begin{aligned} \mu_i(\cdot) &= \mu(u(\pi_i) - u(r_i)), \qquad u(\pi_i) \geq u(r_i) \\ &= -\lambda_i \mu(u(r_i) - u(\pi_i)), \qquad u(\pi_i) < u(r_i) \end{aligned}$$

for some loss aversion parameter, $\lambda_i > 1$, for dimension $i \in \{1, 2\}$.

Reference Point Alternative 1. Player 1's reference points when the other player allocates g_2 are as in II.4.1 and the reference dependent utilities are:

$$V^a(\pi \mid r^a) = [u(\pi_1) + v(\pi_2)] + \mu(u(\pi_1) - u(r_1^a)) - \lambda_2 \mu(v(r_2^a) - v(\pi_2)) \qquad \text{(II.4.4)}$$

$$V^p(\pi \mid r^p) = [u(\pi_1) + v(\pi_2)] - \lambda_1 \mu(u(r_1^p) - u(\pi_1)) + \mu(v(\pi_2) - v(r_2^p)) \qquad \text{(II.4.5)}$$

Game Form Effect. For any given g_2, at the (interior) optimal allocation, g_1^a with consequence material payoffs, π^a in the appropriation game satisfies the f.o.c.,

$$(\lambda_2 \mu_2'(\cdot) + 1)\gamma v'(\pi_2^a) = (\mu_1'(\cdot) + 1)(1 - \gamma)u'(\pi_1^a) \qquad \text{(II.4.6)}$$

Evaluating the derivative of (II.4.5) w.r.t. g_1, in the provision game, at the optimal allocation in the appropriation game (g_1^a) , and substituting from (II.4.6) yield

$$\begin{aligned} V_{g_1}^p(\pi^a \mid r^p) &= (\lambda_1 \mu_1' + 1)(\gamma - 1)u'(\pi_1^a) + (\mu_2' + 1)\gamma v'(\pi_2^a) \\ &= (\lambda_1 \mu_1' + 1)(\gamma - 1)u'(\pi_1^a) + (\mu_2' + 1)\frac{(\mu_1'(\cdot) + 1)(1 - \gamma)u'(\pi_1^a)}{(\lambda_2 \mu_2'(\cdot) + 1)} \\ &= \left[\frac{\lambda_1 \mu_1' + 1}{\mu_1'(\cdot) + 1} - \frac{\mu_2' + 1}{\lambda_2 \mu_2'(\cdot) + 1}\right](\mu_1'(\cdot) + 1)(\gamma - 1)u'(\pi_1^a) < 0 \end{aligned}$$

where the inequality follows from the free riding condition, $\gamma < 1$ and loss aversion (the first term in the square bracket is larger than 1 whereas the second is smaller than 1). Hence, g_1^a is too large to be optimal in the provision game, that is, $g_1^p < g_1^a$.

Contraction Effect. The low bound on allocations has no effect on the reference point in the appropriation game, so the prediction is no contraction effect. In the provision game, the low bound on allocations affects the reference point, as shown in the TK model, but any allocation larger than *c* still results in a loss in own payoff and gain in other's payoff. So, the gain-loss dimensions are preserved, and for linear $\mu(.)$, the f.o.c. in the contraction game is the same as in the full game, that is, the optimal allocation in the full game is also optimal in the contraction game because the non-binding contraction set contains it.

<u>*Reference Point Alternative* 2.</u> The reference point is $r^a = 2\gamma(W,W)$ in the appropriation game and r^p=(W,W) in the provision game with or without contractions.

Game Form Effect. The effect of game form is ambiguous, and the proof is similar to the proof for the TK model.

Contraction Effect. In the provision game, for any nonbinding contraction, *C* of the set of feasible allocations to the public account:

$$\arg\max_{g_1\in[0,W]} U(g_1 \mid g_2, r^p) = \arg\max_{\pi\in S(g_2)} V(\pi \mid r^p) = \arg\max_{\pi\in T(g_2)} V(\pi \mid r^p) = \arg\max_{g_1\in C} U(g_1 \mid g_2, r^p)$$

Similarly for the appropriation game.

Appendix II.5. Morally Monotonic g-Allocations

II.5.A Moral Reference Points across Games

We provide details for moral reference points of player 1 in two-player provision, appropriation, and general g^e-games.

