Morally Monotonic Choice in Public Good Games*

By James C. Cox[a], Vjollca Sadiraj[b], and Susan Xu Tang[c]

November 2020

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

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

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

* We thank the National Science Foundation (grant number SES-1658743) for research funding and Caleb Cox for helpful comments and suggestions.

J.C. Cox is the corresponding author.

Morally Monotonic Choice in Public Good Games

Abstract. Decades of robust data from public good games with positive and negative externalities challenges internal consistency axioms that comprise rational choice theory. This paper reports an extension of rational choice theory that incorporates *observable* moral reference points. This morally monotonic choice theory is consistent with data in the literature and has idiosyncratic features that motivate new experimental designs. We report experiments on choices in public good games with positive, negative, and mixed-sign externalities, with and without non-binding quotas on extractions or minimum contributions. Data favors choices predicted by moral monotonicity over choices predicted by: (a) conventional rational choice theory; or (b) conventional reference dependent model of loss aversion.

JEL Classifications: C91, D62, H41

Keywords: choice theory, public goods, externalities, crowding out, moral reference points, experiment

1. Introduction

Theory and behavior for voluntary contributions to public goods is a central topic in public economics. The topic is particularly interesting because decades of robust data from experiments is inconsistent with predictions from conventional theory. Some of this inconsistency can easily be remedied. For example, inconsistency with free riding predictions of the *homo economicus* special case model can be remedied by recognizing that conventional rational choice theory is consistent with other-regarding preferences. But some prominent inconsistencies between conventional theory and data are much deeper.

By "conventional theory" we mean general rational choice theory (e.g., Sen 1971, 1986) and its prominent special cases including conventional preference theory (e.g., Hicks 1946; Samuelson 1947; Debreu 1959), revealed preference theory (e.g., Afriat 1967; Varian 1982), and (consequentialist) social preferences models (e.g., Fehr and Schmidt 1999; Bolton and Ockenfels 2000; Cox and Sadiraj 2007).

Conventional rational choice theory predicts that government contributions to public goods, funded by lump-sum taxation, will crowd out voluntary contributions by the taxed individuals on a dollar-for-dollar basis.[1] This prediction is inconsistent with data from multiple studies.[2] Conventional rational choice theory predicts equivalent choices in payoff-equivalent provision and appropriation games. Specifically, conventional theory implies individuals will contribute the same amount to the public account in a provision game (or public good game with positive externalities) as the agent leaves in the public account in an appropriation game (or public good game with negative externalities). Alternatively, reference dependent theory of Tversky and Kahneman (1991) predicts larger final allocations to the public account in public good games with negative externalities than in games with positive externalities. Data from most studies is inconsistent with both of these predictions; in a large literature on payoff-equivalent provision and appropriation games (Andreoni 1995 and numerous studies reviewed in Text Appendix 1), the final allocation to the public account is generally larger in a provision game than in a payoff-equivalent appropriation game.[3]

These robust behavioral patterns in public goods games violate the internal consistency requirements that characterize conventional theory. But what are we to make of this? Sen (1993, p. 495) wrote: "Internal consistency of choice has been a central concept in demand theory, social choice theory, decision theory, behavioral economics, and related fields. It is argued here that this idea is essentially confused, and there is no way of determining whether a choice function is consistent or not without referring to something external to choice behavior (such as objectives, values, or norms)." The challenge is how to develop choice theory that "refers to something external to choice behavior" while preserving the central feature that makes a theory of choice empirically testable: quantitative restrictions on observable variables.

We extend rational choice theory (Properties α and β in Sen 1971, 1986) to include *observable* moral reference points. This approach of "morally monotonic choice" is applied in Cox et al. (2020) to dictator games with giving and taking opportunities. We here model best

[1] The crowding out prediction of conventional theory is *not* confined to the *homo economicus* special case of the theory. The prediction holds for conventional preference theory, *per se*, whether or not the preferences are selfish or other-regarding. Bernheim (1986) extends the crowding out prediction to distortionary taxes.

[2] See, for examples: Abrams and Schmitz 1978, 1984; Andreoni 1993; Bolton and Katok 1998; Clotfelter 1985; Khanna and Sandler 2000; Kingma 1989; Ribar and Wilhelm 2002.

[3] Contributing any positive amount in a provision game or extracting less than the maximum feasible amount in an appropriation game is, of course, inconsistent with the *homo economicus* special case interpretation of conventional preference theory although consistent with (some) other-regarding preferences. But the robust observations of different choice behavior in provision and appropriation games is inconsistent with rational choice theory, and therefore inconsistent with conventional preference theory, *per se*, whether or not the preferences are selfish or other-regarding. Reciprocity can explain violations of consistency in *sequential* public good games (Cox, Ostrom, Sadiraj, and Walker 2013) but not in simultaneous public good games, which is the research question in this paper.

response choices of morally monotonic agents for application in richer environments characterized by social dilemmas such as provision and appropriation games.

Previous literature on public goods games with positive and negative externalities is reviewed in Text Appendix 1. Section 2 reports implications of rational choice theory for play in payoff-equivalent provision, appropriation and mixed games with and without restrictions on minimum allocations to the public good. Subsection 2.1 reports conventional theory. Subsection 2.2 develops an extension of conventional theory to include observable moral reference points. We explain how this moral monotonicity theory rationalizes the robust result that over extraction is more severe than free riding in payoff equivalent appropriation and provision games. Furthermore, we explain how this moral monotonicity theory rationalizes the robust result that individuals do not let imposed minimum contributions to a public good crowd out their voluntary contributions on a dollar-for-dollar basis.[4] We derive the implications of moral monotonicity for best response choices from *endogenous* contractions of feasible sets not discussed in previous literature on public goods experiments. These endogenous contractions provide (within-subjects) stress tests of null hypotheses based on conventional choice theory and alternative hypotheses that follow from moral monotonicity.

Section 3 uses data from Andreoni (1995) and Khadjavi and Lange (2015) to test null hypotheses from conventional rational choice theory vs. alternative hypotheses based on moral monotonicity theory. Section 3 also uses data from treatments with contractions of feasible sets in Khadjavi and Lange (2015) to directly test null hypotheses from contraction consistency – Property α in Sen (1971, 1986) – vs. alternative hypotheses based on moral monotonicity that incorporate implications of moral reference points. Data from that experiment provides an incomplete test of conventional vs. moral monotonicity theory because of: (1) limited variation in initial endowments of private and public accounts; (2) imposition of feasible set contractions that are exogenously determined, hence possibly binding on choices; and (3) path dependence and possible reciprocity from repeated decisions. Sections 4 and 5 report a new experiment to address such questions.

In section 4 we report design of the new experiment that elicits play of the game and beliefs about play by others to endogenously implement non-binding contractions of feasible sets in a subsequent round. Conventional theory implies such non-binding contractions will not

[4] The warm glow model (Andreoni 1990) is consistent with the incomplete crowding out observed in experiments.

change *total* allocation to the public account, i.e., one-for-one crowding out. In contrast, moral monotonicity theory implies non-binding contractions will increase an individual's total allocation to the public account because of the induced change in observable moral reference points. Design of the new experiment includes provision and appropriation games and several mixed games constructed by systematic variation in initial endowments of private and public accounts. Conventional theory implies play will be the same in all of these games whereas moral monotonicity theory implies distinct play in each game. Section 5 reports tests of null hypotheses from conventional theory vs. alternatives from moral monotonicity theory.

Section 6 reports a maximization approach to analyzing data from Andreoni (1995), Khadjavi and Lange (2015), and our new experiment. The specific objective ("utility") function imposes strong regularity conditions on choices beyond the implications of R-Consistency and Moral Monotonicity Properties that characterize our theory. But the special case application illustrates how traditional maximization methods can apply moral monotonicity theory to data. Section 7 concludes the discussion and reports the implications of warm glow model (Andreoni 1990) and conventional reference dependent model of loss aversion (Tversky and Kahneman 1991).

2. Theory of Play in Provision, Appropriation, and Mixed Games

We begin with reporting, in subsection 2.1, implications of conventional rational choice theory for provision, appropriation, and mixed games with and without non-binding contractions of feasible sets. Best response choice implications of moral monotonicity theory across games are reported in subsection 2.2.

2.1 Conventional Choice Theory

Let T and S denote feasible sets in the payoff space. Let T^* and S^* denote choice sets, the sets of elements chosen from T and S. Rational choice theory (Samuelson 1938; Chernoff 1954; Arrow 1959; Sen, 1971, 1986, 1993) requires that choices satisfy the consistency properties.

CONSISTENCY PROPERTIES. For all feasible sets $T \subseteq S$:

Property α: $x \in S^* \cap T \;\Rightarrow\; x \in T^*$

Property β: $\forall x, y \in T^*,\, x \in S^* \;\Leftrightarrow\; y \in S^*$

Property α states that any allocation x in the choice set of S is also in the choice set of any subset T of S that contains x. Property β states that if T is a subset of S, and intersection of choice sets, T^* and S^*, is not empty, then all of the elements of T^* are elements of S^*.

Properties α and β characterize rational choice for finite sets because they are equivalent to existence of a weak order, that is, an ordering of choices that is complete and transitive (Sen 1971, 1986). As stated in Proposition 1 below, these properties imply equivalence of (best response) choice sets in provision and appropriation games with identical feasible sets of payoffs, as well as invariance of choice sets for certain types of contractions of feasible sets. This invariance implication is used to inform the design of experiments, as explained in section 4.

A general description of the games with social dilemmas that are the subject of our study is as follows. Each player, i (=1,…,n) chooses an allocation (w_i, g_i) of an amount W of a scarce resource between two accounts: w_i to individual i's private account and g_i to the public account that is shared with n-1 other players. As in conventional linear public good games, let $\gamma \in (1/n, 1)$ denote the marginal per capita rate of return from the public account. When the total amount of others' allocation to the public account is G_{-i}, individual i's money payoff is $\pi_i = w_i + \gamma(g_i + G_{-i})$. Player i's allocation of the total resource, W between the two accounts is uniquely determined by her allocation g_i to the public account because, by non-satiation, $w_i + g_i = W$. The distinguishing characteristic of the provision, appropriation and mixed games is the initial, endowed allocation of the resource W across the two accounts. The initial per capita endowment, g^e in the public account uniquely identifies the 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 created; appropriation game, $g^e = W$ where a public good can be destroyed; and mixed games, $g^e \in (0, W)$ where both creation and destruction of a public good are feasible.

