Warm-Glow Giving and Freedom to be Selfish*

Özgür Evren† and Stefania Minardi‡

February 14, 2013

Abstract

Warm-glow refers to other-serving behavior that is valuable for the actor *per se*, apart from its social implications. We provide axiomatic foundations for warm-glow by viewing it as a form of preference for larger choice sets driven by one's desire to have freedom to act selfishly. Specifically, an individual who experiences warm-glow values the availability of selfish options even if she plans to act unselfishly. Briefly put, warm-glow necessitates free will. Our theory accommodates the empirical findings on motivation crowding out and provides clear-cut predictions for empirically distinguishing between warm-glow and other motivations for prosocial behavior, a task of obvious importance for policy. The choice behavior implied by our theory subsumes Riker and Ordeshook (1968) on voting and Andreoni (1990) on the provision of public goods.

JEL Classification: D11, D64, D81

Keywords: Altruism, Warm-Glow, Freedom of Choice, Philanthropy, Charitable Giving, Public Goods, Voter Turnout

*We are grateful to Efe Ok and Debraj Ray for their continuous guidance and support. We would also like to thank David Dillenberger, Itzhak Gilboa, Anna Gumen, Farhad Husseinov, Edi Karni, Massimo Marinacci, Sujoy Mukerji, Jawwad Noor, David Pearce, Leonardo Pejsachowicz, Philipp Sadowski, and Andrei Savochkin for very helpful comments and suggestions. When this project started, the authors were Ph.D. students at New York University, Department of Economics. We owe special thanks to this institution. All errors are our own.

†New Economic School, Nakhimovsky Prospekt 47, Suite 1721, Moscow 117418, Russia. e-mail: oevren@nes.ru

‡Economics and Decision Sciences Department, HEC Paris, 1, rue de la Libération, 78351, Jouy-en-Josas, France. e-mail: minardi@hec.fr

1 Introduction

In the last decades, there has been a surge of interest in models of prosocial behavior that depart from the traditional approach which explains such behavior with the classical notion of altruism. It has been argued, for instance, that charitable donations may be motivated by a desire for status, acclaim or self-satisfaction (e.g., Arrow, 1972; Becker, 1974; Andreoni, 1989, 1990; Glazer and Konrad, 1996; Bénabou and Tirole, 2006). In the voting literature, Riker and Ordeshook (1968) provide a remarkable example, maintaining that the act of voting in a democracy can be perceived as a civic duty and that performing this duty may lead to a feeling of satisfaction for voters. More recently, Coate and Conlin (2004) and Feddersen and Sandroni (2006) envision citizens who deem voting as an ethical duty whenever this is justified from a rule-utilitarian perspective.

Often, the applied models do not formally elaborate on the connections between such alternative forms of individual satisfaction and the prosocial actions in question, perhaps because the underlying factors are highly complicated.[1] Rather, a typical model proposes a decision maker who derives an intrinsic utility from a prosocial action, apart from the instrumental value of that action as a means of influencing others' welfare. Since the seminal papers of Andreoni (1989, 1990), the notion of *warm-glow payoff* refers to such intrinsic utility associated with prosocial actions. In these papers, Andreoni shows that the empirical findings on public good provision can be satisfactorily explained in a model of warm-glow, as opposed to classical models of altruism which lead to questionable predictions.[2] Riker and Ordeshook (1968), on the other hand, is, perhaps, a more striking example of a warm-glow model, for in their setup the intrinsic utility associated with the act of voting (as a civic duty) corresponds to a warm-glow payoff that is completely independent of instrumental concerns. This structure of their model enables Riker and Ordeshook to explain voter turnout in large elections where voting is typically a time-consuming, costly activity that is highly unlikely to influence others' welfare.

Despite their merits, applied warm-glow models are often criticized as being *ad hoc*,[3] presumably because the models are silent about what drives the warm-glow

[1]Among few notable exceptions, Bénabou and Tirole (2006) study the interplay among altruistic motivation, material incentives, and concerns for social reputation in a game-theoretic model. They show that the introduction of monetary incentives may crowd out prosocial behavior by weakening the reputational motivation for such behavior.

[2]We elaborate more on Andreoni's findings in Section 4.

[3]As Andreoni (2006) writes on p. 1222, "Putting warm-glow into the model is, while in-

payoff associated with a prosocial action, while we do not yet have a foundational work that supports the idea that taking a prosocial action can be intrinsically valuable.[4]

In this paper, we provide utility representation theorems for some classes of preference relations over sets of social allocations. While these representations allow us to formalize the notion of warm-glow payoff, the conceptual interpretation of the assumptions that we impose on preference relations does not necessitate one to presume that prosocial actions may be valuable *per se*. In this sense, we show that the relevance of the notion of warm-glow for human behavior is a necessary consequence of more primitive properties of preference relations. In particular, one of our key assumptions is based on the idea that people may be concerned with their freedom to be selfish, a phenomenon that is supported by considerable evidence, as we discuss below. Thus, a particular implication of our findings is that in an important class of applied warm-glow models, the associated behavior can be seen essentially as a form of preference for freedom to be selfish.

Our starting point is that a prosocial action can be valuable *per se* only if it is an act of free will. While this would seem self-evident on many occasions,[5] a plethora of evidence shows that restricting one's freedom to behave selfishly may actually motivate that person to take more selfish actions. One set of evidence in this direction comes from economic experiments. For example, Falk and Kosfeld (2006) examine a principal-agent problem in which the agent chooses a productive activity a that incurs costs for herself while increasing the principal's payoff. The key feature of the experiment is that the principal determines the set of options available to the agent. Falk and Kosfeld find that if the principal imposes a lower bound for a, then, compared to the case in which the principal does not impose such a restriction, a majority of the agents select a lower level of a . As Falk and Kosfeld note (pp. 1611-1612), a potential reason for this finding is that "agents do not like to be restricted, and perceive control as a negative signal of distrust." Similarly, Fehr and Rockenbach (2003) and Houser, Xiao, McCabe, and Smith

tuitively appealing, an admittedly *ad hoc* fix." However, as Andreoni (2006) also noted, the experimental findings on public good provision provide strong support for the notion of warm-glow (see, e.g., Andreoni, 1993, 1995; Palfrey and Prisbrey, 1997; Bolton and Katok, 1998; Andreoni and Miller, 2002; Eckel, Grossman, and Johnston, 2005). Momentarily, we will discuss another set of findings which suggest that intrinsic motivations do exist and they may be crowded out by extrinsic incentives.

[4]We will discuss the related decision-theoretic work momentarily.

[5]Indeed, in the aforementioned literature on prosocial behavior, the word "giving" refers to a voluntary act.

(2008) find that trustees may return less when investors use sanctions to enforce demanding outcomes.

Another set of evidence comes from psychological experiments testing Brehm's (1966) theory of reactance, which maintains that a threat to someone's freedom to choose an option may lead that person to a psychological state which involves an enhanced level of attraction to that option. In one experiment of this sort, Horowitz (1968) has subjects listen to the tape-recorded voice of a graduate student who is in need of help because he had made some miscalculations regarding a (separate) sensory-deprivation experiment. Horowitz finds that subjects' tendency to help the graduate student increases significantly if they are told that they are under no obligation to help the graduate student.[6]

In line with these findings, our first representation theorem describes an agent who is concerned with her freedom to be selfish. For a generic social allocation x in $\mathbb{R}^{\ell}_{+}$, let the first component x_1 stand for the private consumption of the agent. Our first theorem delivers a weakly increasing function $U : \mathbb{R}^{\ell}_{+} \times \mathbb{R}_{+} \to \mathbb{R}$ such that the (indirect) utility of a set A of allocations equals

$$V(A) = \max_{x \in A} U(x, \max_{y \in A} y_1). \tag{1}$$

This representation suggests that when faced with the set A, the agent would select the allocation that maximizes $U(x, \max_{y \in A} y_1)$ over A. Accordingly, $U(x, \max_{y \in A} y_1)$ is interpreted as the utility of selecting x from A. Since U is weakly increasing in its last argument, a key implication of the representation is that the utility of selecting a given allocation x increases with the maximum possible private consumption, $\max_{y \in A} y_1$. In turn, this makes (1) a representation of preference for freedom to be selfish as $\max_{y \in A} y_1$ can naturally be viewed as a measure of the agent's freedom to be selfish.

According to this representation, a fine (or any other restriction) on selfish behavior that decreases the maximum possible private consumption actually decreases the utility of selecting any available allocation. Moreover, the resulting decrease in the utility of selecting a prosocial allocation may well be larger than the decrease in the utility of selecting a more selfish allocation. This, in turn, implies that the agent can select a prosocial allocation from a given set but switch

[6] Berkowitz (1973) provides an early review of experiments on people's helping behavior that support the reactance theory. It is also worth mentioning the works of Clee and Wicklund (1980), and Kirchler (1999), which discuss the implications of reactance theory on consumers' behavior and tax payers' attitude, respectively.

to a more selfish allocation upon the introduction of a fine. Hence, the choice behavior implied by (1) is compatible with the aforementioned evidence on the adverse effects of restricting one's freedom to behave selfishly.

While representation (1) is a natural starting point, it is by no means the only representation one can imagine that is compatible with the notion of preference for freedom to be selfish. Our second representation theorem delivers a weakly increasing function $U : \mathbb{R}^\ell_+ \times \mathbb{R}_+ \to \mathbb{R}$ such that the utility of a set A is given by

$$V(A) = \max_{x \in A} U(x, \max_{y \in A} y_1 - x_1). \tag{2}$$

This representation suggests that the agent's utility of selecting a given allocation x is an increasing function of $\max_{y \in A} y_1 - x_1$. In turn, the term $\max_{y \in A} y_1 - x_1$ is the private cost that the agent incurs if she decides to select allocation x from menu A. Put differently, $\max_{y \in A} y_1 - x_1$ tells us how much more private consumption the agent could enjoy if she were not to select x. When viewed in this way, representation (2) establishes an alternative relation between the agent's welfare and her freedom to be selfish.

Just as before, representation (2) is compatible with the evidence on the adverse effects of restricting one's freedom to behave selfishly. The difference between representations (1) and (2) lies in their implications about the effects of subsidizing prosocial actions. To see this point, note that $U(x, \max_{y \in A} y_1)$ increases with x_1 (keeping $\max_{y \in A} y_1$ constant), while $U(x, \max_{y \in A} y_1 - x_1)$ may or may not increase with x_1 because the term $\max_{y \in A} y_1 - x_1$ decreases with x_1. According to representation (1), therefore, a private reward for prosocial behavior can never backfire as the utility of such behavior would increase with the reward. By contrast, representation (2) is also compatible with opposite instances. In line with these observations, we find that the only difference between the characterization of the two representations lies in the welfare effects of private rewards on prosocial actions.

In terms of the implied choice among allocations, the generality added by representation (2) is motivated by mounting evidence pointing to the fact that rewarding prosocial actions sometimes lead to negative outcomes. A remarkable finding of this sort is due to Gneezy and Rustichini (2000), who show that offering monetary rewards to children for volunteer work may decrease their performance. Another example is the work of Mellström and Johannesson (2010), which shows that the supply of female blood donors decreases almost by half when they are offered monetary compensation. Such findings are usually considered as evidence for the view that extrinsic motivations may crowd out intrinsic motivations, which has attracted much attention in economics and psychology.[7] This, in turn, takes us back to the notion of warm-glow.

To see why representation (2) embodies a notion of warm-glow, consider a prosocial allocation x and a more selfish allocation y with $y_1 > x_1$. If both x and y are available, then the act of selecting x can be considered a prosocial action taken by free will, which would lead to utility $U(x, y_1 - x_1)$. If, on the other hand, x is the only available option, the act of selecting x would merely be a necessity leading to utility $U(x, 0)$. Therefore, the amount $U(x, y_1 - x_1) - U(x, 0)$ can be considered as an intrinsic payoff associated with the prosocial act of selecting x over y, while $U(x, 0)$ gives us the instrumental payoff resulting from consumption of x.[8] With this notion of warm-glow in our packet, in Section 4.1, we will show that the implied choice behavior subsumes Andreoni's (1989, 1990) warm-glow model. In Section 4.2, we will extend the scope of representation (2) by reinterpreting the notion of an allocation as a vector that lists the (expected) material payoffs of the individuals in the society. In Section 4.3, we apply this extended representation to the problem of voter turnout in large elections, and show that the associated behavior subsumes the civic-duty model of Riker and Ordeshook (1968). In this way, we lay foundations for two prominent models from different subfields of social choice theory.

In fact, our theory not only lays foundations for the notion of warm-glow, but also arms us with clear-cut predictions that can be utilized to empirically distinguish between alternative motivations for prosocial actions. For example, Andreoni's (1989, 1990) static model cannot distinguish pure altruism from warm-glow motivation with quasilinear utility functions. However, in a two-period consumption-saving problem, we can easily distinguish between these two motivations even with quasilinear temporal utility functions, for in this setup the agents' preferences over budget sets become important. We demonstrate this point in Section 4.1.1 with an example of a two-period bequest-giving problem. In this example, the saving and bequest-giving behavior of an agent is sensitive to re-

[7] The idea that monetary rewards or punishments may crowd out intrinsic motivations dates back to Titmuss (1970), who also considers the specific example of blood donation. Starting with Deci (1975), a group of cognitive psychologists advocate the same view. Frey and Jegen (2001) provide an extensive survey of the evidence collected by economists and psychologists in support of the motivation crowding out theory.

[8] Of course, a similar notion of warm-glow can be derived from representation (1), albeit in a less satisfactory way.

distributive policies if she is motivated by warm-glow, but not if she is altruistic, while the two types behave in exactly the same way if they face the same budget set in period 2. At a more fundamental level, the two motivations entail quite distinct attitudes towards larger choice sets: an agent motivated by warm-glow enjoys the freedom of choice offered by larger sets, while a purely altruistic agent is neutral towards the size of the choice set that she faces.[9]

At this point, we should address an important limitation of representation (2). Just as in Riker and Ordeshook (1968), the notion of warm-glow payoff induced by representation (2) is completely independent of the social implications of the action in question. On one hand, this makes representation (2) a powerful tool in suitable settings, such as voter-turnout models or public goods games in which there is a one-to-one correspondence between agents' self-sacrifice and their contribution to the public goods. On the other hand, this limitation leads to two difficulties. First, if we apply the representation to an arbitrary set of allocations, the implied choice behavior may become awkward on occasion. For example, the agent can enjoy "burning money" or making a marginal contribution to someone at the cost of hurting everybody except that person. We avoid such awkward implications by imposing suitable restrictions on the domain of the agent's preference relation. In Section 3.2.1, we argue that this is a natural modeling choice in view of the experiments on motivation crowding out. The second difficulty is that even in this restricted domain, there seems to be no reason to believe that the intrinsic utility of a prosocial action should truly be independent of its social implications. As a solution to this problem, we also provide a multi-dimensional extension of representation (2), which takes into account the social implications of an action in the calculus of warm-glow. While this is certainly a useful generalization, we postpone it to Appendix A, as it makes the connections between the notions of warm-glow and freedom of choice less transparent.

As a final methodological remark, it should be noted that many applications of our theory focus on how the agent's welfare and allocation choice may change in response to the behavior of another actor (such as the government or a principal) who chooses the menu available to the agent.[10] If such a secondary actor is present, the preference relation over menus that we model in this paper reflects

[9]In fact, we can think of a third motivation characterized by preference for *smaller* sets — namely, the urge to avoid negative emotions associated with selfish behavior such as shame or guilt. Prosocial behavior driven by such negative emotions has been addressed recently in Noor and Ren (2011), and Dillenberger and Sadowski (2012), which we discuss below.

[10]For example, all aforementioned experiments on motivation crowding out are of this sort.

psychological preferences of the agent. For our purposes, it is necessary to distinguish "preference" from "choice," as the latter is more intimately linked with the notions of freedom of choice and warm-glow. Indeed, if the agent has to select a menu among others, her behavior would typically depend on the size of the collection of menus that she faces because of the freedom motive. Similarly, the agent could experience a second form of warm-glow, driven solely by the choice of a particular menu (as opposed to the planned choice of allocation). One could address such concerns only in a model of reference-dependent menu choice (just as the allocation choice implied by our theory is reference dependent[11]). By contrast, in the present paper, we assume that non-instrumental motives influence the agent's preference relation over menus only through the planned choice of allocations. This often makes it difficult to interpret our theory as a model of menu choice, but there are some notable exceptions. For instance, in some cases of interest, such as the bequest-giving example mentioned above, the first-stage behavior of the agent may not be publicly observable, unlike the giving behavior in the second stage. In such choice situations, our theory can be applied without major hesitation, assuming that prosocial behavior leads to warm-glow experience only when it is publicly observable.[12] In Section 6, we will elaborate more on these issues. In particular, we will discuss how the agent's psychological preferences can be recovered by suitably designed survey questions, which put the agent into the position of a passive recipient (of menus).