Provision Game. Initially there is 0 in the public account, (*i.e.*, $g^e = 0$) and W in each private account, so initial endowed payoffs for the two players are $\omega_1^p = \omega_2^p = W$. When player 2 allocates g_2 to the public account, player 1's feasible set in the payoff space is $S(g_2)$.[54] Minimal expectations payoffs in $S(g_2)$, from the perspective of player 1, are as follows. The maximum payoff player 1 can get is when he allocates 0 to the public account, in which case player 2 ends up with $2_*(g_2)=(W-g_2)+\gamma g_2$; this is player 2's minimal expectation payoff in $S(g_2)$ from the perspective of player 1. On the other hand, player 2's maximum payoff occurs when player 1 allocates W to the public account, in which case player 1 ends up with $1_*(g_2)=\gamma(W+g_2)$; this is player 1's minimal expectation payoff in $S(g_2)$ from the perspective of player 1. So, from the perspective of player 1, the moral reference point for feasible set $S(g_2)$ in the provision game is

$$\begin{aligned} r_1^1(g_2 \mid g^e=W, c=0) &= (\omega_1^W, 1_*(g_2)) = \left(W, \gamma(W+g_2)\right) \\ r_1^1(g_2 \mid g^e=W, c=0) &= (\omega_2^W, 2_*(g_2)) = \left(W, W-g_2+\gamma g_2\right) \end{aligned} \qquad \text{(II.5.1)}$$

Note that all variables on the right-hand-side of (II.3.1) are observable in an experiment.

<u>*Contractions in Provision Game*</u>. In the presence of a required minimum contribution, c, the maximum payoff player 1 can get is when he allocates c to the public account, in which case player 2 ends up with $2_*(g_2)=(W-g_2)+\gamma(g_2+c)$. On the other hand, player 2's maximum payoff remains when player 1 allocates W to the public account, hence $1_*(g_2)=\gamma(W+g_2)$. Therefore the moral reference point from the perspective of player 1 for feasible set $T(g_2)$ in the provision game with contraction, C=$\{c,\cdots,W\}$ is

$$\begin{aligned} r_1^1(g_2 \mid g^e=W, c>0) &= (\omega_1^W, 1_*(g_2)) = \left(W, \gamma(W+g_2)\right) \\ r_1^1(g_2 \mid g^e=W, c>0) &= (\omega_2^W, 2_*(g_2)) = \left(W, W-g_2+\gamma(g_2+c)\right) \end{aligned} \qquad \text{(II.5.2)}$$

Appropriation Game. Initially there is $2W$ in the public account and 0 in the private account of each player, so initial endowed payoffs of the two players are $\omega^a = 2\gamma(W,W)$**.** Suppose player 2's transfer is $x_2 \in [-W,0]$. Player 1's feasible set in the payoff space is $S^a(x_2)$. The maximum payoff player 1 can get is when he appropriates the maximum allowed (i.e., $x_1=-W$) in which case player 2 ends up with $2_*(x_2)=\gamma(W+x_2)-x_2$; this is player 2's minimal expectation payoff for $S^a(x_2)$ from the perspective of player 1. On the other hand, player 2's maximum payoff

[54] See Figure 1 in the main text for an illustration of $S(5)$, initial endowed payoffs, minimal expectation payoffs and moral reference point.

occurs when player 1 appropriates nothing, in which case player 1 ends up with $1_*(x_2)=\gamma(2W+x_2)$; this is player 1's minimal expectation payoff for $S^a(x_2)$ from the perspective of player 1. Hence, the moral reference point of player 1 in the appropriation game is

$$\begin{aligned} r_1^1(x_2 \mid g^e=0,t=0) &= (\omega_1^0, 1_*(x_2)) = \left(2\gamma W, \gamma(2W+x_2)\right) \\ r_1^1(x_2 \mid g^e=0,t=0) &= (\omega_2^0, 2_*(x_2)) = \left(2\gamma W, \gamma(W+x_2)-x_2\right) \end{aligned}$$

By (II.2.1), $S^a(x_2)=S^p(g_2=W+x_2)=S(g_2)$ and hence, the moral reference point in the appropriation game in terms of *g* allocations left in the public account is

$$\begin{aligned} r_1^1(g_2=W+x_2 \mid g^e=0,c>0) &= (\omega_1^W, 1_*(g_2)) = \left(2\gamma W, \gamma(W+g_2)\right) \\ r_1^1(g_2=W+x_2 \mid g^e=0,c>0) &= (\omega_2^W, 2_*(g_2)) = \left(2\gamma W, W-g_2+\gamma(g_2+c)\right) \end{aligned} \qquad \text{(II.5.3)}$$