Let g_{-i} be a vector of allocations to the public account by players other than player i. Let π denote the vector of payoffs to all players including i. In our g^e-games, for any given vector of others' allocations, g_{-i} the feasible set of individual i (in the money payoff space) is

(1) $S(g_{-i}) = \{\pi(x, g_{-i}) \mid x \in [0, W], g_{-i} \in [0, W]^{n-1}\}$

See Figure 1 (solid line) for a visualization of $S(g_{-i})$ in *two player* games with W = 10 and $\gamma = 0.75$ and when other's choice results in $g_{-i} = 5$ in the public account.[5] If we let $g_i^b = br(g_{-i})$ denote agent *i*'s best-response allocations when others' *n*-vector of allocations to the public account is g_{-i} then the *n*-vector of payoffs, $\pi(b_i^b, g_{-i})$ belongs to the choice set, $S^*(g_{-i})$, that is

(2) $\pi(g_i^b, g_{-i}) \in S^*(g_{-i}) \subseteq S(g_{-i})$

2.1.a Endowed Resource Allocation and Choice

For any given marginal per capita return, γ and total resource, W the first observation is that Properties α and β imply 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 O.1) because her feasible set in the payoff space, $S(g_{-i})$ remains the same for all g^e.[6]

A second observation (see Online Appendix O.1) is that consistency Properties α and β imply player *i*'s choice set, $S^*(g_{-i})$ remains the same if, instead of $[0, W]$, individual *i* is asked to choose from some contracted subset, $C = [c, W] \subseteq [0, W]$ that contains g_{-i} and *i*'s smallest best response allocation, g_i^b for which $\pi(g_i^b, g_{-i}) \in S^*(g_{-i})$. We call these subsets non-binding contractions.[7] Indeed, for any given *c* such that $0 < c < \min(g_i^b, \min(g_{-i}))$ (*) the feasible set in the payoff space is

(3) $T(g_{-i}) = \{\pi(x, g_{-i}) \mid (x, g_{-i}) \in [c, W]^n\}$

[5] That is, the other player takes 5 from the public account in the appropriation game or adds 5 to it in the provision game. In a mixed game with (8,2) *initial* distribution of W = 10, that is 8 in the private account and 2 in the public account, $g_{-i} = 5$ corresponds to adding 3 to the public account.

[6] Figure 1 illustrates that feasible opportunity set, *S*(5) is identical when other's allocation is 5 whether it is consequence of "taking 5 from the public account" in appropriation game (upper dotted line) or "adding 5 to the public account" in provision game (lower dotted line).

[7] In the provision game, contraction sets of interest correspond to required minimum contribution, $c > 0$. Government contribution to a public good financed by lump sum taxation is one way of implementing such a contraction. In the appropriation game, contraction sets of interest correspond to quota on maximum extraction, $t > 0$. The two types of contracted feasible sets are payoff equivalent when $c = W - t$.

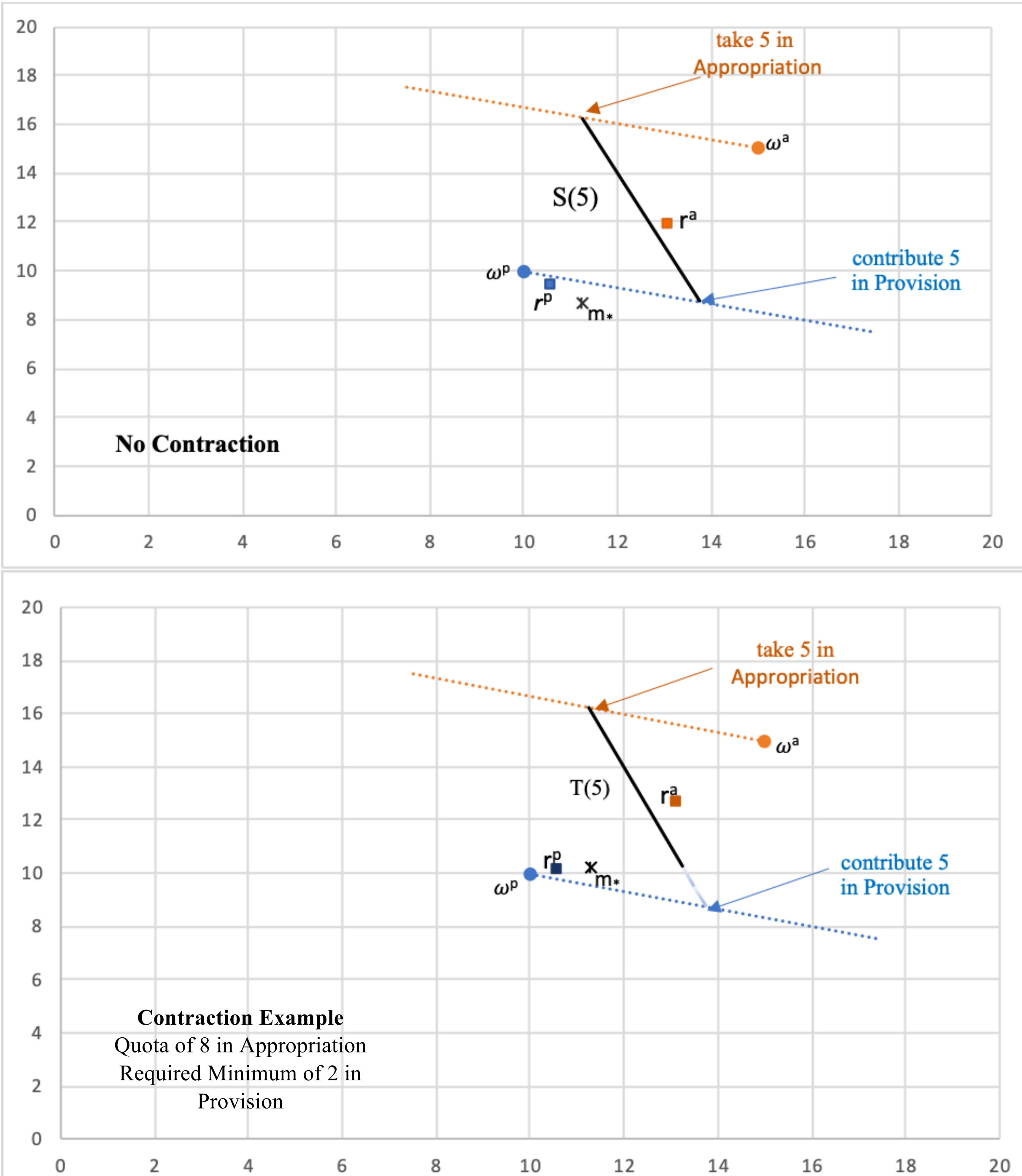

Figure 1. Player 1's Feasible Sets if Player 2 Allocates 5 to the Public Account

Notes: $\gamma = 0.75$. Player 1 and player 2 payoffs on horizontal and vertical axes; feasible sets are discrete points on solid lines, that can result from a belief that player 2 "takes 5" in appropriation, or "contributes 5" in provision; dotted lines correspond to player 2 possible allocations in provision (lower blue line) and appropriation (upper orange line) games; filled circles correspond to initial endowed payoffs; m_* is the minimal expectation point; upper (orange) and lower (blue) rectangles show player 1's moral reference point of the feasible sets (S(5) and T(5)) in appropriation and provision games. Top panel: "No contraction". Bottom Panel: "With Contraction", a quota of 8 in appropriation game or a required minimum contribution of 2 in provision Game.

Note that, $T(g_{-i}) \subset S(g_{-i})$ as $C = [c,W] \subset [0,W]$ and $S^*(g_{-i}) \subset T(g_{-i})$ by specification of the minimum compulsory allocation (*). Therefore, Properties α and β require that $S^*(g_{-i}) = T^*(g_{-i})$, hence *i*'s choice set of allocations in the (non-binding) contraction game remains the same, that is $[0,W]^* = [c,W]^*$, for all *c* that satisfy (*) . As an illustration, consider Figure 1. Suppose player 1's (best response) chosen point at opportunity set, *S*(5) is payoff vector (12,14), which he can ensure by allocating 7 to the public account, bringing the total to 12 units of resource (with value $\gamma 12$). In the contraction game, in the bottom panel, allocations are constrained to [2, 10]. The (12,14) payoff vector is still available in *T*(5) set, so Property α requires it to be in the choice set.

Implications of the two observations are summarized in the following proposition (see Online Appendix O.1 for formal derivation).

Proposition 1. Assume that choice sets in payoff space satisfy Sen's consistency Properties α and β. Let a vector of others' allocations g_{-i} be given and let $g_i^b(g_{-i}) \in [0,W]^*$ be individual *i*'s (best response) smallest *g*-allocation for some g^e-game. Then for all $g_i^*(g_{-1}) \in [0,W]^*$:

a. $g_i^*(g_{-1}) \in [c,W]^*$ for all contractions $[c,W]$ such that $0 \le c < \min(g_i^b, \min(g_{-i}))$

b. $g_i^*(g_{-1}) \in [0,W]^*$ for all initial per capita allocations g^e

The condition on the lower bound, *c* of allocations is to ensure that contractions, *C* are non-binding (see statement (*) above) for player *i*.[8] Part (a) says that individual *i*'s choice is invariant to contractions of the feasible set that constrain allocations to the public account to be no less than non-binding thresholds (that is, amounts *c* that are smaller than agent *i*'s belief about the minimum allocations of others, $\min g_{-i}$ as well as smaller than agent *i*'s smallest allocation, g_i^b

[8] Note that if the lower bound, *c* is binding then by construction individual allocations are weakly increasing in *c*. For example, contraction *T*(5) in Figure 1 is binding for an individual who chooses to allocate 1 in the public account in *S*(5).

from the full set $[0, W]$).[9] Part (b) of Proposition 1 says that choice set of allocations is the same in the provision, appropriation and mixed games.

2.2 Morally Monotonic Choice Theory

Cox et al. (2020) propose a theory of morally monotonic choices, an extension of rational choice theory that incorporates *observable* moral reference points, to model observed choices in dictator games with giving and taking opportunities and in extensive form games with contractions of feasible sets. In this section we derive the implications of an extended moral monotonicity theory for (best response) choices in provision, appropriation, and mixed games and subsequently in section 2.3 apply it to derive implications of efficiency of play in Nash equilibrium.

Theory of morally monotonic choices postulates that choices respond to changes in *observable* moral reference points in a systematic way. First, it is necessary to define moral reference points. The definition of moral reference point incorporates two intuitions into theory of choice: that my ethical constraints on interacting with others in the game we are playing may depend on (a) endowed (or initial) payoffs in the game and (b) the payoff each of us can receive when the others' payoffs are maximized (a.k.a. our "minimal expectation payoffs"). Intuition (a) reflects the idea that larger endowed payoffs entitle one to larger payoffs from playing the game. This initial position (or "property right") effect captures an important feature of everyday life when one is faced by decisions in a fairness game: if we do *not* play the game what would our payoffs be? Intuition (b) reflects the idea that larger ensured minimum payoffs (which correspond to other's most greedy choice) in a game entitle one to larger payoffs from playing the game. This captures a second important feature of everyday life when one is faced by decisions in fairness games: within the environment of our interaction, what is the least each of us can legitimately expect from playing the game? Does my choice offer you a payoff close to your minimum possible (my greediest choice) or significantly more than that amount? Does my choice offer me a payoff much higher than my minimum possible amount or a relatively moderate payoff?

---

[9] This is to ensure that options that are not included in the subset are revealed inferior with respect to both *i*'s beliefs, g_{-i} about others' contributions and player *i*'s smallest choice, g_i^b.