Next, we discuss the related decision-theoretic literature. Section 2 introduces the formal setup, while Section 3 presents the axiomatic foundations and the main representation results. Section 4 discusses the applications of our basic warm-glow representation. Section 5 is devoted to a choice-theoretic study of the implied second-stage behavior for a fixed menu, while Section 6 discusses the distinction between the first and second stages in our theory. Section 7 relates our representations to that of Kreps (1979), and Section 8 concludes. All proofs and some other supplementary material, including the multi-dimensional version

[11]Some special forms of utility functions in our representations, such as additively separable or quasilinear forms, induce choice behavior that is consistent with the Weak Axiom of Revealed Preferences in the second stage, but such forms are too restrictive, especially for the purposes of applied work that is concerned with distinguishing warm-glow motive from pure altruism on the basis of second-stage behavior. In particular, regarding representation (2), Andreoni's (1989, 1990) additional assumptions do not rule out reference dependence. (See also footnote 17 and related discussion in Section 3.)

[12]Admittedly, however, this assumption corresponds to a narrow interpretation of the notion of warm-glow. In particular, we cannot think of a necessary connection between the notion of freedom of choice and observability of the prosocial action in question.

of our warm-glow representation, are relegated to appendices.

1.1 Related Literature

Starting with Sen (1985, 1988), many economists have recognized that there may be a deeper connection between a person's well-being and freedom of choice than that entailed by the traditional, instrumentalist view. In a nutshell, this literature maintains that a large set of alternatives, which offers a certain degree of freedom of choice, may be more valuable to the decision maker than *the* alternative that she selects from that set (Sen, 1985, 1988; Puppe, 1996; Sugden, 1998, among others).[13] The focus of this strand of literature is the measurement of freedom of choice associated with menus.

Dillenberger and Sadowski (2012) study a negative form of prosocial behavior driven by shame associated with selfish acts. Their main representation result describes an agent who exhibits a preference for *smaller* menus. The present paper has further differences in terms of the implied choice among social allocations for a fixed menu. Most remarkably, according to Dillenberger and Sadowski, increasing the private cost of an action can only decrease the likelihood of observing that action, which is at odds with the literature on motivation crowding out. Similarly, their approach cannot explain the act of voting in large elections if this is a costly activity with almost negligible social consequences, as posited by some scholars (e.g., Feddersen and Sandroni, 2006).

Noor and Ren (2011) show, in an experiment, that giving rates in a dictator game increase significantly if payments are offered with delay. To explain this observation, they propose a model of a decision maker who is intrinsically concerned with others' welfare but who faces the temptation to behave selfishly. As in Dillenberger and Sadowski (2012), this is a model of preference for smaller sets that does not accommodate the findings on motivation crowding out.

Cherepanov, Feddersen, and Sandroni (2012), and Saito (2013) are concurrent papers that are also concerned with the foundations of the notion of warm-glow. Cherepanov et al. (2012) propose an abstract model of choice among alternatives, holding fixed the menu that the decision maker faces. Their main point is that, unlike pure altruism, the relevant forms of non-altruistic behavior may lead to violations of the Weak Axiom of Revealed Preferences (henceforth WARP). In

[13] This contention contrasts with the "preference for flexibility" approach that focuses on the *instrumental* value of larger menus driven by choice uncertainty (Kreps, 1979; Dekel, Lipman, and Rustichini, 2001). In Section 7, we relate the present paper to Kreps (1979).

concert with this observation, in our model, the utility function that governs the choice of allocations is menu-dependent. Yet, the choice behavior that corresponds to our model is not within the scope of Cherepanov et al. (2012), because that paper models the "warm-glow payoff" as a fixed number that does not depend on the menu that the agent faces or the allocation that leads to the warm-glow experience. By contrast, in our model, the warm-glow payoff increases with the private cost associated with the allocation in question, which is the key feature that makes our model compatible with the literature on motivation crowding out.

Saito (2013) focuses on preference relations over menus of allocations of lotteries.[14] His representation result delivers an additively separable utility function, which contributes to uniqueness properties of the representation. However, in the second stage, this additively separable structure implies a mode of behavior that cannot be distinguished from that induced by classical altruism for a given menu. In particular, Saito's model is not compatible with any of the applied warm-glow models mentioned above.

2 Setup

We consider an individual, called **the agent**, in a society. There is one private good and, at most, one public good.[15] Set $X := \mathbb{R}^{\ell}_{+}$ where $\ell \geq 2$ is an integer. We refer to an element $x := (x_1, ..., x_\ell)$ of X as an **allocation.** The first component x_1 stands for the agent's private consumption. In turn, any other component x_i represents either the private consumption of another individual i or the amount of the public good (if it exists). Thus, ℓ equals the number of consumption variables related to the decision problem in question, and it can exceed the cardinality of the society by, at most, one.

The agent's preferences are described by a binary relation $\succsim$ over a collection of subsets of X denoted by $\mathcal{A}$. Each set in $\mathcal{A}$ represents a **menu** — that is, a set of allocations from which the agent will make a choice in a subsequent stage. Our representation theorems have certain implications about the agent's second-stage choice behavior. Accordingly, when discussing our axioms on $\succsim$, we will build upon a suitable interpretation of how the agent might be planning to behave in the second stage.

[14] Saito (2013) relates the intrinsic joy associated with a prosocial action to the notion of "pride." His model also accommodates "shame" and "temptation to act selfishly."

[15] In Section 4.2, we discuss how our theory can be extended to include multiple private and public goods.

The collection of all nonempty compact subsets of X is denoted by $\mathcal{K}$. We equip $\mathcal{K}$ with the Hausdorff metric d_H (induced by the Euclidean norm).[16] In representation (1), we will let $\mathcal{A} = \mathcal{K}$, but in the other representations we will take $\mathcal{A}$ as a proper subset of $\mathcal{K}$ by imposing some restrictions on the collection of menus. (See Section 3.2.)

Given a pair of allocations x, y, throughout the paper, $x \geq y$ means $x_i \geq y_i$ for $i = 1, ..., \ell$, while $x > y$ means $x \geq y$ and $x_i > y_i$ for some i. (For Euclidean spaces of other dimensions, the binary relations $\geq$ and $>$ are defined analogously.) Moreover, x_{-1} stands for the vector $(x_2, ..., x_\ell)$.

3 Basic Representation Theorems

In this section, we will provide axiomatic characterizations of representations (1) and (2). We start with a standard rationality requirement.

A1: Weak Order (WO). $\succsim$ is a complete and transitive binary relation on $\mathcal{A}$.

The next axiom states that increasing the size of a menu cannot harm the agent.

A2: Setwise Monotonicity (SM). For any $A, B \in \mathcal{A}$, if $A \supseteq B$, then $A \succsim B$.

A key assumption in the standard model of menu choice is the following:

$$A \cup B \sim A \quad \text{or} \quad A \cup B \sim B. \tag{3}$$

The logic behind this assumption consists of two parts: (i) if the agent can perfectly anticipate which alternative she would select from a menu, she should evaluate this menu solely with that particular alternative; and (ii) if the agent will select a given alternative from $A \cup B$, she must also select it from any subset (A or B) that contains the alternative. Hence, on one hand, property (3) describes a purely *instrumentalist* agent who views a menu solely as a means to her final choice. On the other hand, (3) also entails that the agent's second-stage choice behavior must be consistent with WARP. While the literature on freedom of choice raises normative objections to this instrumentalist view, the aforementioned evidence suggests that the degree of freedom offered by choice sets may influence the agent's

[16] $d_H(A, B) := \max\{\max_{x \in A} \min_{y \in B} \|x - y\|, \max_{y \in B} \min_{x \in A} \|x - y\|\}$ for any pair of nonempty, compact sets $A, B \subseteq \mathbb{R}^\ell$. (Here, $\|\cdot\|$ stands for the Euclidean norm.)

behavior in a way that is not consistent with WARP.[17] Therefore, we allow a menu to be strictly better than any of its subsets in the following particular way.

A3: Weak Instrumentalism (WI). Let $A, B, C \in \mathcal{A}$ and suppose that $C = A \cup B$. If there exists a $y \in A \cap B$ such that $y_1 \geq x_1$ for every $x \in C$, then $C \sim A$ or $C \sim B$.

Normatively, the idea behind this axiom is that if both A and B contain an allocation that secures the maximum possible private consumption offered by $A \cup B$, then joining A and B should not lead to a stronger experience of freedom to be selfish. In such cases, (WI) demands the agent's preferences to be consistent with property (3).

To see why (WI) is compatible with the experiments on freedom to be selfish, consider a set of three allocations $C := \{x, y, z\}$ (that belongs to $\mathcal{A}$) and suppose that $y_1 > z_1 > x_1$. Then, (WI) implies either $C \sim \{x, y\}$ or $C \sim \{y, z\}$. If only the former equivalence holds, we may as well have $\{y, z\} \prec C \succ \{x, z\}$ (in violation of (3)). Indeed, this is the pattern that we would expect if the agent were to select z from $\{x, z\}$ while selecting x from C. In turn, the related experimental findings describe precisely such a switch towards private consumption when subjects face choice sets that restrict their freedom to behave selfishly.

It should also be noted that in the scenario above, (WI) rules out the case $C \succ \{x, y\}$, although menu C provides a higher degree of freedom compared to $\{x, y\}$. In other words, the presence of z should not influence the agent's welfare despite the fact that $z_1 > x_1$. Rather, what matters is the maximum possible private consumption in a given menu and the allocation that the agent plans to select from that menu.[18] In line with this point, in Appendix B, we will show that for every menu A in $\mathcal{A}$ there exist a pair of allocations $x, y \in A$ such that $A \sim \{x, y\}$. Here, either x or y maximizes the private consumption of the agent over A, while the other allocation is interpreted as the agent's planned choice from A. (See Claim 2 and expression (10) in Appendix B.)

Although (WI) gives a special role to the maximum possible private consumption as a determinant of the agent's welfare, the axiom is silent about the nature of this relationship. How does the agent's welfare depend on the maximum possi-

[17] This is particularly evident in the experiment of Falk and Kosfeld (2006).

[18] In this regard, our approach is akin to that of several other papers, including Gul and Pesendorfer (2001) and Dillenberger and Sadowski (2012), although these papers are concerned with modeling preference for smaller menus. It is also worth noting that (WI) is a novel axiom, but in the papers that we just mentioned, an analogous property can be deduced from other axioms. Needless to say, our axioms are independent of each other.

ble private consumption, given the allocation that she plans to select? Our next axiom answers this question.

A4: Freedom to Be Selfish (FS). Let $x, y, y' \in X$ be such that $y_1 \geq x_1$, and suppose that $\{x, y\}$ and $\{x, y'\}$ both belong to $\mathcal{A}$. Then, $\{x\} \prec \{x, y\} \succ \{y\}$ and $y'_1 \geq y_1$ imply $\{x, y'\} \succsim \{x, y\}$.

In line with our earlier discussion, if $y_1 \geq x_1$, our interpretation of the pattern $\{x\} \prec \{x, y\} \succ \{y\}$ is that the agent plans to select x from $\{x, y\}$, and that this act is more valuable than the mere consumption of x because of the freedom to select y in the former case. Moreover, the logic of (WI) implies that in such cases, the utility of selecting x from $\{x, y\}$ should depend solely on y_1, the maximum possible private consumption. In turn, if this relationship is positive — i.e., if the freedom to choose a higher level of private consumption leads to higher welfare — then the agent should be better off whenever $\{x, y\}$ is replaced with a menu of the form $\{x, y'\}$ with $y'_1 \geq y_1$. Indeed, the agent can always select x from $\{x, y'\}$ while enjoying a stronger perception of freedom to be selfish than that entailed by $\{x, y\}$. This is the content of the axiom.

Observe that if $\{x\} \prec \{x, y\} \succ \{y\}$ and $y_1 \geq x_1$, the additional utility of selecting x from $\{x, y\}$, as opposed to the mere consumption of x, can be seen as the warm-glow payoff associated with the former act. Remarkably, (FS) implies that this notion of warm-glow is independent of how the agent's choice of an allocation compares with other available allocations in terms of other individuals' welfare. For instance, if $x_i - y_i$ is substantially larger than $x_i - y'_i$ for $i = 2, ..., \ell$, in terms of what the others receive, selecting x over y can be viewed as a much more generous act than selecting x over y'. If the intrinsic value of a prosocial action for the agent depends on the social consequences of that action, then the agent would typically prefer $\{x, y\}$ to $\{x, y'\}$.

From a conceptual point of view, this is certainly a serious limitation. In Appendix A, we will relax (FS), thereby providing an extension of our theory that incorporates such social concerns into the calculus of warm-glow. It should be noted, however, that the applied warm-glow literature is concerned mainly with phenomena that cannot be explained by one's concerns for others, presumably because the classical notion of altruism is already based on such concerns. Therefore, it will come as no surprise that the second-stage choice behavior compatible with (FS) is rich enough to subsume several, prominent models. (See Section 4.) Aside from being simpler, a further advantage of our current approach is that it takes as primitive the notion of preference for freedom to be selfish and derives the notion of warm-glow payoff as a necessary consequence. The more general model that we propose in Appendix A makes the connections between the notions of warm-glow and freedom of choice less transparent.

Our next axiom rules out negatively interdependent preferences over singletons.

A5: Singleton Monotonicity (SiM). $\{x\} \succsim \{y\}$ for any $x, y \in X$ with $x \geq y$.

It is worth noting that (SiM) also allows for a purely selfish attitude over singletons, as would be represented by the function $\{x\} \to x_1$.

We proceed with a standard continuity property.

A6: Continuity (C). For each $A \in \mathcal{A}$, the sets $\{B \in \mathcal{A} : B \succsim A\}$ and $\{B \in \mathcal{A} : A \succsim B\}$ are closed in $\mathcal{A}$.

The axioms that we have introduced so far are necessary conditions for both representations (1) and (2). The difference embodied in these representations lies in how the utility of $\{x, y\}$ changes with x when $\{x\} \prec \{x, y\} \succ \{y\}$ and $y_1 \geq x_1$. Next, we clarify this difference, and state full characterizations of the two representations.

3.1 A Representation of Preference for Freedom to Be Selfish

Representation (1) requires the following axiom, which we state for the case of a preference relation over $\mathcal{K}$.

A7: Coordinatewise Monotonicity (CM). Let $x, x', y \in X$ be such that $y_1 \geq x_1$. Then, $\{x\} \prec \{x, y\} \succ \{y\}$ and $x' \geq x$ imply $\{x', y\} \succsim \{x, y\}$.

Following our interpretation of (FS), this axiom formalizes the idea that if $x' \geq x$, and if the agent plans to selects x from a menu, then replacing x with x' would make the agent better off. Surely, such a change does not have negative implications for the agent's freedom to behave selfishly and would only improve the set of available allocations from a material point of view. Therefore, (CM) appears to be a natural property from an intuitive point of view.

Unfortunately, however, this property is not compatible with the evidence provided by the literature on motivation crowding out. For example, consider a pair of allocations $x, y \in X$, where x represents a prosocial allocation. Suppose that $\{x\} \prec \{x, y\} \succ \{y\}$, which, as usual, means that the agent plans to select x over y. As we discussed earlier, the related experiments suggest that offering an ill-advised private reward r for x could backfire; that is, the agent could select y from $\{(x_1 + r, x_{-1}), y\}$. But in this case, it would only be natural to expect the pattern $\{(x_1 + r, x_{-1}), y\} \sim \{y\}$, which contradicts (CM) given that $\{y\} \prec \{x, y\}$.

Putting aside this problem for a while, we now state our first representation theorem, which is based on (CM).

Theorem 1. *A binary relation* $\succsim$ *on* $\mathcal{A} := \mathcal{K}$ *satisfies the axioms* (A1)-(A7) *if and only if there exists a weakly increasing and continuous function* $U_{\mathrm{f}} : X \times \mathbb{R}_+ \to \mathbb{R}$ *such that, for each* $A, B \in \mathcal{K}$,

$$A \succsim B \quad \textit{iff} \quad \max_{x \in A} U_{\mathrm{f}}(x, \max_{y \in A} y_1) \geq \max_{x \in B} U_{\mathrm{f}}(x, \max_{y \in B} y_1).$$

In what follows, we refer to such a function U_{f} as an **f-index** for $\succsim$.

The next item in our agenda is to provide a characterization of representation (2) by modifying (CM) in line with the experiments on motivation crowding out.

3.2 Basic Warm-Glow Representation

Let us denote by $\mathcal{K}_{\mathcal{P}}$ the collection of all sets $A \in \mathcal{K}$ that satisfy the following two properties:

(i) A consists of Pareto efficient allocations; that is, $x, y \in A$ and $x \geq y$ imply that $x = y$.

(ii) There exists a $y^* \in A$ such that $x_i \geq y_i^*$ for every $x \in A$ and $i = 2, ..., \ell$.