Contractions in Appropriation Game. In the presence of a quota, $t(<W)$ on the amount extracted, the maximum payoff player 1 can get is when he takes all he can (i.e., $x_1=-t$) from the public account, in which case player 2 ends up with $2_*(x_2)=\gamma(2W+x_2-t)-x_2$. Player 2's maximum payoff remains when player 1 takes nothing from the public account, hence $1_*(x_2)=\gamma(2W+x_2)$. Player 1's moral reference point in Appropriation game with quota *t*, at opportunity set $T^a(x_2)$ is

$$\begin{aligned} r_1^1(x_2 \mid g^e=0,t>0) &= (\omega_1^0, 1_*(x_2)) = \left(2\gamma W, \gamma(2W+x_2)\right) \\ r_1^1(x_2 \mid g^e=0,t>0) &= (\omega_2^0, 2_*(x_2)) = \left(2\gamma W, \gamma(2W+x_2-t)-x_2\right) \end{aligned}$$

In terms of *g* allocations, contraction $[-t,0]$ in appropriation game is equivalent to allocations from $C=[c, W]$ where $c=W-t$.[55] Substitute $g_2=W+x_2$ and $t=W-c$ in the last two statements

$$\begin{aligned} r_1^1(g_2=W+x_2 \mid g^e=0,t>0) &= \left(2\gamma W, \gamma(W+g_2)\right) \\ r_1^1(g_2=W+x_2 \mid g^e=0,t>0) &= \left(2\gamma W, W-g_2+\gamma(g_2+c)\right) \end{aligned} \qquad \text{(II.5.4)}$$

General g^e-Games. Generalizing the above to a two-player g^e-game is straightforward. The initial distribution of the total resource, $2W$ is $2g^e \in [0,2W]$ in the public account and $W-g^e$ in each player's private account. Each player starts the game with a payoff $W-g^e$ from her private account plus $\gamma(2g^e)$ from the public account, so $\omega^e=W-g^e+2\gamma g^e$. The minimal expectation payoffs for feasible set $T(g_2)$ are

[55] Quota on extractions, $x\in[-t,0]$ implies $W-t\le W+x\le W$, which in terms of *g* allocations is $W-t\le g\le W$.

$$1_*(g_2) = \gamma\left(W + g_2\right) \textit{ and } 2_*(g_2) = W - g_2 + \gamma(g_2 + c)$$

Hence, from the perspective of player 1, the moral reference point for feasible set $T(g_2)$ in a g^e-game with contraction is

$$\begin{aligned} r_1^1(g_2 \mid g^e, c) &= (\omega_1^e, 1_*(g_2)) = \left(W + (2\gamma - 1)g^e, \gamma(W + g_2)\right) \\ r_2^1(g_2 \mid g^e, c) &= (\omega_2^e, 2_*(g_2)) = \left(W + (2\gamma - 1)g^e, W - g_2 + \gamma(g_2 + c)\right) \end{aligned} \qquad \text{(II.5.5)}$$

II.5.B (Best Response) Morally Monotonic Choice

Proof of Proposition 4 Let the g^e-game and the vector of others' allocations, g_{-1} be given.[56] Let $br(g_{-1} \mid g^e, c) = C^*(g_{-1} \mid g^e)$ be player 1's best response allocation set. In the payoff space, player 1's corresponding feasible set is $T(g_{-1}) = \{\pi(g_1, g_{-1}) : g_1 \in C\}$, the choice set is

$$T^*(g_{-1} \mid r^c) = \{\pi(g_1, g_{-1}) : g_1 \in C^*(g_{-1} \mid g^e)\} \qquad \text{(II.5.6)}$$

and player 1's moral reference point, r^c such that $r_i^c = (\omega_i^e, i_*^c)$, where[57]

$$\omega_1^e = \omega_i^e = W - g^e + \gamma n g^e,\ 1_*^c = \gamma(W + G_{-1}) \text{ and } i_*^c = W - g_i + \gamma(G_{-1} + c), \forall i > 1 \qquad \text{(II.5.7)}$$