Given the opportunity set S and initial endowments ω, Cox et al. (2020) define moral reference point as a weighted average of minimal expectation payoffs, $m_*(S)$ and initial endowments:

(4) $$r(S,\omega)=0.5m_*(S)+0.5\omega$$

In the following subsection we explain how observable moral reference points are identified in provision and appropriation games.

2.2.a Moral Reference Points in Provision and Appropriation Games

Separate detailed explanations of moral reference points in provision, appropriation, and mixed games with contraction ($c>0$) or without contraction ($c=0$) are contained in Online Appendix O.2.A. We here present a succinct explanation for all g^e-games with or without contraction (i.e., $c\geq 0$). Without any loss of generality we focus on player 1 in a g^e-game. Let the others' vector of transfers, x_{-1} between private and public accounts (in a provision, appropriation, or mixed game) be given. Use mapping $g=\psi(x)=x+g^e$ to convert vector of transfers, x to vector of allocations, g to the public account, and verify that player 1's feasible (payoff) set is $T^e(x_{-1})=T(g_{-1}\,|\,g_{-1}=g^e+x_{-i})$). The maximum feasible payoff of player 1 is when he is the most greedy and his g allocation is the minimum required amount, c which in $T(g_{-1})$ leaves player k with payoff

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

where $G_{-1}=\sum_{j\neq 1}g_j = ng^e+\sum_{j\neq 1}x_j$. The maximum feasible amount of other's payoff, (as a consequence of player 1's choice) is when all player 1's resource W is in the public account. Hence, player 1's minimal expectation payoff is

(6) $$1_*=\gamma\left(W+G_{-1}\right)$$

By the definition of moral reference point, as in (4), player 1's moral reference point at $T(g_{-1})$ in the g^e-game is

(7) $$\begin{aligned} r_1^e &= 0.5(\omega^e+1_*)=0.5(\omega^e+\gamma(W+G_{-1})) \\ r_k^e &= 0.5(\omega^e+k_*)=0.5(\omega^e+W-g_k+\gamma(G_{-1}+c)), \quad \forall k\neq 1 \end{aligned}$$

where endowed payoff is $\omega^e = W - g^e + \gamma n g^e$. Table 1 reports observable minimal expectation payoffs and observable moral reference points for both players in the special case of two-player games. Figure 1 shows player 1's minimal expectation points, m_* and initial endowed payoffs, ω and moral reference points, r in appropriation and provision games without contractions (top panel) and with contractions (bottom panel).

Table 1. Moral Reference Point in a Two-Player g^e - Game

	Player 1 Perspective at $T(g_2)$	Player 2 Perspective at $T(g_1)$
	Minimal Expectation Payoff	
Player 1 Payoffs	$\gamma(W+g_2)$	$(W-g_1)+\gamma(g_1+c)$
Player 2 Payoffs	$(W-g_2)+\gamma(g_2+c)$	$\gamma(W+g_1)$
	Moral Reference Point	
Player 1 Payoffs	$0.5\omega^e+0.5\gamma(W+g_2)$	$0.5\omega^e+0.5\big((W-g_1)+\gamma(g_1+c)\big)$
Player 2 Payoffs	$0.5\omega^e+0.5\big((W-g_2)+\gamma(g_2+c)\big)$	$0.5\omega^e+0.5\gamma(W+g_1)$

Notes: W is total amount of resource, γ is the marginal per capita return from allocations to the public account, $g_i = g^e - x_i^e$ where x_i^e is i's decision in the g^e-game and g_i is the corresponding contribution in the provision game. c = 0 when there is no contraction. $\omega^e = W + g^e(2\gamma - 1)$ is the endowed payoff in g^e-game. Note that both dimensions of moral reference point increase in initial g^e in the public account. Other's dimension of moral reference point increase in c but own dimension does not vary with c.

Replace "1" with "i" in statement (7) to get player i's moral reference point across our g^e games and contractions, $C = [c, W]$ and recall that $n\gamma > 1$ to get:

Observation 1. At any given $T(g_{-i})$, player i's moral reference point has the properties:

(a) The (own) i-coordinate increases in the initial amount, g^e in the public account but does not vary with the contraction level, c;

(b) Each of the coordinates for other players ($k \neq i$) increases in the initial amount, g^e in the public account and increases with contraction level, *c*.

2.2.b Choice Theory with Moral Reference Points

Let *T* and *S* be finite feasible (payoff) sets of a decision maker, player *i*. Let t^r and s^r denote the moral reference points for *T* and *S*, and let $T^*(t^r)$ and $S^*(s^r)$ be player *i*'s choice sets. Cox et al. (2020) incorporate moral reference points into Sen's (1971, 1986) consistency axioms, Properties α and β, as follows. Assume that choice sets are *not empty* and that all feasible sets *T* and *S* satisfy R-consistency properties.

R-CONSISTENCY PROPERTIES. For all $T \subseteq S$, if $t^r = s^r$ then

Property α_R: $x \in S^* \cap T \;\Rightarrow\; x \in T^*$

Property β_R: $\forall x, y \in T^*$, $x \in S^* \;\Leftrightarrow\; y \in S^*$

These statements parallel Sen's Properties α and β but are stated here only for sets with the same moral reference point.

We here state new properties of morally monotonic choices for moral reference point changes. Notation: Let $N = \{1, ..., n\}$; $K = \{k : t_k^r - s_k^r = \delta > 0\} \subseteq N$; and i denotes decision maker's (own) coordinate. A partial order, $>$ of sets *X* and *Y* on *R*, is defined as follows: $X \rhd Y$ if and only if $\inf(X) \geq \inf(Y)$ and $\sup(X) \geq \sup(Y)$.

MORAL MONOTONICITY PROPERTIES. For all $T = S$, if $K \neq \emptyset$ and $t_j^r = s_j^r, \forall j \notin K$, then

Property M_R^a: $T_k^*(t^r) \rhd S_k^*(s^r)$ if $K = \{k\}$

Property M_R^b: $T_i^*(t^r) \rhd S_i^*(s^r)$ if $K = N$

Property M_R^a states that if some player *k* (who may or may not be distinct from player *i*) is favored by the moral reference point then player *i*'s choice set becomes more favorable to player *k*, in the sense that player *k*'s smallest and largest payoffs are both (weakly) larger in player *i*'s choice sets $T^*(t^r)$ than in $S^*(s^r)$.[10] Property M_R^b states that if the moral reference point becomes equally more favorable to all players then player *i*'s choice set favors herself.

2.2.c Example of Morally Monotonic "Utility" Function

We here report an objective function that can be used in constrained maximization applications of moral monotonicity. As demonstrated in Online Appendix O.2.C, the following function of a vector of *n* monetary payoffs π and moral reference points r satisfies R-Consistency and Moral Monotonicity Properties.

(8) $$U_j(\pi \mid r) = \sum_{k=1}^{n} w_k(r)u(\pi_k) \text{ with weights } w_k(r) = \theta(\sigma_k r_k) / \sum_{i=1}^{n} \theta(\sigma_i r_i),\ \sigma_j > 1 = \sigma_{k \neq j},$$

for some increasing $u(\cdot)$ and increasing positive $\theta(\cdot)$ that satisfies $\theta(y+z) = \theta(y)\theta(z)$. Section 6 presents a parametric special case of *U*, that is easy to apply, in which $u(\cdot)$ is a natural exponential function of payoffs and the weights $w(\cdot)$ are normalized natural exponential functions of moral reference points.

2.2.d Implications of Moral Monotonicity for Best Response Choice in Provision, Appropriation and Mixed Games with Contractions

Without any loss of generality we focus on player 1 to derive implications of moral monotonicity for individual choices given others' play in our g^e-games.

Contraction Effect. As noted in subsection 2.2.a (Observation 1), at any given set $T(g_{-1})$, the moral reference point with contraction is *c*-invariant for own dimension but others' coordinates increase in *c*. Moral monotonicity requires that player 1's choice will leave the other player with larger extreme payoffs in the g^e-game with contraction (than in the game without contraction), which player 1 can do by increasing his (best response extreme) allocations to the public account. This feature of moral monotonicity reflects the intuition that a player's resolution of a social dilemma will depend on how much payoff consequences of allocations differ from those for the most selfish feasible action. Imposition of a positive minimum required allocation, *c* raises the reference point for calibrating the extent of free riding from 0 to *c*. Therefore, contrary

[10] Property M_R^a stated here is equivalent to Property M_R in Cox, et al. (2020) for the class of games included in that paper, games in which moral reference point dimensions for two players cannot simultaneously change in the same direction.

to crowding out, moral monotonicity implies a non-binding lower bound on an individual's contributions causes an increase in the individual's total contribution to the public good.[11]

Initial Endowment Effect. As noted in subsection 2.2.a (Observation 1), at any given set $S(g_{-1})$ both dimensions of moral reference point increase in initial g^e. Use statement (7) in subsection 2.2.a to verify that for any two g^e-games with initial (per capita) allocations $g^s > g^t$ in the public account, for all $k = 1,...,n$

(9) $$r_k^s - r_k^t = 0.5(n\gamma - 1)(g^s - g^t) > 0$$

Therefore, the moral reference dimension for player 1 increases by the same amount as for others, which motivates player 1 to aim for larger (extreme) final payoff in the game with the larger g^e. A similar statement holds for players $2,\cdots,n$. These findings are summarized as follows.

Proposition 2. Assume that choice sets in payoff space satisfy R-Properties. Let a vector of others' allocations g_{-i} be given and let $g_i^b(g_{-1}) \in [0,W]^*$ be individual i's (best response) smallest *g*-allocation for some g^e-game. Then

a. $[c,W]^* \rhd [0,W]^*$, for all contractions [*c*, *W*] such that $0 \le c < \min(g_i^b, \min(g_{-i}))$

b. $[0,W]^{*t} \rhd [0,W]^{*s}$, for all g^e-games with initial (per capita) allocations $g^s > g^t$

Proof. See Online Appendix O.2.B.

<u>*2.3 Extreme Nash Equilibria*</u>

Implications of Propositions 1 and 2 for effects of quotas and per capita initial common pool resource on extreme Nash equilibria when best responses are increasing in others' allocations are summarized in Proposition 3. We say best response allocations are increasing in others' allocations if $br_i(\hat{g}_{-i}) > br_i(g_{-i})$ for any vectors of others' allocations such that $(\hat{g}_{-i}) \ge (g_{-i})$.[12]

---

[11] Recall that, in terms of the appropriation game, an extraction quota, *t* is equivalent to a minimum contribution in amount of $W - t$.

[12] $\ge$ is the conventional partial order in R^n; that is, $x \ge y$ if $x_i \ge y_i$, for all *i*=1,…,n.

Proposition 3. If best response allocations are increasing in others' allocations then extreme (the largest and the smallest) Nash equilibrium allocations

a. do not vary with c and g^e for conventional rational choice

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

Proof. See Online Appendix O.3.