In our next representation, we will assume that $\mathcal{A} = \mathcal{K}_{\mathcal{P}}$. Note that when ℓ equals 2, property (ii) trivially follows from (i). In this case, for any $A \in \mathcal{K}$ that consists of efficient allocations, there is a unique allocation y^* in A such that $y_1^* = \max\{x_1 : x \in A\}$, which simultaneously satisfies $y_2^* = \min\{x_2 : x \in A\}$. Property (ii) filters higher dimensional sets that have an analogous feature: *For each* $A \in \mathcal{K}_{\mathcal{P}}$, *there exists a unique allocation* $y^*(A)$ *in* A *such that* $y_1^*(A) = \max\{x_1 : x \in A\}$. *Moreover,* $y^*(A)$ *is also the unique allocation in* A *that satisfies (ii).* (We omit the proof of this simple observation.) In what follows, the **most selfish** option in a menu $A \in \mathcal{K}_{\mathcal{P}}$ refers to $y^*(A)$. The crucial implication of (ii) is that in the second stage, if the agent decides to select an allocation x with $x_1 < y_1^*(A)$, the private consumption that she gives up is converted into a public good or the private consumption of some other agents, without reducing the goods available to any other agent. Thus, $y^*(A)$ can also be seen as the *least generous* option available to the agent, in terms of her influence on others' consumption.

These restrictions on $\mathcal{A}$ lead to some technical problems because they reduce the power of property (C). We will solve these problems with the help of an additional continuity axiom that we will introduce shortly. In turn, the restrictions that we impose on $\mathcal{A}$ serve to avoid the aforementioned conceptual difficulties that come about when representation (2) is applied on an arbitrary set of allocations.

Our next representation requires the following modification of (CM).

A7*: Coordinatewise Monotonicity* (CM*). Let $x, x', y \in X$ be such that $y_1 \geq x_1$ and $\{x, y\} \in \mathcal{K}_{\mathcal{P}}$. Then, $\{x\} \prec \{x, y\} \succ \{y\}$ and $x' \geq x$ imply that $\{x', (a, y_{-1})\} \succsim \{x, y\}$ for every $a \geq y_1 + (x'_1 - x_1)$.

To gain insight, suppose, again, that the agent is offered a reward r if she decides to select a prosocial allocation x over another allocation y, which would effectively transform the choice set of the agent from $\{x, y\}$ to $\{(x_1 + r, x_{-1}), y\}$. Unlike (CM), property (CM*) allows for those cases in which such a reward can make the agent worse off, in line with the related experiments, as discussed earlier. More specifically, the axiom tells us that rather than "rewards," the agent would certainly enjoy "presents" that are not contingent on her behavior. Put formally, we can be sure that $\{(x_1 + r, x_{-1}), (y_1 + r, y_{-1})\} \succsim \{x, y\}$.

Our final axiom is a continuity property that we need because of the restrictions imposed on the domain of $\succsim$.

A8: Extension Continuity (EC). Let $x', y' \in X$ be such that $\{x', y'\} \in \mathcal{K} \backslash \mathcal{K}_{\mathcal{P}}$. Then, for any $A, B \in \mathcal{K}_{\mathcal{P}}$ with $A \succ B$, there exists a neighborhood $\mathcal{N}$ of $\{x', y'\}$ such that one of the following holds:

(i) $A \succ \{x, y\}$ for every $\{x, y\} \in \mathcal{N} \cap \mathcal{K}_{\mathcal{P}}$.

(ii) $\{x, y\} \succ B$ for every $\{x, y\} \in \mathcal{N} \cap \mathcal{K}_{\mathcal{P}}$.

To understand the axiom and its motivation, it should be noted that the representation that we seek requires the existence of a continuous weak order $\succsim'$ on $\mathcal{K}$ which coincides with $\succsim$ on $\mathcal{K}_{\mathcal{P}}$. The existence of such a binary relation $\succsim'$ can be assured only if $\succsim$ is well-behaved on $\mathcal{K}_{\mathcal{P}}$. It turns out that property (EC) is all we need to this end. The axiom tells us that if $A \succ B$, then all doubletons in $\mathcal{K}_{\mathcal{P}}$ that are sufficiently close to a set $\{x', y'\} \in \mathcal{K} \backslash \mathcal{K}_{\mathcal{P}}$ must either be strictly worse than A (which corresponds to the case $B \succsim' \{x', y'\}$) or strictly better than B (which corresponds to the case $\{x', y'\} \succ' B$.)

We are now ready to state our second representation theorem.

Theorem 2. *A binary relation $\succsim$ on $\mathcal{A} := \mathcal{K}_{\mathcal{P}}$ satisfies the axioms* (A1)-(A6), (A7*) *and* (A8) *if and only if there exists a weakly increasing and continuous* *function* $U_{\mathrm{w}} : X \times \mathbb{R}_+ \to \mathbb{R}$ *such that, for each* $A, B \in \mathcal{K}_{\mathcal{P}}$,

$$A \succsim B \quad \textit{iff} \quad \max_{x \in A} U_{\mathrm{w}}(x, \max_{y \in A} y_1 - x_1) \geq \max_{x \in B} U_{\mathrm{w}}(x, \max_{y \in B} y_1 - x_1).$$

Although Theorem 2 simply provides an alternative framework for modeling the notion of preference for freedom to be selfish, in Section 4, we will show that the implied second-stage choice behavior subsumes some prominent warm-glow models. In what follows, **warm-glow representation** refers to the representation notion characterized in Theorem 2. Moreover, **w-index** will refer to a function U_{w} as in the theorem.

To recapitulate, the term $U_{\mathrm{w}}\left(x, \max_{y \in A} y_1 - x_1\right)$ gives us the utility associated with the act of selecting x from a menu A. Accordingly, the least satisfying menu that admits the choice of x is the singleton $\{x\}$. This menu leads to utility $U_{\mathrm{w}}\left(x, 0\right)$, which represents the payoff resulting merely from the consumption of x. In turn, the **warm-glow payoff** associated with the act of selecting x from A is defined as $U_{\mathrm{w}}\left(x, \max_{y \in A} y_1 - x_1\right) - U_{\mathrm{w}}\left(x, 0\right)$. A key feature of the representation is that the warm-glow payoff is a weakly increasing function of the difference between the maximum possible private consumption that the agent can attain and her actual choice of private consumption. We denote by λ the last argument of U_{w}, which corresponds to this difference.

It is important to note that if $U_{\mathrm{w}}(x, \lambda)$ is constant in λ for each x, then the agent never experiences warm-glow payoff. In this case, the representation reduces to the classical utility maximization: $A \succsim B$ iff $\max_{x \in A} U\left(x, 0\right) \geq \max_{x \in B} U\left(x, 0\right)$ for every $A, B \in \mathcal{A}$. Of course, $\succsim$ admits such a w-index if and only if (3) holds for any pair of menus. In turn, the corresponding agent can be considered *purely altruistic* (unless U_{w} is merely a function of x_1). Another case of special interest is when U_{w} depends only on x_1 and λ. This corresponds to a *purely egoistic* agent who is motivated solely by warm-glow and her private consumption.

We proceed with a technical note that proves useful in what follows.

Lemma 1. *Let* $\succsim$ *be a binary relation on* $\mathcal{A} := \mathcal{K}_{\mathcal{P}}$ *that satisfies the axioms* (A1)-(A6), (A7*) *and* (A8). *Then, for any continuous function* $V : \mathcal{K}_{\mathcal{P}} \to \mathbb{R}$ *that represents* $\succsim$, *there exists a w-index* U_{w} *such that* $V(A) = \max_{x \in A} U_{\mathrm{w}}(x, \max_{y \in A} y_1 - x_1)$ *for every* $A \in \mathcal{K}_{\mathcal{P}}$.[19]

We close this section with a discussion of the restrictions that we imposed on

[19]The proof of this lemma is implicit in the proof of Theorem 2. (See Appendix B.)

the domain of $\succsim$.

3.2.1 On the Assumption $\mathcal{A} = \mathcal{K}_{\mathcal{P}}$

Unlike Theorem 1, the second-stage choice behavior implied by our warm-glow representation leads to some awkward conclusions when applied to an arbitrary set of allocations. The issue is that if a set A contains an allocation x' and another allocation of the form $(x'_1 - r, x'_{-1})$ for some $r > 0$, the solution of the optimization problem $\max_{x \in A} U_{\text{w}}(x, \max_{y \in A} y_1 - x_1)$ may well be $(x'_1 - r, x'_{-1})$, despite the fact that this allocation is Pareto dominated by x'. In an earlier version of this paper, we proposed an alternative warm-glow representation that restricts the calculus of warm-glow to efficient allocations. However, this is, at best, only a partial solution to the underlying difficulty. For example, it would be an equally awkward situation if the agent were to select an allocation of the form $(x'_1 - r, x'_2 + \varepsilon, x'_3 - r', ..., x'_\ell - r')$ over x', which can hardly be seen as a prosocial action, especially if r' is not so much smaller than ε. At a fundamental level, these difficulties stem from the fact that property (CM*) is designed to model a decision maker who enjoys making a self-sacrifice on occasion (regardless of the social consequences of the action in question). For our purposes, it is necessary to allow for such preferences because in the experiments on motivation crowding out, the private reward (coming from the experimenter's pocket) typically has negligible social consequences, and yet it proves detrimental for the agent by reducing the private cost of prosocial behavior. That is, the strength of warm-glow that people experience does, indeed, seem to depend on their private sacrifice (presumably, in addition to the social implications of the action in question).[20]

In view of these observations, it seems to be in order to restrict our attention to those cases in which the agent's private sacrifice leads to an unambiguous improvement in others' welfare, no matter how small this improvement might be. This is precisely the role of assumptions (i) and (ii), which define the class of relevant menus, $\mathcal{A} = \mathcal{K}_{\mathcal{P}}$.

It should also be noted that this framework remains rich enough to cover many interesting applications. For example, in applied models of charity, the agent often has an initial endowment of the private good, and the choice set A

[20] The representation that we will propose in Appendix A incorporates social concerns into the calculus of warm-glow, but this is simply a more general version of Theorem 2 that still relates the strength of warm-glow experience to the agent's private sacrifice. Hence, in line with our remarks above, this more general approach is also of little help with regard to the problem at hand.

in question consists of all allocations that the agent can obtain by distributing her endowment among the ℓ consumption variables, given other factors such as government transfers, prices, and the technology that transforms the private good into the public good. Such choice sets are within the scope of our analysis, for by privately consuming all her endowment, the agent typically can maximize her private consumption while minimizing her contributions to all other variables. In addition, in the voting literature, the case of paternalistic citizens provides another suitable setting that has attracted considerable attention. (More on this in Section 4.)

4 On Applied Warm-Glow Models

4.1 Andreoni's Model

Andreoni (1989, 1990) studies a game on public good provision between a set of individuals $\{1, ..., I\}$. He assumes that there is one public good and one private good and that one unit of the private good can be converted into one unit of the public good with a linear technology. Each individual i is endowed with an amount w_i of the private good (or, equivalently, w_i units of dollars) that she can allocate between her private consumption, x_i, and her gift to the public good, g_i. Moreover, the government levies lump sum taxes τ_i. So, $G := \sum_{i=1}^{I} g_i$ is the total private contributions to the public good, and $T := \sum_{i=1}^{I} \tau_i$ is the total tax receipts that are fully used for the provision of the public good. A generic agent — say, agent 1 — takes as given the private consumption and gifts of others, $(\overline{x}_2, \overline{g}_2), ..., (\overline{x}_I, \overline{g}_I)$, and chooses a consumption-gift pair (x_1, g_1) so as to solve an optimization problem of the following form:

$$\max \mathcal{U}(x_1, G + T, g_1) \quad \text{subject to} \quad \begin{aligned} & x_1 + g_1 + \tau_1 = w_1 \\ & \text{and } 0 \leq x_1 \leq w_1 - \tau_1. \end{aligned} \tag{4}$$

Here, $\mathcal{U}$ is a weakly increasing function on $\mathbb{R}^3_+$, which captures altruistic concerns[21] and warm-glow experience by its second and third arguments, respectively.

[21]As Andreoni (1989, 1990) points out, the private consumption of a given individual would act as if it were a public good from others' perspective when others are altruistic in the classic sense. Therefore, in the literature on philanthropy, it is customary to view one's concern for the public good as a form of altruism. By the same token, the models often take into account either one's concern for others' private consumption, as in Roberts (1984), or one's concern for the public good, as in Andreoni (1989, 1990), but not both. For conceptual clarity, in the present paper, we have chosen to refer to a public good separately.

In our terminology, then, agent 1 faces the menu

$$A := \{(x_1, \overline{x}_2, ..., \overline{x}_I, G+T) : 0 \leq x_1 \leq w_1 - \tau_1,\ x_1 + g_1 + \tau_1 = w_1,$$
$$G + T = \tau_1 + g_1 + \textstyle\sum_{i=2}^{I} \tau_i + \overline{g}_i\}.$$

Clearly, with $X := \mathbb{R}_+^{I+1}$, this menu belongs to $\mathcal{K}_\mathcal{P}$ and the most selfish allocation, $y^*(A)$, equals $(w_1 - \tau_1, \overline{x}_2, ..., \overline{x}_I, \tau_1 + \sum_{i=2}^{I} \tau_i + \overline{g}_i)$. Thus, upon solving for g_1 in the budget constraint, we see that $g_1 = y_1^*(A) - x_1$. That is, g_1 is simply the last argument of a w-index in our terminology. The function $U_{\mathrm{w}}(x, \lambda) := \mathcal{U}(x_1, x_{I+1}, \lambda)$, defined on $X \times \mathbb{R}_+$, qualifies as a w-index, and the agent's allocation choice maximizes $U_{\mathrm{w}}(x, y_1^*(A) - x_1)$ over the menu A.[22] To summarize, second-stage behavior implied by our warm-glow representation subsumes Andreoni's model.

The main contribution of Andreoni's model is that, under suitable assumptions, it makes the equilibrium amount of the public good sensitive to fiscal policies and income distribution. This differs from the corresponding models of pure altruism, which predict that government grants and subsidies should crowd out voluntary contributions dollar-for-dollar and that the total supply of the public good should be independent of the income distribution.[23] Andreoni's approach is supported by substantial empirical evidence on incomplete crowding out (Abrams and Schmitz, 1978, 1984; Clotfelter, 1985; Steinberg, 1989) and non-neutrality of income distribution (Hochman and Rodgers, 1973).

While Andreoni's findings are based on some reasonable assumptions on the form of the utility indices, these assumptions may be restrictive from a foundational point of view. For instance, Andreoni assumes that the private consumption and the gift of an agent are both strictly increasing functions of her wealth, which rules out quasilinear utility indices. Indeed, it can easily be seen that in problem (4), the allocation choice implied by the purely *altruistic* utility index $\mathcal{U} = u(x_1) + G + T$ would simply coincide with that induced by the purely *egoistic*

[22]Although we have set $\ell := I + 1$ for the domain of the preference relation, in this particular setup, we could also let $\ell = 2$ since the agents cannot influence the private consumption of others.

[23]For theoretical findings on crowding out under pure altruism, see Warr (1982), Roberts (1984), Bernheim (1986), and Andreoni (1988), among others. Neutrality of income distribution under pure altruism has been demonstrated by Warr (1983) and Bergstrom et al. (1986). However, these findings are subject to some exceptions: if only a subset of the agents make donations, government spending and income distribution may influence the equilibrium amount of the public good (Bergstrom et al., 1986). Moreover, under alternative tax schemes (as opposed to the lump-sum taxes that we discussed above), government subsidies may also be effective (Andreoni and Bergstrom, 1996).

utility index $\mathcal{U} = u(x_1) + g_1$. On a related note, Bergstrom, Blume, and Varian (1986, Section 2) emphasize that with quasi-homothetic utility indices, income transfers would be neutral even in a model of impure altruism like Andreoni's. (For a related finding, see, also, Proposition 2 of Andreoni (1990)). Finally, we should recall that taking into account the boundary solutions further complicates the task of distinguishing pure from impure altruism (see footnote 23).

In view of these remarks, our menu-based approach not only provides foundations for Andreoni's model, but also arms us with clear-cut distinctions between purely altruistic agents and those motivated by warm-glow. Indeed, to test the hypothesis of pure altruism, in an economic experiment, one can simply check whether the subjects violate property (3) systematically. Moreover, our axioms on preferences over menus can be tested with suitably-designed survey questions (see Section 6). Next, we show that when applied to consumption-saving problems, our theory can replicate the qualitative predictions of Andreoni (1990) even if these predictions would no longer be valid in an analogous static model.

4.1.1 Bequest Giving with Quasilinear Utility Indices

Consider two generations within a family: parents and an heir. In period 1, the parents allocate their wealth, w_0, between their private consumption, x_0, and saving, $w_1 = w_0 - x_0$. At the beginning of period 2, they receive an income support $\rho(w_1)$, which is financed by a tax on the heir. We assume that $\rho : \mathbb{R}_+ \to [0, w_0]$ is a differentiable function. The parents allocate their adjusted income between their period 2 consumption, x_1, and a bequest, $g_1 = w_1 + \rho(w_1) - x_1$. The heir's initial wealth also equals w_0. She moves last and consumes all of her adjusted income, $x_2^* = w_0 + g_1 - \rho(w_1) = w_0 + w_1 - x_1$. We now examine the parents' behavior in a subgame perfect equilibrium.

First of all, the menu that the parents face in period 2 takes the form

$$A(x_0, w_0) := \{(x_1, x_2^*) : 0 \leq x_1 \leq w_1 + \rho(w_1),\ x_2^* = w_0 + w_1 - x_1\}.$$

This menu belongs to $\mathcal{K}_{\mathcal{P}}$ with $X := \mathbb{R}^2_+$, and the most selfish allocation is given by $(w_1 + \rho(w_1), w_0 - \rho(w_1))$. Thus, we also see that $g_1 = y_1^*(A(x_0, w_0)) - x_1$.