Part 1. Effect of (nonbinding) constraint *c*. Let $g_1^c \in C$ and $g_1^o \in \{0, \cdots, W\}$ denote the largest player 1's (best response) allocations in $C^*(g_{-1} \mid g^e) \triangleright W^*(g_{-1} \mid g^e)$ and $W^*(g_{-1} \mid g^e)$ in the g^e-game with constraints, *c* and 0 (i.e., no contraction), respectively. We show that $g_1^c \geq g_1^o$. Proof for the smallest allocations is similar. Note that, as others' payoffs increase in g_1, for all $i \in N \setminus \{1\}$

$$\begin{aligned} \pi_i(g_1^o, g_{-1}) &= \max S_i^*(g_{-1} \mid r^o) = \max\{\pi_i : \pi \in S^*(g_{-1} \mid r^o)\} \\ \pi_i(g_1^c, g_{-1}) &= \max T_i^*(g_{-1} \mid r^c) = \max\{\pi_i : \pi \in T^*(g_{-1} \mid r^c)\} \end{aligned} \qquad \text{(II.5.8)}$$

It suffices to show that $\pi_k(g_1^c, g_{-1}) \geq \pi_k(g_1^o, g_{-1})$ for some player k>1, as that together with *k*'s payoff increasing in g_1 require $g_1^c \geq g_1^o$. The proof consists of the following two steps.

---

[56] Without any loss of generality, the proof is written for player 1.

[57] For the full game (no contractions) the moral reference point r^o corresponds to *c*=0. To make reading easier, when there is no contraction, we'll use notation $S(\cdot)$ instead of $T(\cdot \mid c{=}0)$.

Step 1. Consider the following two scenarios, *a* and *d*. In both scenarios, player 1's feasible set in the payoff space is $X = S(g_{-1})$ but the initial endowed payoffs are different. In scenario *a*, the initial endowed payoff is $\omega_i^a = \omega_i^e, \forall i \in N$. In scenario *d*, the initial endowed payoff is $\omega_1^d = \omega_1^e$ for player 1 and $\omega_i^d = \omega_i^e + \gamma c$ for all $i > 1$. Player 1's moral reference points, r^a and r^d, in scenarios *a* and d are: for all $i \in N$, $r_i^a = (\omega_i^a, i_*(X))$ and $r_i^d = (\omega_i^d, i_*(X))$.

For problems (X, r^d) and (X, r^a), $K \neq \varnothing$ as $\delta_i(r^d, r^a) = (\omega_i^d - \omega_i^a) + (i_*^d - i_*^a) = \gamma c, \forall i > 1$ and $\delta_1(r^d, r^a) = (\omega_1^d - \omega_1^a) + (1_*^d - 1_*^a) = 0$. By M-Monotonicity, $X_k^*(r^d) \triangleright X_k^*(r^a)$ for some player $k \in N \setminus \{1\}$, which implies that *k*'s payoffs, $\pi_k^d = \max X_k^*(r^d)$ and $\pi_k^a = \max X_k^*(r^a)$, satisfy

$$\pi_k^d \geq \pi_k^a \tag{II.5.9}$$

Next, $X = S(g_{-1})$ and $r^a = r^o$, so by M-Consistency, $X^*(r^a) = S^*(g_{-1} \mid r^o)$ and by (II.5.8)

$$\pi_k^a = \pi_k(g_1^o, g_{-1}) \tag{II.5.10}$$

Last, there exists some allocation in $\{0, \cdots, W\}$, call it g_1^d, such that $\pi_k^d = \pi_k(g_1^d, g_{-1})$, as by construction, $X = S(g_{-1})$. Hence, by (II.5.9)

$$\pi_k(g_1^d, g_{-1}) \geq \pi_k(g_1^o, g_{-1}) \tag{II.5.11}$$

It follows from (II.5.11) and *k*'s payoff increasing in g_1 that $g_1^d \geq g_1^o$ which together with $g_1^o > c$ (for nonbinding *c*) imply $g_1^d \in C$ and

$$\pi_k^d = \pi_k(g_1^d, g_{-1}) \in T(g_{-1}) \cap X^*(r^d) \tag{II.5.12}$$

Step 2. For feasible problems $(T(g_{-1}), r^c)$ and (X, r^d): (i) $T(g_{-1}) \subseteq X$ as $X = S(g_{-1})$; (ii) $\delta_1(r^c, r^d) = (\omega_1^e - \omega_1^d) + (1_*^c - 1_*^d) = 0$ and $\delta_i(r^c, r^d) = (\omega_i^e - \omega_i^d) + (i_*^c - i_*^d) = -\gamma c + \gamma c = 0, \forall i > 1$; (ii) $T(g_{-1}) \cap X^*(r^d) \neq \varnothing$, by (II.3.12). By M-Consistency, $T^*(g_{-1} \mid r^c) = T(g_{-1}) \cap X^*(r^d)$, hence