2.4. Summary of Testable Implications of Choice Theories

Proposition 1.b tells us that conventional rational choice theory implies the choice set is the same set for all provision ($g^e = 0$), appropriation ($g^e = nW$), and mixed ($g \in (0, nW)$) games with the same feasible set of choices. For the special case in which choice sets are singletons, conventional theory implies the allocation to the public account is the same amount in all payoff-equivalent provision, appropriation, and mixed games. In this way, conventional theory precludes the effects on choices described as "warm glow" and "cold prickle" reported by Andreoni (1995) and many subsequent authors (see Text Appendix 1). In contrast, Proposition 2.b states that moral monotonicity implies a specific partial ordering of choice sets: as the initial endowment in the public account varies from 0 (provision game) to an intermediate value (mixed game) to the total available resource (appropriation game), the supports of choice sets shift towards smaller allocations to the public account. For the special case in which choice sets are singletons, moral monotonicity implies the allocation to the public account will be largest in the provision game, intermediate in a mixed game, and smallest in the appropriation game.

Proposition1.a makes clear that conventional rational choice theory implies the choice set is invariant to non-binding contractions of the feasible set. Such contractions include quotas on extractions from the public account in an appropriation game and floors on minimum contributions in a provision game. Floors on minimum contributions to the public account have been studied in the literature on crowding out of voluntary contributions by public contributions to a public good funded by lump-sum taxation. Proposition 1.a makes clear that complete crowding out is predicted by conventional rational choice theory. In contrast, Proposition 2.a states that moral monotonicity implies a specific partial ordering of choice sets: as the non-binding lower bound on individual allocations to the public account is increased, the supports of choice sets shift towards larger allocations to the public account. Moral monotonicity implies incomplete crowding out. Data from many studies are inconsistent with the prediction of conventional theory but consistent with the prediction from moral monotonicity theory (see footnote 2).

2.5 Tests Reported Below

Section 6 reports a utility function maximization approach to analyzing implications of moral monotonicity for choice in appropriation, provision, and mixed games. That representation of the theory will impose strong regularity conditions on preferences beyond the implications of R-Consistency and Moral Monotonicity Properties. But the special case application will illustrate how traditional maximization methods can apply moral monotonicity theory to data.

In sections 3 and 5 we apply the general theoretical results in Propositions 1 and 2 to data from experiments reported by Andreoni (1995) and Khadjavi and Lange (2015) and to data from the experiment reported herein.

3. Testing Conventional Theory vs. Moral Monotonicity with Existing Data

We here ask whether data from Andreoni's (1995) seminal experiment and data from a recent experiment by Khadjavi and Lange (2015) differ significantly from the predictions of conventional rational choice theory in the direction predicted by moral monotonicity. Similarities and contrasts between these experiments provide robustness checks on our conclusions.

Both experiments include 10 rounds and pay subjects their total earnings from all rounds at the end of a session. Andreoni uses groups of size 5 with random rematching between rounds. Khadjavi and Lange use groups of size 4 with fixed matching over rounds. In both experiments, subjects are informed between rounds of the amount of their own payoff and the total allocation to the public account. Andreoni uses evocative subject instructions that highlight the positive externalities in the provision game and the negative externalities in the appropriation game. Khadjavi and Lange use neutral wording in subject instructions. Andreoni's experiment includes payoff equivalent provision and appropriation games. Khadjavi and Lange's experiment includes these games and also 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 also include a treatment with exogenous contraction of the feasible set in a mixed game that places a lower bound on individual allocations to the public account.

Table 2 reports linear random effects estimation with Andreoni's (1995) data (left two columns) and Khadjavi and Lange's (2015) data (right two columns) where individual allocations to the public account g_i is the dependent variable.

Table 2. Individual Allocations to Public Account in Previous Experiments

Dep. Var: g_i Allocation	**Andreoni (1995)**		**Khadjavi and Lange (2015)**	
	Period 2-10	Period 6-10	Period 2-10	Period 6-10
Total Others' allocation in the previous period, G_{-i}	0.058***	0.053***	0.132***	0.125***
	(0.016)	(0.020)	(0.017)	(0.022)
g^e [-]	-0.131***	-0.110**	-0.103*	-0.090*
	(0.046)	(0.046)	(0.054)	(0.054)
c [+]			0.508***	0.584***
			(0.093)	(0.104)
(D) Period 6-10	-5.408***		-1.665***	
	(1.173)		(0.306)	
Constant	17.330***	11.534***	5.186***	3.388***
	(2.518)	(2.347)	(1.056)	(0.984)
R-Squared (overall)	0.111	0.065	0.350	0.368
Subjects	80	80	40	40
Observations	720	400	160	160

Notes: Random effects estimation. Predicted signs for moral monotonicity in square brackets. Andreoni (1995) experiment consists of groups of five randomly rematched at the beginning of each of 10 rounds, no contractions. In Khadjavi and Lange experiment, groups of four are fixed, play is for 10 rounds, one exogenous contraction. Standard errors in parentheses; clustered at group level for Khdajavi and Lange data. *** p<0.01, ** p<0.05, * p<0.1

Right hand variables include G_{-i}, the total allocation to the public account in the previous period and g^e, the per capita endowment in the public account in the current period, which is 0 in the provision game and the total per capita endowment in the appropriation game. Conventional theory (Proposition 1.b) implies the estimated coefficient for g^e is zero whereas moral monotonicity (Proposition 2.b) implies it is negative. Estimates of the coefficient for g^e are significantly negative with data from both experiments.

A central question that cannot be addressed with Andreoni's data is subjects' responses to contractions of feasible sets. The Khadjavi and Lange experiment includes contractions. As reported in the right two columns of Table 2, the positive estimated effect of lower bound on public account allocations, c is consistent with moral monotonicity theory. But the positive sign could result from design of the experiment because the exogenous contraction bounds, c may be binding for some subjects. More experimentation is needed to test this central difference between conventional theory and moral monotonicity theory.

4. New Experimental Design with Endogenous Contractions

Proposition 1.a, based on conventional rational choice theory, implies that choices are invariant to imposition of non-binding lower bounds on allocations to the public account. In contrast, Proposition 2.b based on moral monotonicity predicts that imposition of such non-binding lower bounds will increase (best response extreme) allocations to the public account because they favor others by increasing their minimal expectations points (that are *observable* features of feasible sets).[13] We here describe our experimental design for a stress test of moral monotonicity theory with non-binding contractions.[14]

We design a two-player experiment with provision, appropriation and mixed games. We observe individuals' chosen allocations in the full game (baseline) and elicit subjects' beliefs about other's allocation. Observed chosen allocations and elicited beliefs are used to inform non-binding 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 for play of conventional rational choice theory vs. moral monotonicity.

---

[13] For the special case of singleton choice sets, conventional theory (resp. moral monotonicity) implies the allocation to the public account is invariant (resp. increasing) with non-binding constraints on minimum allocations to the public account.

[14] This experiment was approved by the Institutional Review Board at Georgia State University.

We give every subject an initial allocation of 10 "tokens" between a public account (with g^e) and a private account (with 10- g^e). Each token has value $1 in the private account and value $1.50 in the public account (or $0.75 for each of two subjects). The classic (contribute only) provision game corresponds to $g^e = 0$ while the payoff equivalent (extract only) appropriation game corresponds to $g^e = 10$. Three payoff equivalent mixed games, with both contributions and extractions being feasible, correspond to g^e = 2 or 5 or 8.

Table 3. 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[a]	[c, 10][b]	[0, 10]	[-2, 8]	[-5, 5]	[-8, 2]	[-10, 0]	[-t, 0][c]
Feasible Allocations in Public Account[a]	[c, 10]	[0, 10]	[0, 10]	[0, 10]	[0, 10]	[0, 10]	[10-t, 10]
Design	Within Subjects		Within Subjects			Within Subjects	
Subjects: Order	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] Possible choices in the experiment include discrete amounts in the intervals.
[b] $c = \min_i \{g_i^* - 1, guess(g_{-i})\}$; [c] $t = \max_i \{t_i^* + 1, guess(t_{-i})\}$. B = [0,10]; C (endogenous) contraction

Our design crosses set contractions with two types of externalities: positive only (for a provision game) or negative only (for an appropriation game). In addition, we have treatments (for mixed games) that allow for actions with both positive and negative externalities. In all treatments, the game is between two players and the public account marginal per capita return, γ is 0.75.[15] Table 3 shows parameter configurations, in terms of initial allocations between the two

[15] The social dilemma requires $0.5 < \gamma < 1$.

accounts, used in each treatment. The decision task consists of allocating $W = 10$ tokens between the private account and public account. Different subjects participated in the provision game, mixed game and appropriation game treatments. Each subject made three decisions without feedback on others' choices and was paired with a different other subject in each of the three decision tasks. In addition, after making each decision every subject was asked to report own expectation ("guess") about the other's decision; correct guesses were paid $2 and incorrect guesses were not paid. One of the three decisions was randomly selected for payoff at the end of each experiment session. After all choices and guesses had been entered, subjects were asked to complete a questionnaire. 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.[16]

In the provision game (with $g^e = 0$), initially all 10 tokens of each player are in his or her private account. The endowed payoff of each subject is $10. The desired allocation between the two accounts can be implemented by transferring (up to 10) tokens from the private to the public account. In the appropriation game (with $g^e = 10$), initially there are a total of 20 tokens in the public account and 0 in each private account, and therefore the endowed payoff of each subject is $15 because each token has value $1.50 in the public account and amounts in that account are split equally. Each subject's desired allocation can be implemented by transferring (up to 10) tokens from the public account to their private account. Similarly, in the mixed $g^e = 5$ game the desired allocation can be achieved by transferring (up to 5) tokens between the two accounts; the endowed payoff here is $12.50 for each subject. Subjects who participated in the mixed games faced tasks in $g^e = 2$, 5, and 8 games in random order.

Provision and 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 "non-binding," as explained below. To control for order effects, half of the subjects participated in BCB design and the other half in CBC design. 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

[16] The questionnaire is available upon request.

choices and beliefs of both players in the previous baseline treatment.[17] To control for "corner set" effect and/or one-sided error the minimum allocation, c was 1 less than the smallest allocation within a pair of subjects.[18] For example, if the allocations of a pair of subjects in the provision game were 3 and 5 and the reported beliefs were 4 and 3 then the set of allocations for the pair in the provision game with contraction would be {2,…,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 beliefs 4 and 3, the contracted feasible set for transfers from the public account to the private account would be {0,1,…,7}.[19]

5. Empirical Play in the New Design

We first look at behavior across provision, appropriation and mixed games with no contractions. Then we analyze behavior in the provision and appropriation games with and without contractions.

5.1 Effects of Endowed Allocations on Choices

Seventy-two subjects participated in (a within-subjects design) mixed-game treatment with each subject making three decisions.[20] 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 made three decisions in appropriation games with and without contraction.[21]

Conventional rational choice theory implies final allocations to the public account (see Proposition 1, part b) are invariant to the endowed allocations, g^e; this is our null hypothesis. Application of moral monotonicity in our games, on the other hand, implies that final and endowed allocations in the public account are inversely related (see Proposition 2, part b); this is our alternative hypothesis. Table 4 reports estimated coefficients with our new data for model

[17] In a CBC session, the contraction sets used in the first C task are the same as in a preceding BCB session.
[18] Exceptions to the "$1 less" criterion are when choices 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 [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 [5,10].
[19] Note that this set, described in terms of the number of tokens allowed to be allocated to the public account (which is our variable of interest and the focus of the data analysis), is {3, 4,…,10}.
[20] That is, one decision in each of the 2-game, 5- and 8-game; the order of the three decision tasks was randomized across subjects.
[21] That is, two (resp. one) decisions in a full (resp. contracted) game or one (resp. two) decisions in a contracted (resp. full) game for ACA or CAC treatments.

specifications similar to the ones reported in Table 2 (for Khadjavi and Lange data). We observe best response allocations that are: (*i*) increasing in guessed other's *g* allocation; (*ii*) decreasing in per capita endowed allocation of the public account, g^e; and (*iii*) increasing in non-binding quotas, *c*.[22] Results (*ii*) and (*iii*) are inconsistent with conventional theory but consistent with moral monotonicity.