The parents' problem in period 1 is to make a choice among the pairs of the form $(x_0, A(x_0, w_0))$. Let W be a utility function over $\{(x_0, A) : x_0 \in \mathbb{R}_+,\ A \in \mathcal{A}\}$ that represents the parents' preferences. In view of Lemma 1, if these preferences restricted to $\{(x_0, A) : A \in \mathcal{A}\}$ satisfy the properties (A1)-(A6), (A7*), and (A8) for each x_0, we can find a w-index $U_{x_0} : \mathbb{R}^3_+ \to \mathbb{R}$ such that

$$W(x_0, A(x_0, w_0)) = \max\{U_{x_0}(x_1, x_2, g_1) : (x_1, x_2) \in A(x_0, w_0)\}. \quad (5)$$

Let us now consider a purely egoistic utility index $U^e_{x_0} = u(x_0) + u(x_1) + g_1$, and a purely altruistic one $U^a_{x_0} = u(x_0) + u(x_1) + x_2$, where $u : \mathbb{R}_+ \to \mathbb{R}_+$ is a function that satisfies the Inada conditions. Just as in the corresponding model of Andreoni, after substituting for g_1 and x^*_2, we immediately see that for any fixed (x_0, w_0), the maximizers of $U^a_{x_0}$ and $U^e_{x_0}$ over the set $A(x_0, w_0)$ coincide. That is, in this setup, we cannot distinguish between the two types of parents based on period 2 behavior.

However, the saving behavior of the two types are typically different, because the income support influences the marginal value of saving for egoistic parents by altering their perception of freedom in period 2. Indeed, among the interior solutions of period 2 (which correspond to large values of adjusted income $w_1 + \rho(w_1)$), the value on the right side of (5) takes the form $u(x_0) + u(\overline{x}_1) + w_1 + \rho(w_1) - \overline{x}_1$ for egoistic parents, while it takes the form $u(x_0) + u(\overline{x}_1) + w_0 + w_1 - \overline{x}_1$ for altruistic parents (here, $\overline{x}_1$ is the number that satisfies $u'(\overline{x}_1) = 1$). Thus, the marginal value of saving equals 1 for altruistic parents and $1 + \rho'(w_1)$ for egoistic parents. In particular, if ρ is a decreasing function of w_1 (which corresponds to a progressive income support), the marginal value of saving for egoistic parents is smaller, and, hence, they save less than altruistic parents. If $w_1 + \rho(w_1)$ is increasing in w_1, this also implies that egoistic parents leave a smaller bequest. Moreover, while ρ is neutral in the case of altruistic parents, the saving of egoistic parents and the consumption of the heir increase with an upward shift in $\rho'(\cdot)$.

One conclusion that follows from this exercise is that dynamic models can sharpen the predictions of earlier static models on warm-glow. In particular, we can explain non-neutrality of redistributive policies even with quasilinear utility functions. Furthermore, the welfare implications of a given policy can differ fundamentally, depending on whether the agents are altruistic or motivated by warm-glow.

4.2 Alternative Sets of Social Outcomes

Before discussing another application, we need to clarify how our theory can be extended to alternative sets of social outcomes. To this end, suppose that the set of allocations X is of the form $X = X_1 \times \cdots \times X_\ell$, where X_i is a separable metric space for each i. Then, under suitable assumptions on the behavior of $\succsim$ over the collection of singletons $\{\{x\} : x \in X\}$, we can find an aggregator $\varphi : \mathbb{R}^\ell \to \mathbb{R}$ and functions $\pi_i : X_i \to \mathbb{R}$ for $i = 1, ..., \ell$, such that $\{x\} \succsim \{y\}$ if and only if $\varphi(\pi_1(x_1), ..., \pi_\ell(x_\ell)) \geq \varphi(\pi_1(y_\ell), ..., \pi_\ell(y_\ell))$.[24] If we abstract from public goods so that x_i corresponds to the private consumption of individual i (which may also be a random variable), just as in Harsanyi's (1953, 1955) theory of utilitarianism, it may be appropriate to interpret π_i as a measure of well-being of individual i from the perspective of the decision maker in question, who acts as a social planner. In fact, that π_i depends solely on x_i would suggest that one view this function as the material payoff of individual i.

Once we agree on this interpretation, we can restate properties (i) and (ii) that define the collection of relevant menus in Section 3.2, as well as the axioms (A1)-(A6), (A7*), and (A8) in terms of the payoff vectors $(\pi_1(x_1), ..., \pi_\ell(x_\ell))$ and utility possibility sets of the form $\{(\pi_1(x_1), ..., \pi_\ell(x_\ell)) : x \in A\} \subseteq \mathbb{R}^\ell$. In particular, we could let $y^{*\pi}(A)$ be an allocation that maximizes the function π_1 over a qualifying menu A, and give the role of $y_1^*(A) - x_1$ in the warm-glow representation of Theorem 2 to the difference $\pi_1(y_1^{*\pi}(A)) - \pi_1(x_1)$. By pursuing this approach, it is a straightforward exercise to obtain an extension of Theorem 2 that delivers a utility representation of the form

$$V_\pi(A) := \max_{x \in A} U\left(\pi_1(x_1), ..., \pi_\ell(x_\ell), \pi_1\left(y_1^{*\pi}(A)\right) - \pi_1(x_1)\right)$$

for a function $U : \mathbb{R}^\ell \times \mathbb{R}_+ \to \mathbb{R}$ (we omit the details of this derivation).

Remark 1. When individuals' utility from private and public goods can be separated from each other, the above argument can also be applied in a framework with a finite number of public goods.

4.3 Voting as a Civic Duty

Explaining voter turnout in large elections has been a major challenge for political economists. The difficulty stems from the fact that when many people vote, the probability of a single voter being decisive (pivotal) is close to zero, whereas voting incurs significant costs. In an earlier attempt to resolve this paradox, Riker and

[24] A large body of literature is devoted to the study of axiomatic foundations of such representations that also demand the aggregator to be additive (see Wakker (1989, Chapter 3) and references therein). In turn, a nonadditive form of the representation can be derived by imposing a weak separability property along the lines of Mak (1984).

Ordeshook (1968) suggest that the act of voting may be valuable *per se*, as the citizens may perceive it as a civic duty.

Suppose that there are two candidates, L and R, and that the agent in question prefers candidate L. Specifically, let us assume that the victory of L will bring a material payoff $\mathfrak{u} > 0$ to our agent, whereas the victory of R is worth 0. Given other voters' behavior, let $p_j > 0$ be the probability of being pivotal for the agent if she votes for candidate j, and let P be the probability of winning for candidate L if she abstains. Finally, let c denote the cost of voting and d the payoff associated with the act of voting, as posited by Riker and Ordeshook.

The implied expected payoff scheme reads as follows:

$$\begin{array}{ll} (P + p_L)\,\mathfrak{u} - c + d & \text{if the agent votes for } L, \\ P\mathfrak{u} & \text{if the agent abstains,} \\ (P - p_R)\,\mathfrak{u} - c + d & \text{if the agent votes for } R. \end{array}$$

Thus, the agent would never vote for R, while the decision between abstaining and voting for L is determined by the following simple rule:

$$\text{vote for } L \quad \text{if and only if} \quad p_L\mathfrak{u} + d \geq c.$$

In particular, no matter how small p_L might be, our agent would vote if $d \geq c$.

While it is widely accepted that voters may be motivated by a sense of duty, some scholars have recently proposed extensions that can explain several other aspects of voters' behavior, as well as the high turnout rates themselves (see, e.g., Coate and Conlin, 2004; Feddersen and Sandroni, 2006). These alternative models are sensitive to the specification of voters' statistical distribution since they relate the turnout rate of a group of individuals to their likelihood of influencing the election outcome.[25] Riker-Ordeshook's approach, however, is compatible with high turnout rates irrespective of how an individual or a group of individuals might influence the election outcome. We will now show how our warm-glow representation in Theorem 2 can reproduce the calculus of voting suggested by Riker and Ordeshook.

Following Section 4.2, let X_i be the space of lotteries over the real line, and π_i be the expectation operator over X_i. Each action a available to the agent in question, individual 1, induces a vector of lotteries $x(a) \in X_1 \times \cdots \times X_\ell$, given the behavior of other $\ell - 1$ voters. So, the agent evaluates action a with the associated

[25] Evren (2012) provides a discussion of the role of voters' distribution in these recent models.

expected payoff vector $(\pi_1(x_1(a)), ..., \pi_\ell(x_\ell(a)))$.

In a paternalistic fashion, the agent believes that the victory of L will contribute to the (material) payoff of everyone in the society, implying that $\pi_i(x_i(\text{vote for } L)) > \pi_i(x_i(\text{vote for } R))$ for every i.[26] Moreover, as before, the victory of candidate L is worth $\mathfrak{u} > 0$ for the agent herself, so that

$$\pi_1(x_1(\text{vote for } L)) = (P + p_L)\,\mathfrak{u} - c \quad \text{and} \quad \pi_1(x_1(\text{abstain})) = P\mathfrak{u}.$$

It follows that when p_L is small, as would be the case in a large election, and if $c > 0$, the agent's expected payoff would be higher if she abstains. On the other hand, the agent believes that if she were to vote for L, she would be contributing to the expected payoff of everyone else, as we just noted. Thus, the menu of lottery vectors $\{x(\text{vote for } L),\ x(\text{abstain})\}$ belongs to $\mathcal{K}_{\mathcal{P}}$ in our extended theory, and $x(\text{abstain})$ is the most selfish option. In turn, the corresponding warm-glow component is given by $\pi_1(x_1(\text{abstain})) - \pi_1(x_1(\text{vote for } L)) = c - p_L\mathfrak{u} > 0$.

As a final step, let us suppose that the w-index of the agent is of the form $U_{\text{w}} = \pi_1 + f(\lambda)$, so that we have a purely *egoistic* agent at hand. Then, according to our extended theory, the agent should solve the following problem:

$$\max\{\pi_1(x_1(\text{vote for } L)) + f(c - p_L\mathfrak{u}),\ \pi_1(x_1(\text{abstain})) + f(0)\}.$$

That is, the agent should vote iff $p_L\mathfrak{u} + f(c - p_L\mathfrak{u}) - f(0) \geq c$. Also note that if f is continuous, $f(c - p_L\mathfrak{u})$ will be approximately equal to $f(c)$ for small values of p_L. Thus, the parameter d in the Riker-Ordeshook model simply corresponds to the warm-glow payoff $f(c - p_L\mathfrak{u}) - f(0) \approx f(c) - f(0)$.

Beyond the technical details, our theory endogenizes the parameter d of Riker and Ordeshook by viewing the act of voting as a selfless action taken by *free will.* Indeed, if citizens were forced to vote — say, by a prohibitively high fine on abstention — it would seem reasonable to assume that they would not attribute an intrinsic value to the act of voting. This is precisely what our model predicts: Given a fine ϕ on abstention, the difference $\pi_1(x_1(\text{abstain})) - \pi_1(x_1(\text{vote for } L))$ reduces to $c - p_L\mathfrak{u} - \phi$, leading to a smaller warm-glow payoff $f(c - p_L\mathfrak{u} - \phi) - f(0)$. That is, a fine on abstention crowds out voters' intrinsic motivation. In a dual fashion, a policy that aims to reduce voting costs would crowd out intrinsic motivations through the same mechanism. As Bénabou and Tirole (2006) also point out, this

[26]The assumption of paternalism is quite common in the voting literature (see, e.g., Feddersen and Sandroni, 2006; Faravelli and Walsh, 2011; Evren, 2012; Myatt, 2012).

phenomenon seems to underlie Funk's (2010) findings, which show that the introduction of mail voting in Switzerland failed to raise the turnout rates in some communities.

5 Uniqueness of Second-Stage Choice Behavior

As we saw in the previous section, applied warm-glow models often focus on the social consequences of individuals' behavior. Thus, it is of major importance to determine the extent of uniqueness of the second-stage choice behavior implied by our representations. This section addresses precisely this issue. For brevity, we will focus on our warm-glow representation, Theorem 2, but the analysis can be readily extended to include the second-stage behavior implied by Theorem 1.

As in Section 3, let $\succsim$ be a binary relation on $\mathcal{A}$ that satisfies properties (A1)-(A6), (A7*), and (A8), and let U be a w-index for $\succsim$. The representation suggests that when faced with a menu $A \in \mathcal{A}$, in the second stage the agent's potential choices would coincide with the following set:

$$\mathbf{C}_U(A) := \left\{ \widehat{x} \in A : U\left(\widehat{x}, y_1^*(A) - \widehat{x}_1\right) = \max_{x \in A} U\left(x, y_1^*(A) - x_1\right) \right\}.$$

Observe that, when $\{x\} \prec \{x, y\} \sim \{y\}$ and $y_1 > x_1$, we cannot pin down how $U(x, y_1 - x_1)$ compares with $U(x, 0)$ and $U(y, 0)$. In particular, depending on the choice of the w-index, we may have either $U(x, y_1 - x_1) = U(y, 0)$ or $U(x, y_1 - x_1) < U(y, 0)$. In both cases, it would follow that the agent may select y from $\{x, y\}$, but whether she could *also* select x depends on the choice of the w-index. However, when $\{x, y\} \succ \{y\}$ and $y_1 > x_1$, we must certainly have $U(x, y_1 - x_1) > U(y, 0)$, so that x can be identified as the unique choice from $\{x, y\}$.

These observations readily extend to arbitrary menus. That is, for any $A \in \mathcal{A}$, if the most selfish option does not belong to $\mathbf{C}_U(A)$, then we have $\mathbf{C}_U(A) = \mathbf{C}_{\widetilde{U}}(A)$ for any other w-index $\widetilde{U}$. In particular, $\mathbf{C}_U(A)$ contains the most selfish option if and only if this is the case for any other w-index. What remains undetermined is if (and which) other allocations can be selected along with the most selfish option when the latter belongs to the choice correspondence:

Proposition 1. *Let U and $\widetilde{U}$ be a pair of w-indices for $\succsim$. Then, for any $A \in \mathcal{A}$,*

(i) $y^*(A) \notin \mathbf{C}_U(A)$ *implies* $\mathbf{C}_U(A) = \mathbf{C}_{\widetilde{U}}(A)$*;*

(ii) $y^*(A) \in \mathbf{C}_U(A)$ *if, and only if,* $y^*(A) \in \mathbf{C}_{\widetilde{U}}(A)$.

The level of identification determined by Proposition 1 seems to be quite satisfactory. In particular, the intersection of all compatible choice correspondences is always nonempty. Put formally, for any $A \in \mathcal{A}$, the set

$$\bigcap \{\mathbf{C}_U(A) : U \text{ is a w-index for } \succsim\}$$

either contains $y^*(A)$ or equals $\mathbf{C}_U(A)$ for an arbitrary w-index U. It also follows that for any pair of w-indices U and $\widetilde{U}$, whenever both $\mathbf{C}_U(A)$ and $\mathbf{C}_{\widetilde{U}}(A)$ consist of single allocations, we must have $\mathbf{C}_U(A) = \mathbf{C}_{\widetilde{U}}(A)$.

Yet it may be of interest to note that we can obtain perfect identification for utility indices that satisfy the following additional property.

Regularity. Let $\{x, y\} \in \mathcal{K}_{\mathcal{P}}$ be such that $U(x, y_1 - x_1) = U(y, 0)$ and $y_1 > x_1$. Then, any neighborhood of $\{x, y\}$ contains a pair of allocations $\{x', y'\} \in \mathcal{K}_{\mathcal{P}}$ such that $U(x', y'_1 - x'_1) > U(y', 0)$ and $y'_1 > x'_1$.

In what follows, we say that a w-index is **regular** if it satisfies the above property.

The notion of regularity is a variant of the local non-satiation property familiar from the classical consumer theory. As we will see momentarily, the regularity notion proves quite general, even outside the classical model. Before presenting some examples in this direction, we state our identification result for regular w-indices:

Proposition 2. *Let* U *and* $\widetilde{U}$ *be a pair of regular w-indices for* $\succsim$*. Then,* $\mathbf{C}_U(A) = \mathbf{C}_{\widetilde{U}}(A)$ *for any* $A \in \mathcal{A}$.

It is a simple exercise to verify that in each of the following cases, the w-index in question is regular.

Example 1 (Classical Altruism). Let U be a w-index that is constant in λ, and suppose that $U(\widehat{x}, 0) > U(x, 0)$ whenever $\widehat{x}_i > x_i$ for $i = 1, ..., \ell$.

Example 2 (Pure Egoism). Let U be a w-index and $u : \mathbb{R}^2_+ \to \mathbb{R}$ be a strictly quasiconcave function such that $U(x, \lambda) = u(x_1, \lambda)$ for every $(x, \lambda) \in X \times \mathbb{R}_+$.[27]

[27] Note that for a given strictly concave function f on $\mathbb{R}_+$, both functions $u_1 = f(x_1) + \lambda$ and $u_2 = x_1 + f(\lambda)$ are strictly quasiconcave on $\mathbb{R}^2_+$. Hence, example (2) also includes such quasilinear functions. Moreover, quasilinearity of the latter form might be especially important in an extended version of our model based on expected material payoffs, as we discussed in Section 4.3 above.