$$\pi_k(g_1^d, g_{-1}) \in T^*(g_{-1} \mid r^c) \tag{II.5.13}$$

Finally, $\pi_k(g_1^c, g_{-1}) \geq \pi_k(g_1^o, g_{-1})$ follows from

$$\pi_k(g_1^c, g_{-1}) \geq \pi_k(g_1^d, g_{-1}) \geq \pi_k(g_1^o, g_{-1})$$

where the first inequality follows from (II.5.8) and (II.5.13) and the second one from (II.5.11).

Part 2. Effect of initial g^e. For any given initial (per capita) allocations $g^s > g^t$, let g_1^{bs} and g_1^{bt} denote player 1's smallest (best response) allocations in these g^e-games when others' vector of allocations is g_{-1}. We show that $g_1^{bt} \geq g_1^{bs}$. Proof for the largest allocations is similar. For the two feasible problems, $(S(g_{-i}), r^s)$ and $(S(g_{-i}), r^t)$, the total change in the moral reference point is $\delta_i(r^s, r^t) = (\omega_i^s - \omega_i^t) + (i_*^s - i_*^t) = (n\gamma - 1)(g^s - g^t) > 0$. So K=N, and by M-Monotonicity, $S_1^*(g_{-1} | r^s) \rhd S_1^*(g_{-1} | r^t)$, implying $\pi_1(g_1^{bs}, g_{-1}) \geq \pi_1(g_1^{bt}, g_{-1})$. Hence $g_1^{bs} \leq g_1^{bt}$ as player 1's payoff decreases in own allocation, g_1 (free-riding incentive).

Appendix II.6. Effects of per capita Initial, g^e and Quota, c on Extreme Nash Equilibria

Proof of Proposition 5 *Part* a. By Proposition 4, best responses are invariant to g^e and (nonbinding) c, therefore, Nash equilibrium set is also invariant. *Part* b. We use Tarski (1955) to compare extreme Nash equilibria across g^e. Proof for (nonbinding) quota effect is similar. Let Υ denote the product space, that is $\Upsilon = \times_i \{0, \cdots, W\}$ and $(\Upsilon, \geq)$ denote the lattice with conventional, increasing partial order, ≥. For any initial (per capita) allocation $t = g^e$ in the public account, let $f^t(g) = (f_i^t(g_{-i}) \in \{0, \cdots, W\} \,|\, i \in N)$ where $f_i^t(\cdot)$ is i's largest (best response) allocation, that is

$$f_i^t(g_{-i}) = \max\{g_i \,|\, g_i \in br_i^t(g_{-i})\}$$

Since $f_i^t(g_{-i})$ is increasing in others' allocations and Υ is a complete lattice, the largest Nash equilibrium is[58]

$$\alpha^t = \sup_\Upsilon E^t = \{g \in \Upsilon \,|\, f^t(g) \geq g\}$$

For any two initial (per capita) allocations t and s, such as $t > s$, by Proposition 4, best response largest allocations are smaller in the t-game than in the s-game, which implies $E^t \subseteq E^s$ and

---

[58] See Tarski (1955). Nash set is a subset of E^t and $f(\alpha^t) = \alpha^t$ as follows. Existence of $\alpha^t \in \Upsilon$ follows from $(\Upsilon, \geq)$ being a complete lattice. For all $g \in E^t$, $\alpha^t \geq g$ and increasing $f^t(\cdot)$ imply $f^t(\alpha^t) \geq f^t(g) \leq g$; that is $f^t(\alpha^t)$ is an upper bound of E^t, hence $f^t(\alpha^t) \geq \alpha^t$. By increasing $f^t(\cdot)$, $f^t(f^t(\alpha^t)) \geq f^t(\alpha^t)$ implying $f^t(\alpha^t) \in \mathrm{E}^t$, hence $\alpha^t \geq f^t(\alpha^t)$.

therefore $\sup_{\Upsilon} E^s \geq \sup_{\Upsilon} E^t$. For g^e-effect on the smallest Nash equilibrium, β^t replace $f_i^t(g_{-i})$ with $h_i^t(g_{-i}) = \min br_i^t(g_{-i})$, E^t with $L^t = \{g \in \Upsilon \mid g \geq h^t(g)\}$ and $\beta^t = \inf_{\Upsilon} L^t$.