Table 4. New Experiment: Best Response Allocations to the Public Account

Dep. Variable: g_i Allocation	**(1)**	**(2)**
Guessed Other's allocation	0.589***	0.568***
	(0.050)	(0.049)
g^e [-]	-0.048*	-0.047*
	(0.028)	(0.028)
c [+]	0.356***	0.404***
	(0.073)	(0.068)
Constant	1.089***	0.959
	(0.226)	(0.945)
Demographics	no	yes
R-Squared (overall)	0.455	0.485
Subjects	232	232
Observations	696	606

Notes: Linear estimators with standard errors clustered at subject level. 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 in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Note that positive estimated coefficients for others' contributions, as reported in Table 2, reveal increasing best response allocations. So Table 3 results (*ii*) and (*iii*) are consistent with implications of moral monotonicity stated in Proposition 3, that extreme equilibria decrease in initial allocation of the public account but increase in (non-binding) quota.

[22] Signs of tobit estimates are the same.

In the following two subsections we look more closely at individual data.

5.1.a Types of Externalities and Choice.

When feasible allocations consist of integers from [0,10], average number of tokens allocated to the public account in provision, mixed and appropriation games are, respectively, 4.01, 3.64 and 3.09,[23] suggesting adverse effect of initial per capita allocation to the public account on resolution of social dilemmas. For free-riding, measured as observed public account allocations of 0 or 1, the provision game elicits least free-riding (30.13%) whereas the appropriation game elicits the most free-riding (52.2%); the free-riding figure for the mixed games is between (42.59%).[24] For statistical inferences we use Kolmogorov-Smirnov test for distributions of public account allocations and Pearson chi2 test for free riding behavior.[25] Choices 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;

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

(iii) Similar public account allocations (p-values=0.384) but less free-riding (p-value=0.075) in mixed than 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).*

Result 1 is inconsistent with conventional rational choice theory but consistent with moral monotonicity.

---

[23] 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.

[24] Figures (in %) for full free-riding (i.e., $g = 0$, allowing for no decision errors) are: 21 (provision), 39 (mixed) and 48 (appropriation).

[25] To ensure independence, when a subject made more than one decision per treatment (e.g. in a BCB session subjects are making two choices in appropriation game), tests are applied to the average of the subject's g allocations. Use of all data (rather than average choices at the individual level) in our tests, produces similar results but p-values are smaller. For the distributions of g allocations in mixed and provision games the p-value is 0.007 (*KS* test) whereas for free-riders, p-value is 0.00 (Pearson chi2 test).

5.1.b Endowed per capita g^e Allocation Effects in Mixed Games (Within-Subjects Analysis)

There is some variation in the means (3.9, 4.2 and 5.1) of guessed other's public account allocations with respect to initial endowed allocations, g^e. Propositions 1 and 2 provide statements on (best response) allocation choice implications that follow from conventional choice theory and moral monotonicity conditional on the other's g allocation. To test the empirical validity of these statements we analyze observed g allocations of subjects whose guesses about others' allocations did not change with initial allocation g^e.[26] We constructed a new variable, Δg: the difference between the chosen g allocations observed for different initial g^e allocations (conditional on the guess not changing). For each subject, $\Delta g = g^i - g^j$, where superscripts $i<j$ denote the g^e values from {2,5,8}; that is, Δg is the difference between the public account allocation in the *i*-game and the allocation in the *j*-game. The null hypothesis that follows from Properties α and β of conventional theory is $\Delta g = 0$ (Proposition 1, part a) whereas the alternative hypothesis that follows from moral monotonicity is $\Delta g > 0$ (Proposition 2, part a). The mean of Δg is 0.782 (95% C.I.=[-0.05,1.61]) and the (conventional theory) null hypothesis is rejected by the *t*-test (one-sided p-value=0.032) in favor of the (moral monotonicity) alternative hypothesis.[27] 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*.

Result 2 is inconsistent with conventional rational choice theory but consistent with moral monotonicity.

5.2 Contraction Effects

For any given allocation by the other player, conventional theory requires that choices in the provision game or appropriation game are invariant to a *nonbinding* contraction – a contraction set that contains choices of both players and their beliefs about choices by others. In contrast,

---

[26] There are 56 such choices from 32 (out of 72) subjects.

[27] If we don't include *selfish* subjects (subjects who *always* allocated 0 in the public account), the mean of Δg is 1.229 (95% C.I. =[-0.08, 2.53])

moral monotonicity predicts g allocation is increasing in c for nonbinding contractions $C = \{c,...,10\}$ such that $c > 0$. Tests of these predictions by within-subjects data analysis are reported next.

We constructed a new variable, Δg_i^{cb} that takes its values according to the difference between the subject's observed allocations in the public account from the (non-binding) contracted set, $C=\{c,\ldots,10\}$ and the full set, $B=\{0,\ldots,10\}$. The null hypothesis from conventional theory is that Δg_i^{cb} values are drawn from a distribution with mean 0, provided that the guess of other's contribution did not change. For such cases (that is, subjects with unchanged guesses), the mean of Δg_i^{cb} is significantly larger than 0 in the provision game but not in the appropriation game.[28] As a further check that the preceding test is picking up (full vs. contracted game) treatment effects rather than (first vs. third round) decision-order effects, we also looked at Δg_i^{bb}, the within-subjects difference in g allocations in tasks in which subjects faced the full set, $B=\{0,\ldots,10\}$ in both first and third rounds. Both conventional theory and moral monotonicity require the mean of the distribution of Δg_i^{bb} to be 0. Data fail to reject this null hypothesis as the mean is -0.06 (95% C.I.= [-0.31,0.18], p-value=0.607 (*t*-test)).[29] Our next result:

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

Result 3 is inconsistent with conventional rational choice theory but consistent with moral monotonicity.

6. Maximization Approach to Testing Conventional and Moral Monotonicity Theory

We here report results for a special case of moral monotonicity theory using a tractable parametric objective function.

[28] The 95% C.I. is [0.13, 1.71] (p-value = 0.02; *t*-test) in the provision game and [-0.60, 1.78] (p-value = 0.32, *t*-test) in the appropriation game.

[29] Provision game: 0.2 (mean), p-value = 0.44 (*t*-test); appropriation game: -0.14 (mean), p-value=0.34 (*t*-test).

6.a Special Case Morally Monotonic Best Response Allocations

Without loss of generality, consider choices by agent 1. Assume a parametric form of objective function (8) in which $u(\pi) = 0.5(1 - e^{-\alpha\pi})$, $\theta(r_1) = e^{\sigma r_1}$ and for all $i > 1$, $\theta(r_i) = e^{r_i}$. Maximization of the parametric function yields necessary condition:

(9) $$\sum_{i>1} \mathrm{e}^{r_i - \sigma r_1}\, \mathrm{e}^{-\alpha(\pi_i - \pi_1)} = (1-\gamma)/\gamma$$

Note that $\pi_i - \pi_1 = g_1 - g_i$, substitute it in (9) and solve for g_1 to get

(10) $$g_1^{br} = br(g_{-1} \mid r) = \frac{1}{\alpha}\left(\ln(\frac{\gamma}{1-\gamma}) - \sigma r_1 + \ln(\sum_{i>1} e^{r_i + \alpha g_i}) \right)$$

Details of the derivation of (10) are reported in Online Appendix O.4. It is also demonstrated in that appendix that (10) implies chosen points consistent with the general-case Propositions 2 and 3 in section 2.

6.b Analysis of Experimental Data

We use statement (10) in our estimation of parameters for α and σ using data from Andreoni (1995) and Khadjavi and Lange (2015) and the experiment reported herein. In the Andreoni experiment and the Khadjavi and Lange experiment a subject knows only total allocations by others in the public fund at the end of each round. Therefore, for empirical estimation we assume that, at the beginning of each period *t*+1, $g_{-1t} = G_{-1t}/(n-1)$ for all other players. Hence, statement (10) simplifies to

(11) $$g_{1(t+1)}^{br} = \frac{1}{\alpha}\left(\ln(\frac{\gamma}{1-\gamma}) - \sigma r_{1t} + \ln(n-1) + r_{-1t} + \alpha g_{-1t} \right)$$

where moral reference point specification is as in statement (7). Table 5 reports nonlinear least square estimates of α and σ for all data, as well as separately for games without contraction as in Khadjavi and Lange's experiment contraction is exogenous (and therefore can be binding for some subjects). Estimated parameter for σ is significantly greater than 1 with data from each of the experiments, which is consistent with moral monotonicity.

Table 5. Non-linear Least Squares Estimates for Parametric Objective Function

Parameters	Andreoni (1995)[1]	K&L (2015)[a]		New Experiment	
	All Data	All Data	No Contraction	All Data	No Contraction
σ	1.12***	1.19***	1.22***	1.072***	1.079***
	(0.013)	(0.024)	(0.027)	(0.016)	(0.020)
	[1.10, 1.15]	[1.14, 1.24]	[1.17, 1.27]	[1.04,1.10]	[1.04, 1.12]
α	0.86***	1.72***	1.54***	1.389***	1.415***
	(0.116)	(0.245)	(0.229)	(0.175)	(0.195)
	[0.63, 1.09]	[1.23, 2.20]	[1.09, 2.00]	[1.04, 1.73]	[1.03, 1.80]
Observatio	720	1440	1080	696	554
R-squared	0.41	0.69	0.53	0.745	0.667
Clusters	80	160	120	232	232

Notes: [a]Round 1 data are not included as there is no information on others contributions. Robust standard errors in parentheses. 95% Confidence Interval in square brackets.

7. Conclusion

We respond to Sen's (1993) call for extending choice theory beyond exclusive focus on internal consistency axioms to include considerations external to choice behavior. Our approach incorporates observable moral reference points into contraction consistency properties that define the theory. Moral monotonicity shows promise of wide applicability. In Cox, et al. (2020) an initial development of this moral monotonicity theory was applied to behavior in dictator games with give and take opportunities and second mover play in extensive form games with contractions of feasible sets including investment, moonlighting, carrot, stick and carrot/stick games. We here extend moral monotonicity theory and use it to predict the effect of changes in the environment (type of externality, contractions) on best response choices in public good games, and use it to derive implications for efficiency of (Nash) equilibrium play. We explain that moral monotonicity can rationalize decades of robust data from public good games that is inconsistent with conventional theory.