Example 3 (Impure Altruism). Let U be a w-index such that $U(\widehat{x},\lambda) > U(x,\lambda)$ whenever $\widehat{x}_1 \geq x_1$ and $\widehat{x}_i > x_i$ for $i = 2, ..., \ell$.

In view of these examples, (1) in the classical model, monotonicity of a utility index implies its regularity; (2) under pure egoism, strict quasiconcavity implies regularity; and (3) even an impure form of altruism suffices for regularity. Thus, it appears that one would rarely encounter a non-regular utility index in applications. We should note, however, that it is a nontrivial problem to obtain a characterization of preference relations that admit a regular utility index, for the definition of regularity refers to the condition $U(x, y_1 - x_1) = U(y, 0)$. In turn, this equality implies $\{x, y\} \sim \{y\}$, but the converse does not hold. We do not pursue this problem further in the present paper.

The next remark describes the regularity notion for f-indices of Theorem 1. Propositions 1 and 2 can be extended accordingly for such utility indices, in a straightforward way.

Remark 2 (Regularity of an f-index). Let $\{x, y\} \in \mathcal{A}$ be such that $U_{\mathrm{f}}(x, y_1) = U_{\mathrm{f}}(y, y_1)$ and $y_1 > x_1$. Then, any neighborhood of $\{x, y\}$ contains a pair of allocations $\{x', y'\} \in \mathcal{A}$ such that $U_{\mathrm{f}}(x', y'_1) > U_{\mathrm{f}}(y', y'_1)$ and $y'_1 > x'_1$.

6 First-Stage vs Second-Stage

As we observed in the Introduction, it is often necessary to interpret our theory as a model of psychological preferences over menus, as opposed to a model of menu choice. Indeed, the agent's perception of her freedom in a first-stage choice situation would typically influence her choice behavior, but our model does not accommodate this fact. Similarly, we do not account for the fact that choosing a menu $\{x\}$ over a menu $\{y\}$ may lead to a warm-glow experience (just like the act of choosing x from $\{x, y\}$). However, the agent's psychological preference relation over menus should be free from such behavioral phenomena, as we assume here.[28] We believe that this is a reasonable restriction since, to the best of our knowledge, in the earlier literature all applications of the notion of warm-glow focus on the agent's second-stage behavior. In turn, our theory of first-stage preferences lays the foundations for related forms of second-stage behavior and may allow one

[28] If we were to extend our representations to include non-instrumental motives solely associated with first-stage behavior, our current representations would correspond to utility of selecting a menu A when only A is available. Indeed, in such a situation, the act of selecting the menu A would merely be a necessity, which would not lead to a non-instrumental payoff by itself.

to study welfare implications of second-party actions that can affect the set of alternatives available to the agent.

Moreover, we could elicit the agent's preference relation over menus with suitable survey questions. For example we could ask the following survey question: "Suppose that the government is planning to give you a set of allocations from which you can make a selection in a subsequent stage. Which set would you like to receive, A or B? Your answer to this question will be kept confidential and, hence, cannot influence the government's behavior."[29]

Finally, as noted in the Introduction, if only the second-stage behavior is publicly observable, our theory can also be interpreted as a model of menu choice. Another way of facilitating this interpretation could be to focus on additive forms of our representations. For example, Sarver (2008) faces a similar issue in his model of regret, but the additive form of his representation allows him to argue that even if he were to extend his model to include regret considerations in the first stage, the implied choice behavior would be consistent with his model of preferences over menus. We do not pursue this approach here, as the implied behavior under such additive forms would be consistent with WARP in both stages, while we need to allow for violations of WARP in the second stage.

7 Relations to Kreps' Model of Preference for Flexibility

Following Kreps' (1979) pioneering work, the literature on preference for flexibility also focuses on decision makers who prefer a menu to all of its subsets (see, e.g., Dekel et al., 2001; Epstein et al., 2007; Ahn and Sarver, 2013). This literature attributes violations of property (3) to uncertainty of future preference relations over the set of alternatives. The decision maker in question is concerned solely with her final choices, just as in the case of pure altruism. Yet she still exhibits a preference for larger sets since, on occasion, she cannot precisely predict which alternative she would select from a given menu in period 2. In particular, instances of the form $B \prec B \cup \{x\} \succ \{x\}$ correspond precisely to those cases in which the agent is unsure whether she would select x or an element of B when faced with $B \cup \{x\}$.

[29] In a different setup, Gilboa, Maccheroni, Marinacci, and Schmeidler (2010) propose a similar method based on the agent's statements to recover her psychological preference relation, which is assumed to be incomplete.

By stark contrast, unless x is the most selfish option in $B \cup \{x\}$, whenever $B \prec B \cup \{x\}$, our model predicts that the agent would *certainly* select x from $B \cup \{x\}$. In turn, if x is the most selfish option, $B \cup \{x\} \succ \{x\}$ implies that the agent would *not* select x from $B \cup \{x\}$. These observations point to a clear-cut distinction between the two models, to the extent that in period 2, one can verify the random choice behavior that the preference-for-flexibility model predicts. More generally, holding fixed the most selfish option, the second-stage behavior implied by our theory is within the scope of the standard choice model, whereas the preference-for-flexibility approach predicts a stochastic behavior, as in McFadden and Richter (1990) or McFadden (2005).

If one focuses solely on preferences over menus, the difference between the two approaches becomes less stark. In particular, we can think of the following utility representation, which lies at the intersection of our Theorem 1 and Kreps' (1979) representation theorem:

$$V_K(A) := \alpha \max_{x \in A} u^a(x) + (1-\alpha) u^e(y_1^*(A)),$$

where $\alpha \in (0,1)$, $u^a : \mathbb{R}_+^k \to \mathbb{R}$ is weakly increasing (and continuous), and $u^e : \mathbb{R}_+ \to \mathbb{R}$ is strictly increasing. This representation describes an agent who believes that in period 2, she may wish either to act unselfishly as guided by the function u^a (which will happen with probability α) or to select the most selfish option in a purely egoistic manner. The corresponding preference relation satisfies the properties (A1)-(A7). Yet, this representation rules out the patterns that we accommodate by (A7*) (if u^a is increasing as we assumed).

It should also be noted that Kreps' representation requires the submodularity axiom, which posits that $A \cup C \sim A \cup B \cup C$ whenever $A \sim A \cup B$. By contrast, in both our representations, we may well have $A \cup C \prec A \cup B \cup C$ and $A \sim A \cup B$, as the most selfish options (i.e., reference points) in $A \cup B$ and $A \cup B \cup C$ may be different. For example, according to our warm-glow representation, we may have $\{x, y\} \sim \{y\}$ for some x, y with $y_1 > x_1$, but the agent may strictly prefer a menu of the form $\{x, y, z\}$ to $\{y, z\}$ whenever $z_1 > y_1$, as the stronger warm-glow experience associated with the sacrifice $z_1 - x_1$ may convince the agent to select x uniquely from $\{x, y, z\}$.

8 Concluding Remarks

This paper provides a foundation for the notion of warm-glow giving by relating it to the concept of preference for freedom to be selfish. The key insight is that warm-glow necessitates free will: the agent values the availability of selfish options even if she plans to behave prosocially. We provide representation results that are consistent with the evidence on motivation crowding out and, in particular, on the adverse effects of restricting one's freedom to behave selfishly. Compared to Theorem 1, the warm-glow representation in Theorem 2 is more successful in accommodating the evidence on adverse effects of rewarding prosocial behavior. In fact, the second-stage behavior implied by this basic warm-glow representation subsumes that modeled by Riker and Ordeshook (1968) and Andreoni (1989, 1990).

According to this basic warm-glow representation, the agent experiences warm-glow only when her actions (might) help others. However, the strength of the warm-glow experience (i.e., the warm-glow payoff) is solely a function of the agent's private cost, irrespective of how strongly the other agents are influenced. In Appendix A, we provide an extension of this representation in which the warm-glow payoff is an increasing function of the agent's contribution to every other individual's payoff, as well as of her private cost.

If one views "giving" as an act of free will, as opposed to a compulsory transfer of resources, our model can simply be seen as a theory of "preference for giving." On the other hand, social pressure or negative feelings such as shame and guilt may also motivate other-serving actions, even if the decision maker in question dislikes such modes of behavior. As we noted in the Introduction, earlier models of such negative forms of prosocial behavior imply preferences for smaller sets and entail distinct behavior in the second stage.

Appendix A. Multi-Dimensional Warm-Glow Representation

For our multi-dimensional warm-glow representation, we need to replace (FS) and (CM*) with the following axiom.

A9: Other-Regarding Warm-Glow (OWG). Let $x, x', y, y' \in X$ be such that $y_1 \geq x_1$, and suppose that $\{x, y\}$ and $\{x', y'\}$ both belong to $\mathcal{K}_{\mathcal{P}}$. Then, $\{x\} \prec \{x, y\} \succ \{y\}$, $x' \geq x$, $y'_1 - x'_1 \geq y_1 - x_1$ and $x'_{-1} - y'_{-1} \geq x_{-1} - y_{-1}$ imply $\{x', y'\} \succsim \{x, y\}$.

By letting $x = x'$, we see that the following property is a particular consequence of (OWG): For $\{x, y\}, \{x, y'\} \in \mathcal{K}_{\mathcal{P}}$,

$$\{x\} \prec \{x, y\} \succ \{y\}\,,\ y'_1 \geq y_1 \text{ and } y_{-1} \geq y'_{-1} \quad \text{imply} \quad \{x, y'\} \succsim \{x, y\}\,. \qquad (*)$$

This property is a multi-dimensional version of (FS) that relates the warm-glow payoff associated with an action to the social consequences of that action. Intuitively, $(*)$ implies that selecting x over y' would "surely" make the agent better off than selecting x over y only if the former act involves a bigger gift to others as well as a bigger private sacrifice. If we take for granted that the agent may enjoy the act of giving, $(*)$ appears to be a useful generalization of (FS). On the other hand, when our model is viewed as a theory of preference for freedom of choice, the merit of the added generality becomes questionable. For example, if $y_1 = y'_1$ and $y_{-1} > y'_{-1}$, from $\{x, y\}$ to $\{x, y'\}$ the implied change can simply be seen as an increase in the freedom to hurt others. It seems to us that preference for this form of freedom would not be so compelling in a model without negatively interdependent preferences.

It is also important to note that (OWG) implies (CM*), as can be seen by letting $y'_{-1} = y_{-1}$ in the former axiom. Next, we show why (OWG) holds in our basic model.

Claim 1. *If* $\succsim$ *satisfies* (FS), (CM*), (SiM), (SM) *and* (WO) *(over* $\mathcal{A} := \mathcal{K}_{\mathcal{P}}$*), then it also satisfies* (OWG).[30]

Proof. Let $\{x, y\}$ and $\{x', y'\}$ belong to $\mathcal{K}_{\mathcal{P}}$ and suppose that $\{x\} \prec \{x, y\} \succ \{y\}$ while $y_1 \geq x_1$. Notice that $\{x, (y_1, \mathbf{0})\}$ and $\{x', (y'_1, \mathbf{0})\}$ also belong to $\mathcal{K}_{\mathcal{P}}$, where $\mathbf{0} := (0, ..., 0) \in \mathbb{R}^{\ell-1}$. Moreover, by (FS), we must have $\{x, (y_1, \mathbf{0})\} \succsim \{x, y\}$. In turn, this implies that $\{x\} \prec \{x, (y_1, \mathbf{0})\} \succ \{(y_1, \mathbf{0})\}$ because $\succsim$ is transitive and $\{y\} \succsim \{(y_1, \mathbf{0})\}$ by (SiM). Hence, applying (CM*) yields $\{x', (y'_1, \mathbf{0})\} \succsim \{x, (y_1, \mathbf{0})\}$ for $x' \geq x$ and $y'_1 - x'_1 \geq y_1 - x_1$. Finally, note that $\{x', y'\} \succsim \{x', (y'_1, \mathbf{0})\}$. Indeed, if $\{x', (y'_1, \mathbf{0})\} \sim \{x'\}$ or $\{x', (y'_1, \mathbf{0})\} \sim \{(y'_1, \mathbf{0})\}$ that $\{x', y'\} \succsim \{x', (y'_1, \mathbf{0})\}$ follows from (SM), (SiM) and transitivity in an obvious way. In turn, if $\{x'\} \prec \{x', (y'_1, \mathbf{0})\} \succ \{(y'_1, \mathbf{0})\}$ we can apply (FS) to reach the same conclusion. $\square$

Definition 3. A binary relation $\succsim$ on $\mathcal{K}_{\mathcal{P}}$ admits a **multi-dimensional warm-glow representation** if there exists a weakly increasing and continuous function

[30] We have not been able to determine if the conclusion of this claim remains true upon replacing (FS) with property $(*)$. However, the answer seems to be negative because in the statement of (OWG), we do not necessarily have $y_{-1} \geq y'_{-1}$.

$U_{\mathrm{mw}} : X \times \mathbb{R}^{\ell}_{+} \to \mathbb{R}$ such that, for any $A \in \mathcal{K}_{\mathcal{P}}$, the function

$$V_{\mathrm{mw}}(A) := \max_{x \in A} U_{\mathrm{mw}}(x, \max_{y \in A} y_1 - x_1, x_2 - \min_{y \in A} y_2, x_3 - \min_{y \in A} y_3, ..., x_\ell - \min_{y \in A} y_\ell)$$

represents $\succsim$. We refer to such a function U_{mw} as an **mw-index** for $\succsim$.

Theorem 3. *A binary relation* $\succsim$ *on* $\mathcal{A} := \mathcal{K}_{\mathcal{P}}$ *satisfies the axioms* (A1)-(A3), (A5), (A6), (A8) *and* (A9) *if and only if it admits a multi-dimensional warm-glow representation.*

Appendix B. Proofs

B1. Proof of Theorem 1

As it is straightforward, we omit the "if" part of the proof of Theorem 1. To prove the "only if" part, let $\succsim$ be a binary relation on $\mathcal{A} := \mathcal{K}$ that satisfies (A1)-(A7).

For each $A \in \mathcal{K}$, pick a point $y^*(A) \in \mathcal{K}$ such that $y^*_1(A) = \max\{x_1 : x \in A\}$. Define $\mathbf{A} := \{\{x, y^*(A)\} : x \in A\}$, and note that $\mathbf{A}$ (equipped with the Hausdorff metric) is homeomorphic to A, and compact in particular. Thus, $\succsim$ admits a maximal set in $\mathbf{A}$ by (C). That is, there exists an allocation $\overline{x}(A)$ in A such that $\{\overline{x}(A), y^*(A)\} \succsim \{x, y^*(A)\}$ for every $x \in A$. The following claim proves a related observation that we mentioned earlier.

Claim 2. *For any* $A \in \mathcal{K}$*, we have* $A \sim \{\overline{x}(A), y^*(A)\}$.

Proof. Fix a set $A \in \mathcal{K}$, and let $\{x^1, ..., x^n, ...\}$ be a countable, dense subset of A. For every $n \in \mathbb{N}$, put $A^n := \{x^1, ..., x^n\} \cup \{\overline{x}(A), y^*(A)\}$. Then, (SM) implies $A^1 \succsim \{\overline{x}(A), y^*(A)\}$. Moreover, by (WI), we have $A^1 \sim \{x^1, y^*(A)\}$ or $A^1 \sim \{\overline{x}(A), y^*(A)\}$. As $\{x^1, y^*(A)\} \precsim \{\overline{x}(A), y^*(A)\}$, either equivalence implies $A^1 \precsim \{\overline{x}(A), y^*(A)\}$, that is, $A^1 \sim \{\overline{x}(A), y^*(A)\}$. Similarly, either $A^2 \sim \{x^2, y^*(A)\}$ or $A^2 \sim A^1$, and in both cases, we have $A^2 \sim \{\overline{x}(A), y^*(A)\}$. Inductively, it follows that $A^n \sim \{\overline{x}(A), y^*(A)\}$ for every n. Moreover, since the sequence $A^1, A^2, ...$ is uniformly bounded in Euclidean norm and increases with respect to set inclusion, it is well known that $A^n \to \mathrm{cl}(\bigcup_{n=1}^{\infty} A^n)$ in Hausdorff metric (see, e.g., Dekel et al., 2001, Lemma 5). In turn, $\mathrm{cl}(\bigcup_{n=1}^{\infty} A^n)$ equals A by construction. Hence, (C) implies $A \sim \{\overline{x}(A), y^*(A)\}$, as we sought. $\square$

For future use, note that in the above proof we have not utilized (FS) or (CM). Next, we prove another general claim that does not require these two axioms.

Claim 3. *Let $x, x', y, y' \in X$ be such that $x' \geq x$ and $y' \geq y$. Then, $\{x', y'\} \succsim \{x, y\}$ unless $\{x\} \prec \{x, y\} \succ \{y\}$.*

Proof. Suppose that $\{x\} \prec \{x, y\} \succ \{y\}$ does not hold. By (SM), this means that either $\{x, y\} \sim \{x\}$ or $\{x, y\} \sim \{y\}$. Moreover, (SiM) and (SM) imply $\{x\} \precsim \{x'\} \precsim \{x', y'\}$ and $\{y\} \precsim \{y'\} \precsim \{x', y'\}$. As $\succsim$ is transitive, we conclude that $\{x, y\} \precsim \{x', y'\}$. $\square$

In the following claim, we utilize (FS) and (CM) to strengthen the conclusion of Claim 3.