Appendix II.7. Special Case Objective Function Derivation and Application

Let the (best response) allocation be determined by the maximization of

$$\max_{g_1} U(g_1 \mid r, g_{-1}) = \sum_{i \in N} w_i(r) u(\pi_i) = \frac{\mathrm{e}^{\sigma(r_{11}+r_{12})}}{E} u(\pi_1) + \sum_{k>1} \frac{e^{r_{k1}+r_{k2}}}{E} u(\pi_k) \qquad \text{(II.7.1)}$$

where $\sigma > 1$, $u(x) = (1 - e^{-\alpha x}), \alpha > 0$ and $E = \mathrm{e}^{\sigma(r_{11}+r_{12})} + \sum_{k>1} \mathrm{e}^{r_{k1}+r_{k2}}$.

Recall that, $\pi_i = W - g_i + \gamma(G_{-i} + g_i)$, $r_{i1} = \omega_i^e = W - g^e + \gamma n g^e$, $r_{12} = 1_*^c = \gamma(W + G_{-1})$ and $\forall k > 1$ $r_{k2} = k_*^c = W - g_k + \gamma(G_{-1} + c)$. Player 1's optimal (interior) allocation is determined by

$$\sum_{k>1} \frac{w_k u'(\pi_k)}{w_1 u'(\pi_1)} = \sum_{k>1} \mathrm{e}^{(r_{k1}+r_{k2})-\sigma(r_{11}+r_{12})} \, \mathrm{e}^{-\alpha(\pi_k - \pi_1)} = 1/\gamma - 1 \qquad \text{(II.7.2)}$$

Verify that $\pi_k - \pi_1 = g_1 - g_k$, substitute it in (II.7.2) and solve for g_1 to get

$$g_1^*(\mathrm{g}_{-1}, r) = br(g_{-1} \mid r) = \frac{1}{\alpha}\left(\ln(\frac{\gamma}{1-\gamma}) - \sigma(r_{11} + r_{12}) + \ln(\sum_{k>1} e^{r_{k1}+r_{k2}+\alpha g_k}) \right) \qquad \text{(II.7.3)}$$

and note that:

a. Consistent with Proposition 4.a, $g_1^*(\cdot)$ increases in *c* as for all $k \in N \setminus \{1\}$, r_{k2} increases in *c*

b. Consistent with Proposition 4.b, $g_1^*(\cdot)$ decreases in g^e. Indeed, take any two g^e-games with initial (per capita) allocations, *s* and *t* in the public account such that $s > t$. For all $i \in N$, $r_{i1}^s = r_{i1}^t + (\gamma n - 1)(s - t)$. Use statement (II.7.3), $\sigma > 1, \alpha > 0$ and $\gamma \in (1/n, 1)$ to verify that $g_1^*(\cdot \mid s) - g_1^*(\cdot \mid t) = (1 - \sigma)(n\gamma - 1)(s - t)/\alpha \leq 0.$

Appendix II.8. Questionnaire

Thank you very much for participating in our decision experiment. We would like to ask you a few questions. Your privacy is protected because your name will not appear on this questionnaire or on your decision tables.

1. What year are you in school? Freshman __ Sophomore __ Junior ___ Senior __ Grad. __
2. What is your intended or declared major? ___________________________

3. What is your current grade point average? ________

4. In what year were you born? Year ________

5. What is your gender? Female ____ Male _____

6. What is your race? Asian ___ Black/IIAfrican American ___ White ____ Other ___

8. What is your religious affiliation? ________________ No religion _____

9. Most people would stop and help a person whose car is disabled
Disagree Strongly __ Disagree Slightly __ Agree Slightly __ Agree Strongly __

10. People are usually out for only their own good
Disagree Strongly __ Disagree Slightly __ Agree Slightly __ Agree Strongly __

11. Most people inwardly dislike putting themselves out to help other people
Disagree Strongly __ Disagree Slightly __ Agree Slightly __ Agree Strongly __

12. I have given money to a stranger who needed it (or asked me for it)
Never ____ Once_____ More than once____ Often_____ Very often ___

13. I have done volunteer work for charity
Never ____ Once_____ More than once____ Often_____ Very often ___

14. I have helped a classmate who I did not know that well with an assignment when my knowledge was greater than his or hers
Never ____ Once_____ More than once____ Often_____ Very often ___

15. Do you share your secrets with some of of your close friends?
Never ____ Once_____ More than once____ Often_____ Very often ___