Conventional theory predicts one-for-one crowding out of voluntary contributions by imposed minimum contributions to a public good. This prediction is inconsistent with data from most studies in a large literature. In contrast, moral monotonicity implies non-binding lower bound on an individual's contribution causes an increase in the individual's total contribution to the public good, as has generally been observed. Warm glow model (Andreoni 1990, p. 469) makes this same prediction. Conventional theory predicts allocations to a public account will be the same in payoff-equivalent provision and appropriation games. Alternatively, reference dependent theory of Tversky and Kahneman (1991) predicts smaller allocation to the public account in a provision game than in a payoff-equivalent appropriation game, as shown in Online Appendix O.5. Predictions from both theories are inconsistent with data from a large literature. In contrast, moral monotonicity implies larger allocation to the public account in a provision game than in a payoff-equivalent appropriation game, as has been reported in Andreoni (1995), Khadjavi and Lange (2015), and many papers discussed in Text Appendix 1.[30]

We report an experiment with payoff-equivalent provision, appropriation, and mixed games that discriminates between null hypotheses implied by conventional theory and one-sided alternatives provided by moral monotonicity. A novel feature of our experiment is *endogenous* (*non-binding*) contractions of feasible sets that contain a subject's previous choice as well as belief about other's choice as interior points. Observed play and elicited beliefs are used to inform contractions of the sets of allocations that exclude only alternatives that have not been chosen nor believed in being chosen by subjects that are matched in a subsequent play of a contracted game. Conventional choice theory predicts that such exclusion of "irrelevant alternatives" will have no effect on chosen allocations. In contrast, moral monotonicity predicts that the non-binding constraints on choices embodied in the contractions will affect choices because they change players' moral reference points.

Data from provision, appropriation and mixed games, with and without contractions, is mostly inconsistent with predictions from conventional theory but consistent with predictions from moral monotonicity.

[30] This empirical result was interpreted in Andreoni (1995) as reflecting effects of warm glow (in the provision game) vs. cold prickle (in the appropriation game) but a formal model was not reported.

Text Appendix 1: Related Literature on Payoff-Equivalent 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 choices 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).[31] 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, 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

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

significant effects of punishments or reported emotions.[32] 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). 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 central result in Andreoni 1995) and that (2) their mixed frame data does not differ significantly from data for their provision game.

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

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 contraction restrictions of rational choice theory (Sen 1971, 1986). While the Khadjavi and Lange design allows for *exogenous* contraction in the mixed game our design introduces endogenous contractions known to include previous choices in (provision or appropriation) games in addition to elicited beliefs about others' choices. Such endogenous contractions are essential to ascertaining whether behavior in provision, appropriation, and mixed games exhibits monotonicity in moral reference points.

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. Int. Econ. Rev. 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. Q. J. Econ. 110, 1–21. https://doi.org/10.2307/2118508

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. Am. Econ. Rev. 90, 166–193. https://doi.org/10.1257/aer.90.1.166

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

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. Econ. Lett. 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., 2020. Moral Costs and Rational Choice: Theory and Experimental Evidence. Exp. Econ. Cent. Work. Pap. Ser.

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

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

Cubitt, R.P., Drouvelis, M., Gächter, S., 2011. Framing and Free Riding: Emotional Responses and Punishment in Social Dilemma Games. Exp. Econ. 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., Gächter, S., Hennig-Schmidt, H., 2011. The Framing of Games and the Psychology of Play. Games Econ. Behav. 73, 459–478. https://doi.org/10.1016/J.GEB.2011.02.003

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

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

Fujimoto, H., Park, E.-S., 2010. Framing Effects and Gender Differences in Voluntary Public Goods Provision Experiments. J. Socio. Econ. 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.

Khadjavi, M., Lange, A., 2015. Doing Good or Doing Harm: Experimental Evidence on Giving and Taking in Public Good Games. Exp. Econ. 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.

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

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

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

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

Sen, A.K., 1971. Choice Functions and Revealed Preference. Rev. Econ. Stud. 38, 307–317. https://doi.org/10.2307/2296384

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

Sonnemans, J., Schram, A., Offerman, T. (1998). Public good provision and public good prevention: the effect of framing. J. Econ. Behav. Organ. 34, 143-161.

Tarski, Alfred, 1955. A Lattice-Theoretical Fixpoint Theorem and Its Applications. Pacific J. Math. 5, 285-309.

Tversky, Amos, and Daniel Kahneman, 1991. Loss aversion in Riskless Choice: A Reference-Dependent Model. Q. J. Econ. 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. J. Environ. Econ. Manage. 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

Appendix O.1: Payoff Equivalence of g^e-Games and Conventional Theory

Recall that, the g^e-game means the initial, endowed allocations are ng^e to the public account and $W-g^e$ to the private account of each player $i=1,\ldots,n$. We first show that g^e-games are payoff equivalent to the provision game, which is the g^0-game. Then we use this result and consistency axioms to prove statements in Proposition 1.

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

$$\pi_i^0(g)=W-g_i+\gamma\sum_{k=1}^{n}g_k$$

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

$$\psi:[-g^e,W-g^e]\to[0,W] \text{ s.t. } \psi(x)=g^e+x \qquad \text{(A1.1)}$$

between transfers, x and g allocations will be used to establish payoff equivalence across games.[33] Indeed, for any vector of transfers, $x\in[-g^e,W-g^e]^n$ in a g^e-game, individual *i*'s payoff is

$$\pi_i^e(x)=(W-g^e)-x_i+\gamma(ng^e+\sum_{k=1}^{n}x_k)$$

Use (A1.1) mapping of *x* transfer vector to *g* allocation vector, $g_i=\psi(x_i)=g^e+x_i$ for all $i=1,\ldots,n$, and verify that *i*'s payoff is exactly the same as the payoff in the provision game with contribution vector *g*,

[33] In a g^e-game, verify that when individual *i* takes all he can (i.e., $x_i=-g^e$) from the public account, then his *g*-allocation is 0, which agrees with $\psi(-g^e)=0$. When *i* adds all he can (i.e., $x_i=W-g^e$) to the public account, then *i*'s *g*-allocation is W (the initial g^e plus the transfer), which agrees with $\psi(W-g^e)=W$.

$$\pi_i^e(x) = W - (g^e + x_i) + \gamma \sum_{k=1}^{n} (g^e + x_k) = W - g_i + \gamma \sum_{k=1}^{n} g_k = \pi_i^0(g) \qquad \text{(A1.2)}$$

Hence, we write payoffs in terms of g allocations and the e-superscripts will be dropped. We use $*$-superscripts to denote choice sets, that is, X^* denotes the choice set when the opportunity set is X.

Proposition 1. Assume that choice sets in payoff space satisfy Properties α and β. Let a vector of others' allocations g_{-i} be given and let $g_i^b(g_{-i}) \in [0,W]^*$ be individual i's (best response) smallest g-allocation for some g^e-game. Then

(i) $g_i^b(g_{-i}) \in [c,W]^*$ for all feasible sets $[c,W]$ such that $0 \le c < \min(g_i^b, \min(g_{-i}))$

(ii) $g_i^b(g_{-i}) \in [0,W]^*$ for all g^e-games.

Proof. Let $T(g_{-i}) = \{\pi(g) : g_i \in [c,W]\}$ be player 1's feasible set in the final payoff space when the vector of others' allocations is g_{-i} and let $br(g_{-i} \mid g^e, c) = [c,W]^*$ be player i's best response choice correspondence. Hence, in the payoff space $T^*(g_{-i}) = \{\pi(g) : g_i \in [c,W]^*\}$. To simplify notation, we use S when c=0 and T when c>0 and, as the vector of others' allocations, g_{-i} is given, we use g_i^b to refer to elements from choice set, $br(g_{-i} \mid g^e, c)$.

Part (a). By notation, if $g_i^b \in [0,\mathrm{W}]^*$ then $\pi(g_i^b, g_{-i}) \in S^*(g_{-i})$. It follows from the supposition, $c \in [0, \min(g_i^b, g_{-i})]$ that $g_i^b \in [c,W]$, and therefore $\pi(g_i^b, g_{-i}) \in T(g_{-i}) \subset S(g_{-i})$. By Properties α and β, $\pi(g_i^b, g_{-i}) \in T^*(g_{-i})$, hence $g_i^b \in [c,W]^*$.

Part (b). By payoff equivalence, $S^e(x_{-i}) = S^0(g^e + x_{-i}) = S(g_{-i})$. Thus, across g^e-games, in the payoff space the feasible set is always $S(g_{-i})$ and the choice set is always $S^*(g_{-i})$. Hence, allocation choice set $[0,W]^*$ is invariant to initial per capita allocation g^e in the public account.

Appendix O.2. Morally Monotonic g-Allocations

2.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 $\omega^p = (W, W)$. When player 2 allocates g_2 to the public account, player 1's feasible set in the payoff space is $S(g_2)$.[34] 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 ($g_2$-conditional) minimal expectation payoff 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 ($g_2$-conditional) minimal expectation payoff from the perspective of player 1. Hence, moral reference point for player 1 at opportunity set $S(g_2)$ in the provision game is

$$r^p(g_2) = 0.5m_*(g_2) + 0.5\omega^p = 0.5\left(W + \gamma(W + g_2), 2W - g_2 + \gamma g_2\right) \qquad \text{(A2.1)}$$

Note that all variables on the right-hand-side of (A2.1) 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 for player 1 at opportunity set $T(g_2)$ in the provision game with contraction, $[c, W]$ is

$$r^{pc}(g_2) = 0.5m_*(g_2) + 0.5\omega^p = 0.5\left(W + \gamma(W + g_2), 2W - g_2 + \gamma(g_2 + c)\right) \qquad \text{(A2.2)}$$

---

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

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 at $S^a(x_2)$ from the perspective of player 1. On the other hand, player 2's maximum payoff 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 at $S^a(x_2)$ from the perspective of player 1. Hence, at $S^a(x_2)$ moral reference point of player 1 in the appropriation game is

$$r^a(x_2) = 0.5\big(\gamma(4W + x_2), -x_2 + \gamma(3W + x_2)\big)$$

Our interest is on individual choices in Appropriation and Provision games when the individual faces the same set of final payoffs. By (A1.1), $S^a(x_2) = S^p(g_2 = W + x_2) = S(g_2)$ and moral reference point in appropriation game in terms of *g* allocations is

$$r^a(g_2 = W + x_2) = 0.5\big(\gamma(3W + g_2), W - g_2 + \gamma(2W + g_2)\big) \qquad \text{(A2.3)}$$

<u>Contractions in Appropriation Game</u>. 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

$$r^a(x_2) = 0.5\big(\gamma(4W + x_2), -x_2 + \gamma(4W + x_2 - t\big)$$

In terms of *g* allocations, contraction, $[-t, 0]$ in appropriation game is equivalent to allocations from *C*=[*c*, *W*] where $c = W - t$.[35] Substitute $g_2 = W + x_2$ in $r^a(x_2)$ to get

$$r^a(g_2 = W + x_2) = 0.5\big(\gamma(3W + g_2), W - g_2 + \gamma(2W + g_2 + c)\big) \qquad \text{(A2.4)}$$

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

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 at opportunity set $T(g_2)$, remain

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

Hence, player 1's moral reference point in a g^e-game with contraction is

$$r^{ec}(g_2) = 0.5\big(W - g^e + \gamma(2g^e + W + g_2),\ 2W - g^e - g_2 + \gamma(2g^e + g_2 + c)\big) \qquad \text{(A2.5)}$$

It should be noted that contraction has no effect on own coordinate of the moral reference point but it favors the other player as player 2's coordinate increases in *c*. Both coordinates increase in initial per capita allocation, g^e in public account as $2\gamma - 1 > 0$. Note that all variables on the right-hand-side of (A2.5) are observable in an experiment.