Claim 4. *Let $x, x', y, y' \in X$ be such that $x' \geq x$ and $y' \geq y$. Then, $\{x', y'\} \succsim \{x, y\}$.*

Proof. In view of Claim 3, without loss of generality we can assume $\{x\} \prec \{x, y\} \succ \{y\}$. By relabeling if necessary, let us also assume that $y_1 \geq x_1$. Then, (FS) implies $\{x, y'\} \succsim \{x, y\}$ since $y'_1 \geq y_1$. It remains to show that $\{x', y'\} \succsim \{x, y'\}$. To this end, by applying Claim 3 again, we can assume $\{x\} \prec \{x, y'\} \succ \{y'\}$. But then, as $y'_1 \geq x_1$ and $x' \geq x$, the desired conclusion follows from (CM): $\{x', y'\} \succsim \{x, y'\}$. $\square$

As is well-known, when endowed with the Hausdorff metric, the space of all nonempty, compact subsets of $\mathbb{R}^\ell$ is separable. Then, as a subspace, $\mathcal{K}$ is also separable. Hence, Debreu's classical theorem implies that there exists a continuous function $V : \mathcal{K} \to \mathbb{R}$ such that $A \succsim B$ iff $V(A) \geq V(B)$, for every $A, B \in \mathcal{K}$. The following claim characterizes the main feature of f-indices compatible with V. (Throughout the remainder of the proof, $\mathbf{0}$ denotes the origin of $\mathbb{R}^{\ell-1}$.)

Claim 5. *Let $U : X \times \mathbb{R}_+ \to \mathbb{R}$ be a weakly increasing function. Then, properties* (ii)-(iv) *below hold simultaneously iff property* (i) *holds.*

(i) $\max_{x \in A} U(x, y_1^*(A)) = V(A)$ *for all* $A \in \mathcal{K}$.

(ii) $U(x, x_1) = V(\{x\})$ *for all* $x \in X$.

(iii) $U(x, \lambda) \leq V(\{x, (\lambda, \mathbf{0})\})$ *for all* $(x, \lambda) \in X \times \mathbb{R}_+$ *with* $\lambda \geq x_1$.

(iv) $U(x, \lambda) = V(\{x, (\lambda, \mathbf{0})\})$ *for all* $(x, \lambda) \in X \times \mathbb{R}_+$ *with* $\lambda \geq x_1$ *and* $\{x\} \prec \{x, (\lambda, \mathbf{0})\} \succ \{(\lambda, \mathbf{0})\}$.

Proof. First, suppose that (i) holds. Then, by letting $A := \{x\}$, we immediately see that (ii) must also hold. Now, take any $(x, \lambda) \in X \times \mathbb{R}_+$ with $\lambda \geq x_1$, so that $\lambda = y_1^*(A')$ where $A' := \{x, (\lambda, \mathbf{0})\}$. Then, by (i) and (ii), $V(A') = \max\{U(x, \lambda), U((\lambda, \mathbf{0}), \lambda)\}$, implying that $V(A') \geq U(x, \lambda)$. This verifies (iii).

Moreover, $A' \succ \{(\lambda, \mathbf{0})\}$ simply means $V(A') = \max\{U(x,\lambda), U((\lambda,\mathbf{0}),\lambda)\} > U((\lambda,\mathbf{0}),\lambda)$. In turn, the latter statement is equivalent to $V(A') = U(x,\lambda) > U((\lambda,\mathbf{0}),\lambda)$. In particular, $\{x\} \prec A' \succ \{(\lambda,\mathbf{0})\}$ implies $V(A') = U(x,\lambda)$, which verifies (iv).

Conversely, suppose now that (ii)-(iv) hold, and take any $A \in \mathcal{K}$. Let us write $\overline{x}$ instead of $\overline{x}(A)$, and y^* instead of $y^*(A)$. Then, from Claim 2 and the definition of $\overline{x}$, it follows that

$$V(A) = V(\{\overline{x}, y^*\}) \geq V(\{x, y^*\}) \quad \text{for } x \in A. \tag{6}$$

Moreover, Claim 4 implies

$$V(\{x, y^*\}) \geq V(\{x, (y_1^*, \mathbf{0})\}) \quad \text{for } x \in A. \tag{7}$$

In turn, by property (iii), we also have

$$V(\{x, (y_1^*, \mathbf{0})\}) \geq U(x, y_1^*) \quad \text{for } x \in A. \tag{8}$$

By combining (6)-(8), we see that $V(A) \geq \sup_{x \in A} U(x, y_1^*)$.

To prove the converse inequality, obviously, it suffices to show that

$$V(\{\overline{x}, y^*\}) \leq \max_{x \in \{\overline{x}, y^*\}} U(x, y_1^*). \tag{9}$$

Note that $\max\{V(\{\overline{x}\}), V(\{y^*\})\} = \max\{U(\overline{x}, \overline{x}_1), U(y^*, y_1^*)\} \leq \max_{x \in \{\overline{x}, y^*\}} U(x, y_1^*)$ by property (ii) and weak monotonicity of U. Therefore, (9) trivially holds if $\max\{V(\{\overline{x}\}), V(\{y^*\})\} = V(\{\overline{x}, y^*\})$. Thus, assume $\max\{V(\{\overline{x}\}), V(\{y^*\})\} < V(\{\overline{x}, y^*\})$. Then, (FS) implies $V(\{\overline{x}, y^*\}) \leq V(\{\overline{x}, (y_1^*, \mathbf{0})\})$. Since $V(\{(y_1^*, \mathbf{0})\}) \leq V(\{y^*\})$, it also follows that $\max\{V(\{\overline{x}\}), V(\{(y_1^*, \mathbf{0})\})\} < V(\{\overline{x}, (y_1^*, \mathbf{0})\})$. By (iv), we must then have $U(\overline{x}, y_1^*) = V(\{\overline{x}, (y_1^*, \mathbf{0})\})$, implying that $V(\{\overline{x}, y^*\}) \leq U(\overline{x}, y_1^*)$. This proves (9).

It follows that $V(A) = \sup_{x \in A} U(x, y_1^*) = \max_{x \in \{\overline{x}, y^*\}} U(x, y_1^*)$. Finally, the latter equality implies that the function $x \to U(x, y_1^*)$ attains its maximum over A (either at $\overline{x}$ or y^*). This proves the property (i). $\square$

We complete the proof of Theorem 1 with the following claim.

Claim 6. *Define a function $U : X \times \mathbb{R}_+ \to \mathbb{R}$ as $U(x,\lambda) := V(\{x, (\lambda, \mathbf{0})\})$ for all $(x,\lambda) \in X \times \mathbb{R}_+$. The function U is an f-index for $\succsim$.*

Proof. In view of Claim 5, it suffices to show that U is a weakly increasing, continuous function that satisfies the properties (ii)-(iv) in Claim 5. That U is weakly increasing is an obvious consequence of Claim 4 and the definition of V. Moreover, since V and the function $(x,\lambda) \to \{x,(\lambda,\mathbf{0})\}$ are both continuous, so is U. To verify property (ii) in Claim 5, fix an $x \in X$. Note that $\{x,(x_1,\mathbf{0})\} \succsim \{x\}$ by (SM), and $\{x\} = \{x,x\} \succsim \{x,(x_1,\mathbf{0})\}$ by Claim (4). Thus, $\{x,(x_1,\mathbf{0})\} \sim \{x\}$, implying that $U(x,x_1) := V(\{x,(x_1,\mathbf{0})\}) = V(\{x\})$, as we seek. Finally, note that U trivially satisfies the properties (iii) and (iv) in Claim 5. $\square$

Since Theorem 3 is more general than Theorem 2, we shall deduce the latter result from the former. We proceed to:

B2. Proof Theorem 3

We omit the "if" part of the proof which is a routine exercise. Let $\succsim$ be a binary relation on $\mathcal{K}_{\mathcal{P}}$ that satisfies (WO), (SM), (WI), (SiM), (C), (EC) and (OWG).

As in the text, for each $A \in \mathcal{K}_{\mathcal{P}}$ let $y^*(A)$ denote the unique element of A such that $y_1^*(A) = \max\{x_1 : x \in A\}$ and $y_i^*(A) = \min\{x_i : x \in A\}$ for $i = 2,...,\ell$. Notice that if a set A belongs to $\mathcal{K}_{\mathcal{P}}$, any compact subset of A that contains $y^*(A)$ also belongs to $\mathcal{K}_{\mathcal{P}}$. Hence, we can repeat the arguments in the proof of Claim 2 to show that for each $A \in \mathcal{K}_{\mathcal{P}}$, there exists a point $\overline{x}(A) \in A$ such that

$$A \sim \{\overline{x}(A), y^*(A)\} \succsim \{x, y^*(A)\} \quad \text{for every } x \in A. \tag{10}$$

Let us now introduce a bit of notation. Throughout the remainder of the appendix, λ stands for a vector of the form $(\lambda_1,...,\lambda_\ell) \in \mathbb{R}_+^\ell$. In turn, $\mathbf{0}$ denotes the origin of $\mathbb{R}^\ell$ or $\mathbb{R}^{\ell-1}$, depending on the context. We define $D := \{(x,\lambda) \in X \times \mathbb{R}_+^\ell : x_{-1} \geq \lambda_{-1}\}$, and $x \diamond \lambda := (x_1+\lambda_1, x_{-1}-\lambda_{-1})$ for every $(x,\lambda) \in D$. Notice that for any $(x,\lambda) \in D$, the point $x \diamond \lambda$ belongs to X, while the set $\{x, x\diamond\lambda\}$ belongs to $\mathcal{K}_{\mathcal{P}}$ iff (i) $\lambda = \mathbf{0}$, or (ii) $\lambda_1 > 0$ and $\lambda_{-1} > \mathbf{0}$. For any two vectors $a = (a_1,...,a_k)$, $b = (b_1,...,b_k)$, we set $a \vee b := (\max\{a_1,b_1\},...,\max\{a_k,b_k\})$ and $a \wedge b := (\min\{a_1,b_1\},...,\min\{a_k,b_k\})$. It should be noted that since $x_{-1} \geq \lambda_{-1}$ and $x'_{-1} \geq \lambda'_{-1}$ imply $x_{-1} \vee x'_{-1} \geq \lambda_{-1} \vee \lambda'_{-1}$ and $x_{-1} \wedge x'_{-1} \geq \lambda_{-1} \wedge \lambda'_{-1}$, the set D is a lattice: $(x \vee x', \lambda \vee \lambda')$ and $(x \wedge x', \lambda \wedge \lambda')$ belong to D for $(x,\lambda),(x',\lambda') \in D$.

It is clear that, as in the proof of Theorem 2, there exists a continuous function $V : \mathcal{K}_{\mathcal{P}} \to \mathbb{R}$ such that $A \succsim B$ iff $V(A) \geq V(B)$, for every $A, B \in \mathcal{K}_{\mathcal{P}}$. The following claim is an analogue of Claim 5 for the present setup.

Claim 7. *Let $U : D \to \mathbb{R}$ be a weakly increasing function. Then, properties* (ii)-(iv) *below hold simultaneously iff property* (i) *holds.*

(i) $\max_{x \in A} U\left(x, y_1^*(A) - x_1, x_{-1} - y_{-1}^*(A)\right) = V(A)$ *for all* $A \in \mathcal{K}_{\mathcal{P}}$*.*

(ii) $U(x, \mathbf{0}) = V(\{x\})$ *for all* $x \in X$*.*

(iii) $U(x, \lambda) \leq V(\{x, x \diamond \lambda\})$ *for all* $(x, \lambda) \in D$ *such that* $\{x, x \diamond \lambda\} \in \mathcal{K}_{\mathcal{P}}$*.*

(iv) $U(x, \lambda) = V(\{x, x \diamond \lambda\})$ *for all* $(x, \lambda) \in D$ *such that* $\{x, x \diamond \lambda\} \in \mathcal{K}_{\mathcal{P}}$ *and* $\{x\} \prec \{x, x \diamond \lambda\} \succ \{x \diamond \lambda\}$*.*

Proof. As in the proof of Claim 5, (ii) and (iii)-(iv) easily follow from (i) upon letting $A := \{x\}$ and $A' := \{x, x \diamond \lambda\}$, respectively.

To prove the converse implication, suppose that (ii)-(iv) hold. Take any $A \in \mathcal{K}_{\mathcal{P}}$, and set $\overline{x} := \overline{x}(A)$, $y^* := y^*(A)$. The expression (10) simply means that

$$V(A) = V(\{\overline{x}, y^*\}) = \max_{x \in A} V(\{x, y^*\}). \tag{11}$$

Moreover, for each $x \in A$ if we set $\lambda_x := (y_1^* - x_1, x_{-1} - y_{-1}^*)$, we see that $(x, \lambda_x) \in D$ and $\{x, x \diamond \lambda_x\} = \{x, y^*\} \in \mathcal{K}_{\mathcal{P}}$. Thus, property (iii) implies $V(\{x, y^*\}) \geq U(x, \lambda_x)$ for every $x \in A$. From (11), it then follows that $V(A) \geq \sup_{x \in A} U(x, \lambda_x)$, as we seek.

As in Claim 5, we will complete the proof by showing that

$$V(\{\overline{x}, y^*\}) \leq \max_{x \in \{\overline{x}, y^*\}} U(x, \lambda_x). \tag{12}$$

Since $\max\{V(\{\overline{x}\}), V(\{y^*\})\} = \max\{U(\overline{x}, \mathbf{0}), U(y^*, \mathbf{0})\} \leq \max_{x \in \{\overline{x}, y^*\}} U(x, \lambda_x)$, without loss of generality we can assume $\max\{V(\{\overline{x}\}), V(\{y^*\})\} < V(\{\overline{x}, y^*\})$. But then, as $\overline{x} \diamond \lambda_{\overline{x}} = y^*$, property (iv) implies $U(\overline{x}, \lambda_{\overline{x}}) = V(\{\overline{x}, \overline{x} \diamond \lambda_{\overline{x}}\}) = V(\{\overline{x}, y^*\})$, which proves (12). $\square$

In what follows, we will define a function U on the set D and then extend it to $X \times \mathbb{R}_+^\ell$. Of course, the construction of U on D will build upon Claim 7, but first we need to extend the function V to include the doubletons in the closure of $\mathcal{K}_{\mathcal{P}}$. Set

$$\overline{\mathcal{K}}_{\mathcal{P}_2} := \text{cl}\{\{x, y\} : x, y \in X, \ \{x, y\} \in \mathcal{K}_{\mathcal{P}}\}.$$

We proceed with some technical observations.

Claim 8. *$\mathcal{K}_{\mathcal{P}}$ is a connected set.*

Proof. Let us show that $\mathcal{K}_{\mathcal{P}}$ is path-connected. Fix a pair of sets $A, B \in \mathcal{K}_{\mathcal{P}}$ and a point $x \in X$. Note that for any $C \in \mathcal{K}_{\mathcal{P}}$ and $\alpha \in [0, 1]$, the set $\alpha C + (1 - \alpha)\{x\} :=$

$\{\alpha z + (1-\alpha)x : z \in C\}$ belongs to $\mathcal{K}_{\mathcal{P}}$. For each $t \in [0,1]$, put

$$f(t) := \begin{cases} (1-2t)A + 2t\{x\} & \text{if } 0 \le t \le 1/2, \\ (2-2t)\{x\} + (2t-1)B & \text{if } 1/2 < t \le 1. \end{cases}$$

As we just noted, the function f maps $[0,1]$ into $\mathcal{K}_{\mathcal{P}}$. It is also clear that f is a continuous mapping, for $\lim_{t\to 0.5^+} f(t) = \lim_{t\to 0.5^-} f(t) = \{x\}$. Finally, note that $f(0) = A$ and $f(1) = B$. This completes the proof. $\square$

Claim 9. *Let $\{x,y\} \in \overline{\mathcal{K}}_{\mathcal{P}_2}$, and take a sequence $\{x^n, y^n\}$ in $\mathcal{K}_{\mathcal{P}}$ that converges to $\{x,y\}$. Then:*

(i) *$V(\{x^n, y^n\})$ converges to a finite number.*

(ii) *$\lim V(\{x^n, y^n\}) = \lim V(\{x'^n, y'^n\})$ for any other sequence $\{x'^n, y'^n\}$ in $\mathcal{K}_{\mathcal{P}}$ that converges to $\{x,y\}$.*

Proof. By relabeling if necessary, assume $y_1^n > x_1^n$ and $x_{-1}^n > y_{-1}^n$ for every n. Note that by definition of the Hausdorff metric, $\lim x^n$ and $\lim y^n$ exist, and we have $\{x,y\} = \{\lim x^n, \lim y^n\}$.

We shall now show that the sequence $V(\{x^n, y^n\})$ is bounded. Since they are convergent, the sequences x^n and y^n are bounded. Let $x' \in \mathbb{R}^{\ell}_{++}$ be such that $x' \ge x^n \vee y^n$ for every n. Clearly, the set $\{2x', (3x'_1, x'_{-1})\}$ belongs to $\mathcal{K}_{\mathcal{P}}$. Moreover,

$$\{2x', (3x'_1, x'_{-1})\} \succsim \{x^n, y^n\} \quad \text{for every } n. \tag{13}$$

Indeed, for any n, if $\{x^n, y^n\} \sim \{x^n\}$ or $\{x^n, y^n\} \sim \{y^n\}$, (13) follows from (SiM), (SM) and (WO) as in Claim 3. On the other hand, if $\{x^n\} \prec \{x^n, y^n\} \succ \{y^n\}$, then (OWG) implies (13), for $2x' \ge x^n$, $3x'_1 - 2x'_1 \ge y_1^n - x_1^n$ and $2x'_{-1} - x'_{-1} \ge x_{-1}^n - y_{-1}^n$. In turn, (13) simply means that $V(\{2x', (3x'_1, x'_{-1})\}) \ge V(\{x^n, y^n\})$ for every n.

Since $V(\{x^n, y^n\})$ is bounded, the proof of (i) will be complete if we can show that for any two convergent subsequences $V(\{x^{n_k}, y^{n_k}\})$ and $V(\{x^{n_l}, y^{n_l}\})$, we have

$$\lim_k V(\{x^{n_k}, y^{n_k}\}) = \lim_l V(\{x^{n_l}, y^{n_l}\}). \tag{14}$$

Assume by contradiction that $\lim_k V(\{x^{n_k}, y^{n_k}\}) > \lim_l V(\{x^{n_l}, y^{n_l}\})$ for a pair of convergent subsequences $V(\{x^{n_k}, y^{n_k}\})$, $V(\{x^{n_l}, y^{n_l}\})$. Pick two numbers $\varepsilon_1, \varepsilon_2$ such that $\lim_k V(\{x^{n_k}, y^{n_k}\}) > \varepsilon_1 > \varepsilon_2 > \lim_l V(\{x^{n_l}, y^{n_l}\})$. Then,

$$V(\{x^{n_k}, y^{n_k}\}) > \varepsilon_1 > \varepsilon_2 > V(\{x^{n_l}, y^{n_l}\}) \quad \text{for all sufficiently large } k \text{ and } l. \tag{15}$$

Since V is a continuous function on the connected set $\mathcal{K}_{\mathcal{P}}$, the set $V(\mathcal{K}_{\mathcal{P}})$ is an interval. Thus, (15) implies that there exists a pair of sets $A, B \in \mathcal{K}_{\mathcal{P}}$ such that $V(A) = \varepsilon_1$ and $V(B) = \varepsilon_2$. Moreover, we obviously have $\lim_k\{x^{n_k}, y^{n_k}\} = \{x, y\} = \lim_l\{x^{n_l}, y^{n_l}\}$. Then, in view of (15), it follows that any neighborhood $\mathcal{N}$ of $\{x, y\}$ contains a pair of sets $\{x^{n_k}, y^{n_k}\}, \{x^{n_l}, y^{n_l}\}$ such that $\{x^{n_k}, y^{n_k}\} \succ A \succ B \succ \{x^{n_l}, y^{n_l}\}$. Since this contradicts (EC), we conclude that part (i) of the claim holds.

To prove part (ii), take any other sequence $\{x'^n, y'^n\}$ in $\mathcal{K}_{\mathcal{P}}$ that converges to $\{x, y\}$. By part (i) of the claim, $\lim V(\{x'^n, y'^n\})$ exists. Suppose now that this limit is distinct from $\lim V(\{x^n, y^n\})$, say $\lim V(\{x^n, y^n\}) > \lim V(\{x'^n, y'^n\})$. Then, just as in the proof of (14), we can find two sets $A, B \in \mathcal{K}_{\mathcal{P}}$ such that $\{x^n, y^n\} \succ A \succ B \succ \{x'^n, y'^n\}$ for all sufficiently large n. In turn, this contradicts (EC) as we have $\lim\{x^n, y^n\} = \{x, y\} = \lim\{x'^n, y'^n\}$. This completes the proof. $\square$

For each $\{x, y\} \in \overline{\mathcal{K}}_{\mathcal{P}_2}$, set

$$\overline{V}(\{x, y\}) := \lim V(\{x^n, y^n\})$$

for every sequence $\{x^n, y^n\}$ in $\mathcal{K}_{\mathcal{P}}$ that converges to $\{x, y\}$. In view of Claim 9, $\overline{V}(\cdot)$ is a well-defined function on $\overline{\mathcal{K}}_{\mathcal{P}_2}$. Moreover, for any $\{x, y\} \in \mathcal{K}_{\mathcal{P}}$, we have $\overline{V}(\{x, y\}) = V(\{x, y\})$ since a constant sequence that equals $\{x, y\}$ trivially converges to $\{x, y\}$. In particular, $\overline{V}(\{x\}) = V(\{x\})$ for every $x \in X$.

It should also be noted that

$$\{x, x \diamond \lambda\} \in \overline{\mathcal{K}}_{\mathcal{P}_2} \text{ for any } (x, \lambda) \in D. \tag{16}$$

To see this point, pick any $(x, \lambda) \in D$, and a strictly positive, $\ell - 1$ dimensional vector e. For each $n \in \mathbb{N}$, set $x^n := (x_1, x_{-1} + \frac{1}{n}e)$ and $y^n := (x_1 + \lambda_1 + \frac{1}{n}, x_{-1} - \lambda_{-1})$. Then, clearly, $\{x^n, y^n\}$ is a sequence in $\mathcal{K}_{\mathcal{P}}$ that converges to $\{x, x \diamond \lambda\}$, as we seek.

In view of (16), for any $(x, \lambda) \in D$ the set $\{x, x \diamond \lambda\}$ belongs to the domain of the function $\overline{V}$. Notice, however, that $\overline{V}(\{x, x \diamond \lambda\})$ may not be a weakly increasing function of λ because we may have $\overline{V}(\{x, x \diamond \lambda\}) = \overline{V}(\{x \diamond \lambda\})$ on occasion, while $\overline{V}(\{x \diamond \lambda\})$ is weakly decreasing with λ_{-1}. In the next claim, we will establish some useful facts about the function $\overline{V}$. In particular, we will uncover some monotonicity properties of $\overline{V}(\{x, x \diamond \lambda\})$ for suitably selected $(x, \lambda) \in D$.

Claim 10. **(i)** *$\overline{V}$ is continuous on $\overline{\mathcal{K}}_{\mathcal{P}_2}$.*

(ii) $\overline{V}(\{x,y\}) \geq \overline{V}(\{x\})$ *for any* $\{x,y\} \in \overline{\mathcal{K}}_{\mathcal{P}_2}$.

(iii) *Take any* $(x,\lambda),(x',\lambda') \in D$ *with* $(x',\lambda') \geq (x,\lambda)$. *Assume further that* $\{x, x \diamond \lambda\}$ *belongs to the closure of the set* $\{\{z,y\} \in \mathcal{K}_{\mathcal{P}} : \{z\} \prec \{z,y\} \succ \{y\}\}$. *Then,* $\overline{V}(\{x', x' \diamond \lambda'\}) \geq \overline{V}(\{x, x \diamond \lambda\})$.

(iv) *Take any* $(x,\lambda),(x',\lambda') \in D$ *with* $(x',\lambda') \geq (x,\lambda)$. *Assume further that* $x' \diamond \lambda' \geq x \diamond \lambda$. *Then,* $\overline{V}(\{x', x' \diamond \lambda'\}) \geq \overline{V}(\{x, x \diamond \lambda\})$.

Proof. We start with the proof of (i). Let $\{x^n, y^n\}$ be a sequence in $\overline{\mathcal{K}}_{\mathcal{P}_2}$ that converges to $\{x,y\}$. By definition of $\overline{V}$, for each n, there exists a set $\{x'^n, y'^n\} \in \mathcal{K}_{\mathcal{P}}$ such that

$$d_H(\{x'^n, y'^n\}, \{x^n, y^n\}) < 1/n \quad \text{and} \quad \left|V(\{x'^n, y'^n\}) - \overline{V}(\{x^n, y^n\})\right| < 1/n. \tag{17}$$

Since $\lim\{x^n, y^n\} = \{x,y\}$, the former expression in (17) implies $\lim\{x'^n, y'^n\} = \{x,y\}$. Thus, $\overline{V}(\{x,y\}) := \lim V(\{x'^n, y'^n\})$. Hence, from the latter expression in (17), it follows that $\lim \overline{V}(\{x^n, y^n\})$ exists, and we have $\lim \overline{V}(\{x^n, y^n\}) = \overline{V}(\{x,y\})$. This proves (i).

Now, take any $\{x,y\} \in \overline{\mathcal{K}}_{\mathcal{P}_2}$, and let $\{x^n, y^n\}$ be a sequence in $\mathcal{K}_{\mathcal{P}}$ that converges to $\{x,y\}$ so that $\{x,y\} = \{\lim x^n, \lim y^n\}$. Without loss of generality, assume $x = \lim x^n$. Then, $\overline{V}(\{x,y\}) := \lim V(\{x^n, y^n\}) \geq \lim V(\{x^n\}) := \overline{V}(\{x\})$, where the weak inequality follows from (SM). This proves (ii).

In the remainder of the proof, let (x,λ) and (x',λ') be points in D such that $(x',\lambda') \geq (x,\lambda)$.

To prove (iii), suppose that

$$\{x, x \diamond \lambda\} = \lim\{x^n, y^n\} \tag{18}$$

for a sequence $\{x^n, y^n\}$ in $\mathcal{K}_{\mathcal{P}}$ such that $\{x^n\} \prec \{x^n, y^n\} \succ \{y^n\}$ for every n. By relabeling if necessary, assume $y^n_1 > x^n_1$ and $x^n_{-1} > y^n_{-1}$ for every n. Since $x_1 = \min\{z_1 : z \in \{x, x \diamond \lambda\}\}$ and $x^n_1 = \min\{z_1 : z \in \{x^n, y^n\}\}$ for each n, from (18) it obviously follows that $x_1 = \lim \min\{z_1 : z \in \{x^n, y^n\}\} = \lim x^n_1$. Similarly, $(x \diamond \lambda)_1 = \lim \max\{z_1 : z \in \{x^n, y^n\}\} = \lim y^n_1$, while for $i = 2, ..., \ell$, we have $x_i = \lim \max\{z_i : z \in \{x^n, y^n\}\} = \lim x^n_i$ and $(x \diamond \lambda)_i = \lim \min\{z_i : z \in \{x^n, y^n\}\} = \lim y^n_i$. Hence,

$$x = \lim x^n \quad \text{and} \quad x \diamond \lambda = \lim y^n. \tag{19}$$

Define a sequence λ^n in $\mathbb{R}^\ell_+$ as $\lambda^n_1 := y^n_1 - x^n_1$ and $\lambda^n_{-1} := x^n_{-1} - y^n_{-1}$. Then, (19) implies that $\lim \lambda_1^n = (x \diamond \lambda)_1 - x_1 = \lambda_1$ and $\lim \lambda_{-1}^n = x_{-1} - (x \diamond \lambda)_{-1} = \lambda_{-1}$, so that $\lim \lambda^n = \lambda$. Moreover, we have $x^n \diamond \lambda^n = y^n$, and hence, $\{x^n\} \prec \{x^n, x^n \diamond \lambda^n\} \succ \{x^n \diamond \lambda^n\}$ for every n.

Now, for each n, set $x'^n := x' \vee x^n$ and $\lambda'^n := \lambda' \vee \lambda^n$. Then, $\lambda_1'^n \geq \lambda_1^n > 0$ and $\lambda_{-1}'^n \geq \lambda_{-1}^n > \mathbf{0}$. Furthermore, $(x'^n, \lambda'^n) \in D$ for each n because the set D is a lattice as we noted earlier. Thus, we see that $\{x'^n, x'^n \diamond \lambda'^n\}$ belongs to $\mathcal{K}_\mathcal{P}$ for each n. In turn, by construction, $x'^n \geq x^n$, $(x'^n \diamond \lambda'^n)_1 - x_1'^n \geq (x^n \diamond \lambda^n)_1 - x_1^n$ and $x_{-1}'^n - (x'^n \diamond \lambda'^n)_{-1} \geq x_{-1}^n - (x^n \diamond \lambda^n)_{-1}$. From (OWG), it therefore follows that $V(\{x'^n, x'^n \diamond \lambda'^n\}) \geq V(\{x^n, x^n \diamond \lambda^n\})$ for every n. Finally, note that since $\vee$ is a continuous operator, we have $\lim x'^n = x' \vee x = x'$ and $\lim \lambda'^n = \lambda' \vee \lambda = \lambda'$, implying that $\lim\{x'^n, x'^n \diamond \lambda'^n\} = \{x', x' \diamond \lambda'\}$. Hence, we conclude that $\overline{V}(\{x', x' \diamond \lambda'\}) := \lim V(\{x'^n, x'^n \diamond \lambda'^n\}) \geq \lim V(\{x^n, x^n \diamond \lambda^n\}) := \overline{V}(\{x, x \diamond \lambda\})$. This completes the proof of (iii).

To prove (iv), suppose now that $x' \diamond \lambda' \geq x \diamond \lambda$**.** Note that if $\overline{V}(\{x, x \diamond \lambda\}) = \max\{\overline{V}(\{x\}), \overline{V}(\{x \diamond \lambda\})\}$, (SiM) and part (ii) of the claim obviously imply $\overline{V}(\{x, x \diamond \lambda\}) \leq \overline{V}(\{x', x' \diamond \lambda'\})$. Therefore, without loss of generality, assume

$$\overline{V}(\{x, x \diamond \lambda\}) > \max\{\overline{V}(\{x\}), \overline{V}(\{x \diamond \lambda\})\}. \tag{20}$$

Pick a sequence $\{z^n, y^n\}$ in $\mathcal{K}_\mathcal{P}$ that converges to $\{x, x \diamond \lambda\}$. Then, clearly, the sequence $\max\{V(\{z^n\}), V(\{y^n\})\}$ converges to $\max\{\overline{V}(\{x\}), \overline{V}(\{x \diamond \lambda\})\}$, while $\lim V(\{z^n, y^n\}) := \overline{V}(\{x, x \diamond \lambda\})$. Therefore, (20) implies that $V(\{z^n, y^n\}) > \max\{V(\{z^n\}), V(\{y^n\})\}$ for all sufficiently large n. But then, $\{x, x \diamond \lambda\}$ belongs to $\mathrm{cl}\{\{z, y\} \in \mathcal{K}_\mathcal{P} : \{z\} \prec \{z, y\} \succ \{y\}\}$. Hence, from part (iii) of the claim, it follows that $\overline{V}(\{x', x' \diamond \lambda'\}) \geq \overline{V}(\{x, x \diamond \lambda\})$, as we seek. $\square$

We are now ready to define a function U on D. For every $(x, \lambda) \in D$, set $(x, \lambda)^\uparrow := \{(x', \lambda') \in D : (x', \lambda') \geq (x, \lambda)\}$, and

$$U(x, \lambda) := \inf\{\overline{V}(\{x', x' \diamond \lambda'\}) : (x', \lambda') \in (x, \lambda)^\uparrow\}.$$

Next, we shall show that:

Claim 11. *U is a weakly increasing, continuous function on D that satisfies the properties* (ii)-(iv) *in Claim 7.*

Proof. Note that for any $(x, \lambda) \in D$, we also have $(x, \lambda) \in (x, \lambda)^\uparrow$, and hence, $U(x, \lambda) \leq \overline{V}(\{x, x \diamond \lambda\})$. In turn, $\overline{V}(\{x, x \diamond \lambda\}) = V(\{x, x \diamond \lambda\})$ for $\{x, x \diamond \lambda\} \in \mathcal{K}_\mathcal{P}$, implying that $U(x, \lambda) \leq V(\{x, x \diamond \lambda\})$. This verifies property (iii) in Claim 7. If, in addition, $\{x\} \prec \{x, x \diamond \lambda\} \succ \{x \diamond \lambda\}$, then Claim 10(iii) implies $\overline{V}(\{x', x' \diamond \lambda'\}) \geq V(\{x, x \diamond \lambda\})$ for every $(x', \lambda') \in (x, \lambda)^{\uparrow}$, so that $U(x, \lambda) \geq V(\{x, x \diamond \lambda\})$. Hence, $U(x, \lambda) = V(\{x, x \diamond \lambda\})$ for any such (x, λ), as demanded by property (iv) in Claim 7. To verify property (ii), first note that $\{x, x \diamond \mathbf{0}\} = \{x, x\} = \{x\}$ for any $x \in X$. As we have just seen, this implies $U(x, \mathbf{0}) \leq V(\{x\})$. In turn, for any $(x', \lambda') \in (x, \lambda)^{\uparrow}$, by Claim 10(ii) and (SiM) we also have $\overline{V}(\{x', x' \diamond \lambda'\}) \geq V(\{x'\}) \geq V(\{x\})$. Hence, $U(x, \mathbf{0}) = V(\{x\})$ for $x \in X$.