2.B (Best Response) Morally Monotonic Choice

Proof of Proposition 2

Let the vector of others' allocations, g_{-i} be given. The contracted subset, $C = [c,W] \subseteq [0,W]$ is chosen to contain g_{-i} and *i*'s smallest allocation, g_i^b for which $\pi(g_i^b, g_{-i}) \in S^*(g_{-i})$. As in Proposition 1, let $T(g_{-i}) = \{\pi(g) : g_i \in [c,W]\}$ be player *i*'s feasible opportunity set in the final payoff space when others' vector of allocations is g_{-i}, and let $br(g_{-i} \mid g^e, c) = [c,W]^*$ be player *i*'s best response choice correspondence. Hence, in the payoff space, $T^*(g_{-i}) = \{\pi(g) : g_i \in [c,W]^*\}$. Without any loss of generality, the proof is written for player 1.

Part 1. Effect of (non-binding) constraint *c*. We show that the largest individual 1's allocation choice in the game with contraction is (weakly) larger than the largest allocation in the full game. Proof for the smallest allocation choice is similar. Let $g_1^{bo} \in [0,W]^*$ be the largest allocation choice in the full game. This implies that, in the payoff space, $\pi(g_1^{bo}, g_{-1}) \in S^*(g_{-1})$ and, as others' payoffs increase in g_1,

$$\pi_i(g_1^{bo}, g_{-1}) = \max\{\pi_i : \pi \in S^*(g_{-1})\}$$

It follows from $0 \le c < \min(g_1^b, \min(g_{-1}))$ that $g_1^{bo} \in [c,W]$, which in the payoff space, implies that $\pi(g_1^{bo}, g_{-1}) \in T(g_{-1})$. Let r^c, r^o denote moral reference points of $T_{g_{-1}}$, *and* $S_{g_{-1}}$ respectively. Use (5) to verify that for all $i > 1$, $r_i^c - r_i^o = \gamma c > 0$ and $r_1^c = r_1^0$.

Step 1. Consider the following scenario. Player 1's feasible set is [0,W], and therefore in payoff space the feasible set is $T^c(g_{-1}) = S(g_{-1})$ but the moral reference point is r_1^c. To get this scenario have $\omega = 2r^c - m_*(S(g_{-i}))$ as initial endowed payoffs. By Property M_R^b, $T_i^{c*}(r^c) \rhd S_i^*(r^o)$ for at least one *i*. That is,

$$\max\{\pi_i : \pi \in T^{c*}(g_{-1})\} \ge \pi_i(g_1^{bo}, g_{-1}) \tag{A2.6}$$

Let g_1^{bc} denote 1's largest choice allocation in [0,W] in this scenario. Because *i*'s payoff is increasing in g_1, we have

$$\pi_i(g_1^{bc}, g_{-1}) = \max\{\pi_i : \pi \in T^{c*}(g_{-1})\} \tag{A2.7}$$

It follows from (A2.6) and (A2.7) that $\pi_i(g_1^{bc}, g_{-1}) \ge \pi_i(g_1^{bo}, g_{-1})$, which together with *i*'s payoff increasing in g_1 imply $g_1^{bc} \ge g_1^{bo}$ (*)

Step 2. It follows from $[c,W] \subset [0,W]$ that $T(g_{-1}) \subset T^c(g_{-1})$, and by construction of the scenario in Step 1, the moral reference points are identical. By R-consistency properties, $T^* = T^{c*} \cap T$. It follows from $g_1^{bc} \ge g_1^{bo} \ge c$ that $g_1^{bc} \in T^*$, which together with (*) concludes the proof.

Part 2. Effect of initial g^e. Take any two g^e-games with initial allocations s > t in the public account. We show that the smallest allocation in *s*-game is smaller than the smallest allocation in *t*-game. Proof for the largest allocation is similar.

In both games, player 1 can choose from [0,W], hence in the payoff space, $S^s(g_{-1} \mid g^e = s) = S^t(g_{-1} \mid g^e = t\} = S(g_{-1})$. Let $g_i^x \in [0,W]^*, x \in \{s,t\}$ be the smallest choice allocation in *x*-game. That leaves player 1 with the largest payoff as own payoff decreases in own allocation,

$$\pi_1(g_1^x, g_{-1}) = \max\{\pi_1(g_1, g_{-1}) : g_1 \in S^*(g_{-1} \mid g^e = x)\}.$$

Use (5) to verify that for all i, $r_i^s - r_i^t = 0.5(n\gamma - 1)(s-t)$ and $r_1^s > r_1^t$, hence by Property M_R^a, $(S^s)_1^* \triangleright (S^t)_1^*$, which implies

$$\pi_1(g_1^s, g_{-1}) \geq \pi_1(g_1^t, g_{-1})$$

and together with player 1's payoff decreasing in g_1, imply $g_1^s \leq g_1^t$.

2.C Example of Objective Function for Morally Monotonic Choice

Without any loss of generality, we use 1 to index the decision maker. As in Cox et al. (2020), let player 1's choices from a feasible set X when the moral reference point is r be determined by maximization of the following weighted sum of payoffs (in utils):

$$X^*(r) = \{\pi^* \in X : U(\pi^* \mid r) \geq U(\pi \mid r), \forall \pi \in X\} \qquad \text{(A2.8)}$$

where $U(\pi \mid r) = \sum_k w_k(r) u(\pi_k)$ with weights $w_k(r) = \theta(\sigma_k r_k) / \sum_j \theta(\sigma_j r_j)$, $\sigma_1 > 1 = \sigma_{k>1}$, for some increasing $u(\cdot)$ and increasing positive $\theta(\cdot)$ that satisfies $\theta(y+z) = \theta(y)\theta(z)$. We show that such choices satisfy R-Consistency Properties.

First, the <u>R-Consistency Properties</u> (α_R and β_R) are clearly satisfied as for any given reference point, $r = s^r = t^r$, and (payoff) choice $x^* \in S^*(r) \cap T$, $U(x^* \mid r) \geq U(y \mid r), \forall y \in S$ implies

a. $U(x^* \mid r) \geq U(y \mid r), \forall y \in T \subseteq S$ and therefore $x^* \in T^*(r)$

b. $U(x^* \mid r) = U(y^* \mid r), \forall y^* \in T^*(r)$ and therefore $y^* \in S^*(r)$

Next, for Moral Monotonicity Properties, let $T(t^r) = S(s^r)$ and $t^* \in T^*(t^r)$ (*) and $s^* \in S^*(s^r)$ (**). It follows from (*) and (**) that

$$\begin{aligned} &(1) \quad \sum_{i=1}^n w_i(t^r) u(t_i^*) \geq \sum_{i=1}^n w_i(t^r) u(s_i^*) \\ &(2) \quad \sum_{i=1}^n w_i(s^r) u(s_i^*) \geq \sum_{i=1}^n w_i(s^r) u(t_i^*) \end{aligned} \qquad \text{(A2.9)}$$

<u>Moral Monotonicity Properties</u> Let $t_k^r - s_k^r = \delta (> 0)$ for all $k \in K \neq \emptyset$ and $t_j^r = s_j^r$ for all $j \notin K$.

Property M_R^a. If $K = \{k\}$ then $T_k^*(t^r) \triangleright S_k^*(s^r)$

In (A2.9), multiply both sides of (1) by $\theta(\sigma t_1^r)+\sum_{i>1}\theta(t_i^r)$ and (2) by $\sum_i \theta(\sigma_i s_i^r)$, and rearrange terms

$$\sum_{i=1}^{n}\theta(\sigma_i t_i^r)(u(t_i^*)-u(s_i^*))\geq 0$$
$$\sum_{i=1}^{n}\theta(\sigma_i s_i^r)(u(s_i^*)-u(t_i^*))\geq 0$$

Add the two inequalities and use $t_j^r = s_j^r$ for all $j\notin K$ to get

$$\left(\theta(\sigma_k t_k^r)-\theta(\sigma_k s_k^r)\right)\left(u(t_k^*)-u(s_k^*)\right)\geq 0$$

It follows from $t_k^r - s_k^r = \delta(>0)$ and monotonicity of $\theta(\cdot)$ that the term in the first bracket is positive. Hence, $u(t_k^*)-u(s_k^*)\geq 0$, and by monotonicity of $u(\cdot)$, $t_k^*\geq s_k^*$.

Property M_R^b. If $K=\{1,...,n\}$ then $T_1^*(t^r)\rhd S_1^*(s^r)$

First note that,

$$\frac{w_i(t^r)}{w_n(t^r)}=\frac{\theta(t_i^r)}{\theta(t_n^r)}=\frac{\theta(s_i^r+\delta)}{\theta(s_n^r+\delta))}=\frac{\theta(s_i^r)\theta(\delta)}{\theta(s_n^r)\theta(\delta)}=\frac{\theta(s_i^r)}{\theta(s_n^r)}=\frac{w_i(s^r)}{w_n(s^r)},\qquad i>1$$

$$\frac{w_1(t^r)}{w_n(t^r)}=\frac{\theta(\sigma_1 t_1^r)}{\theta(t_n^r)}=\frac{\theta(\sigma_1(s_1^r+\delta))}{\theta(s_n^r+\delta))}=\frac{\theta(\sigma_1 s_1^r)\theta(\sigma_1\delta)}{\theta(s_n^r)\theta(\delta)}>\frac{\theta(\sigma_1 s_1^r)}{\theta(s_n^r)}=\frac{w_1(s^r)}{w_n(s^r)} \tag{A2.10}$$

where the inequality follows from monotonicity of $\theta(\cdot)$ and $\sigma_1>1$. Next, divide the first and second inequalities in (A2.9) by player *n*'s weight $w_n(t^r)$ and $w_n(s^r)$ and rearrange terms to get

$$\frac{w_1(t^r)}{w_n(t^r)}\left(u(t_1^*)-u(s_1^*)\right)\geq\sum_{i=2}^{n-1}\frac{w_i(t^r)}{w_n(t^r)}(u(s_i^*)-u(t_i^*))+u(s_n^*)-u(t_n^*)$$
$$\frac{w_1(s^r)}{w_n(s^r)}\left(u(s_1^*)-u(t_1^*)\right)\geq\sum_{i=2}^{n-1}\frac{w_i(s^r)}{w_n(s^r)}(u(t_i^*)-u(s_i^*))+u(t_n^*)-u(s_n^*)$$

add the two inequalities

$$\left(\frac{w_1(t^r)}{w_n(t^r)}-\frac{w_1(s^r)}{w_n(s^r)}\right)\left(u(t_1^*)-u(s_1^*)\right)\geq\sum_{i=2}^{n-1}\left(\frac{w_i(t^r)}{w_n(t^r)}-\frac{w_i(s^r)}{w_n(s^r)}\right)(u(s_i^*)-u(t_i^*))=0$$

where the equality follows from (A2.10), for $i > 1$. By (A2.10) for $i = 1$ the first term on the left-hand-side is strictly positive, and by monotonicity of $u(\cdot)$, $t_1^* \geq s_1^*$.

Appendix O.3. Effects of per capita Initial, g^e and Quota, c on Extreme Nash Equilibria

In this section we provide a general result for comparative statics of extreme Nash equilibria when best responses are increasing in others' allocations.

Notation. We say best response allocations are increasing in others' allocations if $br_i(\hat{g}_{-i}) \rhd br_i(g_{-i})$ for any vectors of others' allocations such that $(\hat{g}_{-i}) \geq (g_{-i})$.[36]

Extreme Equilibria: If best response allocations are increasing in others' allocations then extreme (the largest and the smallest) Nash Equilibria allocations

c. do not vary with c and g^e for conventional rational choice

d. increase in c and decrease in g^e for morally monotonic choice

The intuition for (b) is that an increase in g^e, has a negative direct effect on morally monotonic allocations (Proposition 2), and a negative indirect effect, as lower others' allocations elicit lower own allocation (for increasing best responses).

PROOF. *Part* a. By Proposition 1, best responses are invariant to g^e and (non-binding) c, therefore, Nash equilibrium set is also invariant.

Part b. We use Tarski (1955) to compare extreme Nash equilibria across g^e. Proof for (non-binding) quota effect is similar. Let Υ denote the product space, that is $\Upsilon = \times_i \{0,...,W\}$. and $(\Upsilon, \leq)$ denote the lattice with conventional, increasing partial order, $\leq$. For any $t = g^e$, let $f^t(g) = (f_i^t(g_{-i}) \mid i = 1...n)$ where $f_i^t(\cdot) \in \{0,...,W\} \subset R$ is i's largest best reply allocation, that is

$$f_i^t(g_{-i}) = \max\{g_i \in \{0,...,W\} \mid g_i \in br_i^t(g_{-i})\}$$

[36] $\leq$ is the conventional, increasing partial order in R^n; that is, $x \leq y$ if $x_i \leq y_i$, for all i=1,…,n.

Since best response largest allocation is increasing in others' allocations, and Υ is a complete lattice, the largest Nash equilibrium is[37]

$$\alpha^t = \sup_{\Upsilon} E^t = \{g \in \Upsilon \mid g \le f^t(g)\}$$

For any two, *t* and *s*, such as $t > s$, by Proposition 2, best response largest allocations are smaller in *t* than in *s*, which implies $E^t \subseteq E^s$ and therefore $\sup_{\Upsilon} E^t \le \sup_{\Upsilon} E^s$.

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

Appendix O.4. Special Case Objective Function Derivation and Application

We here report results for a special case of moral monotonicity theory using a parametric objective function.

Let $S(g_{-1}) = \{\pi(g) \mid g_1 \in [c, W]\} \subset R^n$ be player 1's feasible set in final payoff space. Let the choice allocation be determined by maximization of

$$\max_{g_1} U(g_1 \mid r, g_{-1}) = \sum_{i=1..n} w_i(r) u(\pi_i) = \frac{1}{\mathrm{e}^{\sigma r_1} + \sum_{i>1} \mathrm{e}^{r_i}} \left(\mathrm{e}^{\sigma r_1} u(\pi_i) + \sum_{i>1} e^{r_i} u(\pi_i) \right) \qquad \text{(A4.1)}$$

where $u(x) = \frac{1}{\alpha}(1 - e^{-\alpha x})$.

At feasible set, $(S(g_{-1}) \mid s^r)$ individual 1's choice, $g_1^{br} = br(g_{-1} \mid r = s^r)$ is the closest feasible allocations to the one implicitly determined by

$$\sum_{i>1} \mathrm{e}^{r_i - \sigma r_1} \, \mathrm{e}^{-\alpha(\pi_i - \pi_1)} = (1-\gamma)/\gamma \qquad \text{(A4.2)}$$

Note that $\pi_i - \pi_1 = g_1 - g_i$, substitute it in (A4.2) and solve for g_1 to get

---

[37] 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 $(\Upsilon, \le)$ being a complete lattice. For all $g \in E^t$, $g \le \alpha^t$ and increasing $f^t(\cdot)$ imply $g \le f^t(g) \le f^t(\alpha^t)$; that is $f^t(\alpha^t)$ is an upper bound of E^t, hence $\alpha^t \le f^t(\alpha^t)$. By increasing $f^t(\cdot)$, $f^t(\alpha^t) \le f^t(f^t(\alpha^t))$ implying $f^t(\alpha^t) \in E^t$, hence $f^t(\alpha^t) \le \alpha^t$.

$$g_1^{br} = br(g_{-1} \mid r) = \frac{1}{\alpha}\left(\ln(\frac{\gamma}{1-\gamma}) - \sigma r_1 + \ln(\sum_{i>1} e^{r_i + \alpha g_i}) \right) \tag{A4.3}$$

Use statement (7) in the main text for moral reference point and statement (A4.3) here to get the following properties of this best response function, $g_1^{br}(.)$:

a. $g_1^{br}(\cdot)$ decreases in r_1 but increases in $r_i, i \neq 1$. This is consistent with Properties M_R^a and M_R^b as player 1's payoff decreases in g_1 whereas player i's (i >1) payoff increases in g_1.

b. Consistent with Proposition 2.a, $g_1^{br}(\cdot)$ increases in c. For any, $c^+ = c + \Delta c$ for some $\Delta c > 0$, let best response allocations (as in (A4.3)) be g_1^{+br} and g_1^{br}, respectively. By statement (7) in the main text, $r_1(c^+) = r_1(c)$ and $r_i(c^+) = r_i(c) + 0.5\gamma\Delta c$ for all $i \in \{2,\ldots,n\}$, which together with statement (A4.3) above imply

$$g_1^{+br} - g_1^{br} = \frac{1}{\alpha}\left(\ln(\sum_{i>1} e^{r_i(c)+0.5\gamma\Delta c + \alpha g_i}) - \ln(\sum_{i>1} e^{r_i + \alpha g_i}) \right) = \frac{1}{\alpha} 0.5\gamma\Delta c \geq 0$$

c. Consistent with Proposition 2.b, $g_1^{br}(\cdot)$ decreases in g^e. For any, $g^{e+} = g^e + \Delta g^e$ for some $\Delta g^e > 0$, let best response allocations (as in (A4.3)) be g_1^{+br} and g_1^{br}, respectively. By statement (7), $r_i(g^{e+}) = r_i(g^e) + 0.5(\gamma n - 1)\Delta g^e$, for all $i \in \{1,\ldots,n\}$, which together with statement (A4.3), $\sigma > 1$ and $\gamma \in (1/n, 1)$ imply

$$g_1^{+br} - g_1^{br} = \frac{1}{\alpha}(1-\sigma)0.5(n\gamma - 1)\Delta g^e \leq 0$$

Appendix O.5. Data Are Inconsistent with Conventional Loss Aversion

The observed patterns of smaller allocations in appropriation game than in provision game, and larger allocations in games with non-binding contraction are also inconsistent with predictions of the classical loss-aversion reference dependent model of Tversky and Kahneman (1991), that is at the heart of well-known later models of reference dependent choices in the extensive literature. Indeed, for a given vector of others' allocations, g_{-1}, player 1's feasible (payoff) set is $S(g_{-1} \mid r)$. There are two alternatives for the reference point.

Alternative 1. The reference point is the initial vector of payoffs at feasible (payoff) set $S(g_{-1} \mid r)$ before player 1 makes his choice.

In appropriation game, $r^a = (\gamma(W + g_2), W - g_2 + \gamma(W + g_2))$. Note that any point from $S(g_{-1} \mid r^a)$ offers a gain for player 1 as his payoff increases but a loss for player 2 as other's payoff decreases. Using 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 $u(.)$ and $v(.)$ and a loss parameter $\lambda_2 > 1$. The optimal allocation, g_1^a satisfies the f.o.c.,

$$\lambda_2 v'(\pi_2^a)\gamma = (1-\gamma)u'(\pi_1^a) \qquad \text{(A5.1)}$$

In provision game, the payoff vector before player 1 makes his choice, is $r^p = (W + \gamma g_2, W - g_2 + \gamma g_2)$. Note that any point from $S(g_{-1} \mid r^p)$ is a loss for player 1 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

$$\frac{dU^p}{dg_1} = \lambda_1 u'(\pi_1)(\gamma - 1) + \gamma v'(\pi_2)$$

Evaluate this at the optimal allocation, g_1^a in the appropriation game and use (A5.1) to get

$$\frac{dU^p}{dg_1}\Big|_{g_1 = g_1^a} = \lambda_1 u'(\pi_1^a)(\gamma - 1) + \frac{(1-\gamma)}{\lambda_2} u'(\pi_1^a)$$
$$= \left(-\lambda_1 + \frac{1}{\lambda_2}\right)(1-\gamma)u'(\pi_1^a) < 0$$

where the inequality follows from loss aversion, $\lambda_1 > 1 > 1/\lambda_2$. Hence, g_1^a is too large to be optimal in provision game, so $g_1^p < g_1^a$. This implication is inconsistent with the robust result of provision games eliciting larger allocations to the public account than payoff-equivalent appropriation games.

The intuition for this result is as follows. The reference point in appropriation game is northwest of the one in provision game (clearly visible in Figure 1 as r^p is the most southeast point of the solid line whereas r^a is the most northwest point). Let g_1^* be an optimal allocation for conventional rational choice and let π^* be the payoff vector there (referred to as optimal "consumption" point in the literature). The effect of the reference point on optimal payoff vector in appropriation game is to move it northwest of π^* because r^a is northwest, so moving southeast decreases "consumption" utility and is further away from the reference point. Player 1 can accomplish this (decrease his payoff and increase other's payoff) only by increasing his allocation to the public account, hence, $g_1^* \leq g_1^a$. The optimal payoff vector in provision game, must be southeast of π^* because r^p is southeast. A larger payoff for player 1 and smaller for player 2 requires lower allocation to the public account than g_1^*, hence $g_1^p \leq g_1^*$. The implication of TK reference dependent model then is $g_1^p \leq g_1^a$, which is inconsistent with the robust result of provision games eliciting larger allocations to the public account than payoff-equivalent appropriation games.

<u>*Alternative* 2.</u> In this alternative interpretation, the reference point is the initial endowed payoffs: $r^p = (W,W)$ in provision game and $r^a = 2\gamma(W,W)$ in appropriation game. In this case, in either game, the reference point remains the same for all player 1's feasible (payoff) sets $S(g_{-i} \mid r)$ in the game, so implications is that allocation choices are not affected by non-binding contractions. This is inconsistent with observed (best response) allocations increasing in the presence of nonbinding contractions.