To see why U is a weakly increasing function on D, simply note that for any $(x, \lambda), (\widetilde{x}, \widetilde{\lambda}) \in D$ with $(\widetilde{x}, \widetilde{\lambda}) \geq (x, \lambda)$, we have $(\widetilde{x}, \widetilde{\lambda})^{\uparrow} \subseteq (x, \lambda)^{\uparrow}$, which immediately implies $U(\widetilde{x}, \widetilde{\lambda}) \geq U(x, \lambda)$.

It remains to show that U is continuous on D. Let (x^n, λ^n) be a convergent sequence in D and put $(x, \lambda) := \lim(x^n, \lambda^n)$. Pick any $(x', \lambda') \in (x, \lambda)^{\uparrow}$. For every n, put $x'^n := x' \vee x^n$ and $\lambda'^n := \lambda' \vee \lambda^n$. Since D is a lattice, (x'^n, λ'^n) belongs to D for each n. Moreover, $\lim(x'^n, \lambda'^n) = (x' \vee x, \lambda' \vee \lambda) = (x', \lambda')$, implying that $\lim\{x'^n, x'^n \diamond \lambda'^n\} = \{x', x' \diamond \lambda'\}$. Since $\overline{V}$ is continuous, we then have

$$\lim \overline{V}(\{x'^n, x'^n \diamond \lambda'^n\}) = \overline{V}(\{x', x' \diamond \lambda'\}). \tag{21}$$

Also note that $(x'^n, \lambda'^n) \in (x^n, \lambda^n)^{\uparrow}$ for each n, and hence, $U(x^n, \lambda^n) \leq \overline{V}(\{x'^n, x'^n \diamond \lambda'^n\})$. By combining this observation with (21), we see that $\limsup U(x^n, \lambda^n) \leq \overline{V}(\{x', x' \diamond \lambda'\})$. Since (x', λ') is an arbitrary point in $(x, \lambda)^{\uparrow}$, it follows that $\limsup U(x^n, \lambda^n) \leq U(x, \lambda)$.

We will complete the proof by showing that $\liminf U(x^n, \lambda^n) \geq U(x, \lambda)$. Suppose by contradiction that $\liminf U(x^n, \lambda^n) < U(x, \lambda)$. By passing to a subsequence if necessary, assume $\lim U(x^n, \lambda^n)$ exists so that

$$\lim U(x^n, \lambda^n) < U(x, \lambda). \tag{22}$$

Note that by definition of U, for each n, there exists a point $(\widetilde{x}^n, \widetilde{\lambda}^n) \in (x^n, \lambda^n)^{\uparrow}$ such that $0 \leq \overline{V}(\{\widetilde{x}^n, \widetilde{x}^n \diamond \widetilde{\lambda}^n\} - U(x^n, \lambda^n) < 1/n$. It immediately follows that

$$\lim \overline{V}(\{\widetilde{x}^n, \widetilde{x}^n \diamond \widetilde{\lambda}^n\} = \lim U(x^n, \lambda^n). \tag{23}$$

Pick an ℓ dimensional constant vector $\mathbf{m} :=(m, ..., m)$ such that $\mathbf{m} \geq x \vee \lambda$. For every n, put $\widehat{x}^n := \widetilde{x}^n \wedge \mathbf{m}$ and $\widehat{\lambda}^n := \widetilde{\lambda}^n \wedge \mathbf{m}$. Since D is a lattice, $(\widehat{x}^n, \widehat{\lambda}^n)$

belongs to D for each n. The next step is to show that

$$\widetilde{x}^n \diamond \widetilde{\lambda}^n \geq \widehat{x}^n \diamond \widehat{\lambda}^n \quad \text{for every } n. \tag{24}$$

Put $\widetilde{\mathbf{a}} := \widetilde{x}^n \diamond \widetilde{\lambda}^n$ and $\widehat{\mathbf{a}} := \widehat{x}^n \diamond \widehat{\lambda}^n$ for a fixed n. Then, $\widetilde{\mathbf{a}}_1 := \widetilde{x}_1^n + \widetilde{\lambda}_1^n \geq \widetilde{x}_1^n \wedge m + \widetilde{\lambda}_1^n \wedge m := \widehat{\mathbf{a}}_1$. Now, fix any $i \in \{2, ..., \ell\}$ so that $\widetilde{\mathbf{a}}_i := \widetilde{x}_i^n - \widetilde{\lambda}_i^n$ while $\widehat{\mathbf{a}}_i := \widetilde{x}_i^n \wedge m - \widetilde{\lambda}_i^n \wedge m$. Since $\widetilde{x}_i^n \geq \widetilde{\lambda}_i^n$, there are three cases to consider: (1) $\widetilde{x}_i^n \geq \widetilde{\lambda}_i^n \geq m$; (2) $\widetilde{x}_i^n \geq m \geq \widetilde{\lambda}_i^n$; and (3) $m \geq \widetilde{x}_i^n \geq \widetilde{\lambda}_i^n$. In case (1), we have $\widehat{\mathbf{a}}_i = 0 \leq \widetilde{\mathbf{a}}_i$. In turn, case (2) implies $\widehat{\mathbf{a}}_i = m - \widetilde{\lambda}_i^n \leq \widetilde{\mathbf{a}}_i$. Finally, in case (3), $\widehat{\mathbf{a}}_i = \widetilde{\mathbf{a}}_i$. Thus, $\widehat{\mathbf{a}} \leq \widetilde{\mathbf{a}}$, as we sought.

Since we also have $(\widetilde{x}^n, \widetilde{\lambda}^n) \geq (\widehat{x}^n, \widehat{\lambda}^n)$, Claim 10(iv) and (24) imply

$$\overline{V}(\{\widetilde{x}^n, \widetilde{x}^n \diamond \widetilde{\lambda}^n\} \geq \overline{V}(\{\widehat{x}^n, \widehat{x}^n \diamond \widehat{\lambda}^n\}) \text{ for every } n. \tag{25}$$

Now, note that $(\widehat{x}^n, \widehat{\lambda}^n)$ is a bounded sequence, and hence, it has a convergent subsequence $(\widehat{x}^{n_k}, \widehat{\lambda}^{n_k})$. Moreover, as D is a closed set, $(\widehat{x}, \widehat{\lambda}) := \lim_k (\widehat{x}^{n_k}, \widehat{\lambda}^{n_k})$ belongs to D. We shall now show that

$$(\widehat{x}, \widehat{\lambda}) \geq (x, \lambda). \tag{26}$$

Suppose by contradiction that $\widehat{x}_i < x_i$ for some $i \in \{1, ..., \ell\}$. Then, we must also have $\widehat{x}_i < m$, for $m \geq x_i$ by definition of $\mathbf{m}$. Moreover, $\widehat{x}_i < m$ implies $\widehat{x}_i^{n_k} < m$ for all sufficiently large k. But for any such k, we have $\widehat{x}_i^{n_k} = \widetilde{x}_i^{n_k}$, implying that $\widehat{x}_i := \lim_k \widehat{x}_i^{n_k} = \lim_k \widetilde{x}_i^{n_k}$. In turn, $\lim_k \widetilde{x}_i^{n_k} \geq x_i$ since, by construction, $\widetilde{x}^n \geq x^n$ for each n while $\lim x^n := x$. This contradiction shows that $\widehat{x} \geq x$. By the same arguments, we also have $\widehat{\lambda} \geq \lambda$, which proves (26).

(26) simply means that $(\widehat{x}, \widehat{\lambda}) \in (x, \lambda)^{\uparrow}$. Thus, $U(x, \lambda) \leq \overline{V}(\{\widehat{x}, \widehat{x} \diamond \widehat{\lambda}\})$ by definition of U. Finally, note that $\overline{V}(\{\widehat{x}, \widehat{x} \diamond \widehat{\lambda}\}) = \lim_k \overline{V}(\{\widehat{x}^{n_k}, \widehat{x}^{n_k} \diamond \widehat{\lambda}^{n_k}\}) \leq \lim_n \overline{V}(\{\widetilde{x}^n, \widetilde{x}^n \diamond \widetilde{\lambda}^n\})$, where the equality follows from continuity of $\overline{V}$ while the weak inequality follows from (25). Thereby, we see that $U(x, \lambda) \leq \lim_n \overline{V}(\{\widetilde{x}^n, \widetilde{x}^n \diamond \widetilde{\lambda}^n\})$, a contradiction to (22) and (23). This completes the proof. $\square$

The final step is to extend the function U from D to $X \times \mathbb{R}_+^\ell$. For every $(x, \lambda) \in X \times \mathbb{R}_+^\ell$ set $g(x, \lambda) := (x, (\lambda_1, \lambda_{-1} \wedge x_{-1}))$, and note that g is a weakly increasing, continuous function from $X \times \mathbb{R}_+^\ell$ into D. From the corresponding properties of U, it obviously follows that $\overline{U} := U \circ g$ is a weakly increasing, and continuous function on $X \times \mathbb{R}_+^\ell$. Moreover, for any $(x, \lambda) \in D$, we have $g(x, \lambda) = (x, \lambda)$, and hence, $\overline{U}(x, \lambda) = U(x, \lambda)$. In view of Claims 7 and 11, we conclude that $\overline{U}$ is an mw-index for $\succsim$. This completes the proof of Theorem 3.

B3. Proof of Theorem 2

The "if" part of the theorem is a routine exercise. For the "only if" part, let $\succsim$ be a binary relation on $\mathcal{K}_{\mathcal{P}}$ that satisfies the axioms in Theorem 2. Then, as we have shown in Claim 1, $\succsim$ also satisfies the axioms in Theorem 3. Let $\overline{U} : X \times \mathbb{R}^{\ell}_{+} \to \mathbb{R}$ be an mw-index for $\succsim$, and set

$$V(A) := \max \overline{U}(x, y_1^*(A) - x_1, x_{-1} - y_{-1}^*(A)) \quad \text{for } A \in \mathcal{K}_{\mathcal{P}}.$$

The following claim uncovers a key implication of (FS).

Claim 12. $\overline{U}(x, 0, x_{-1}) = \overline{U}(x, 0, \mathbf{0})$ *for each* $x \in X$.

Proof. Fix an $x \in X$. Since $\overline{U}$ is weakly increasing, we trivially have $\overline{U}(x, 0, x_{-1}) \geq \overline{U}(x, 0, \mathbf{0})$. To prove the converse inequality, suppose by contradiction that

$$\overline{U}(x, 0, x_{-1}) > \overline{U}(x, 0, \mathbf{0}). \tag{27}$$

Then, $x_{-1} > \mathbf{0}$, and hence, the sets $A^n := \{x, (x_1 + \frac{1}{n}, \mathbf{0})\}$ and $B^n := \{x, (x_1 + \frac{1}{n}, \frac{n-1}{n} x_{-1})\}$ belong to $\mathcal{K}_{\mathcal{P}}$ for each $n \in \mathbb{N}$. Note that $\overline{U}(x, \frac{1}{n}, x_{-1}) \geq \overline{U}(x, 0, x_{-1})$ for each n, while $\lim \overline{U}((x_1 + \frac{1}{n}, \mathbf{0}), 0, \mathbf{0}) = \overline{U}((x_1, \mathbf{0}), 0, \mathbf{0}) \leq \overline{U}(x, 0, \mathbf{0})$. Thus, (27) implies that $\overline{U}(x, \frac{1}{n}, x_{-1}) > \overline{U}((x_1 + \frac{1}{n}, \mathbf{0}), 0, \mathbf{0})$ for all sufficiently large n. For any such n, we have $V(A^n) = \overline{U}(x, \frac{1}{n}, x_{-1}) > V(\{(x_1 + \frac{1}{n}, \mathbf{0})\})$ by definition of V. In particular,

$$\lim V(A^n) = \lim \overline{U}\left(x, 1/n, x_{-1}\right) = \overline{U}\left(x, 0, x_{-1}\right). \tag{28}$$

In turn, from (27) and (28), it follows that

$$\lim V(A^n) > \overline{U}(x, 0, \mathbf{0}), \tag{29}$$

implying that $\{x\} \prec A^n \succ \{(x_1 + \frac{1}{n}, \mathbf{0})\}$ for all sufficiently large n. But for any such n, (FS) implies $V(B^n) \geq V(A^n)$. Finally note that $\lim B^n = \{x\}$, and hence, $\lim V(B^n) = V(\{x\}) = \overline{U}(x, 0, \mathbf{0})$. Thus, we see that $\overline{U}(x, 0, \mathbf{0}) \geq \lim V(A^n)$, which contradicts (29). $\square$

Now, define a function $U : X \times \mathbb{R}_+ \to \mathbb{R}$, as for each $(x, \lambda_1) \in X \times \mathbb{R}_+$,

$$U(x, \lambda_1) := \overline{U}(x, \lambda_1, x_{-1}).$$

It is clear that U is weakly increasing and continuous on $X \times \mathbb{R}_+$ because of the corresponding properties of $\overline{U}$ on $X \times \mathbb{R}_+^\ell$.

Fix any $A \in \mathcal{K}_\mathcal{P}$, and put $y^* := y^*(A)$. In order to conclude that U is a w-index for $\succsim$, it remains to show that

$$\max_{x \in A} U(x, y_1^* - x_1) = \max_{x \in A} \overline{U}(x, y_1^* - x_1, x_{-1} - y_{-1}^*). \tag{30}$$

Notice that $\max_{x \in A} U(x, y_1^* - x_1) := \max_{x \in A} \overline{U}(x, y_1^* - x_1, x_{-1})$ is greater than or equal to the right hand side of (30) simply because $\overline{U}$ is weakly increasing.

To establish the converse inequality, first note that $U(y^*, 0) := \overline{U}(y^*, 0, y_{-1}^*) = \overline{U}(y^*, 0, \mathbf{0})$ by Claim 12. Thus, we only need to show that $\sup_{x \in A \backslash \{y^*\}} U(x, y_1^* - x_1)$ is less than or equal to the right hand side (30).

Pick any $\widetilde{x} \in A \backslash \{y^*\}$. Then, $\widetilde{x}_{-1} > \mathbf{0}$, and hence, $\{\widetilde{x}, (y_1^*, \mathbf{0})\}$ belongs to $\mathcal{K}_\mathcal{P}$. Note that the right hand side of (30) is greater than or equal to $\overline{U}(\widetilde{x}, 0, \mathbf{0})$ and $\overline{U}(y^*, 0, \mathbf{0})$. In turn, $\overline{U}(y^*, 0, \mathbf{0}) \geq \overline{U}((y_1^*, \mathbf{0}), 0, \mathbf{0})$. Thus, without loss of generality we can assume $\overline{U}(\widetilde{x}, y_1^* - \widetilde{x}_1, \widetilde{x}_{-1}) > \max\{\overline{U}(\widetilde{x}, 0, \mathbf{0}), \overline{U}((y_1^*, \mathbf{0}), 0, \mathbf{0})\}$, which simply means $\{\widetilde{x}\} \prec \{\widetilde{x}, (y_1^*, \mathbf{0})\} \succ \{(y_1^*, \mathbf{0})\}$. But then

$$\max_{x \in A} \overline{U}(x, y_1^* - x_1, x_{-1} - y_{-1}^*) \geq V(\{\widetilde{x}, y^*\}) \geq V(\{\widetilde{x}, (y_1^*, \mathbf{0})\}) \geq \overline{U}(\widetilde{x}, y_1^* - \widetilde{x}_1, \widetilde{x}_{-1}),$$

where the second inequality follows from (FS) while the others follow the definition of V. Since $\widetilde{x}$ is an arbitrary point in $A \backslash \{y^*\}$, we conclude that $\sup_{x \in A \backslash \{y^*\}} U(x, y_1^* - x_1) \leq \max_{x \in A} \overline{U}(x, y_1^* - x_1, x_{-1} - y_{-1}^*)$, as we sought. This completes the proof of Theorem 2.

B4. Additional Proofs

Proof of Proposition 1. Let us write y^* instead of $y^*(A)$, and suppose $y^* \notin \mathbf{C}_U(A)$. Pick any $x \in \mathbf{C}_U(A)$. Then, $U(x, y_1^* - x_1) > U(y^*, 0)$, so that $\{x, y^*\} \succ \{y^*\}$. Since $\widetilde{U}$ is another w-index for $\succsim$, we must then have $\widetilde{U}(x, y_1^* - x_1) > \widetilde{U}(y^*, 0)$. This immediately implies $y^* \notin \mathbf{C}_{\widetilde{U}}(A)$. Now, pick any $x' \in \mathbf{C}_{\widetilde{U}}(A)$, and suppose by contradiction that $x \notin \mathbf{C}_{\widetilde{U}}(A)$. Then, $\widetilde{U}(x', y_1^* - x_1') > \max\{\widetilde{U}(y^*, 0), \widetilde{U}(x, y_1^* - x_1)\}$, so that $\{x', x, y^*\} \succ \{x, y^*\}$. In turn, the latter condition implies $U(x', y_1^* - x_1') > U(x, y_1^* - x_1)$, which contradicts the fact that $x \in \mathbf{C}_U(A)$.

Thus, we have shown that $y^* \notin \mathbf{C}_U(A)$ implies (a) $\mathbf{C}_U(A) \subseteq \mathbf{C}_{\widetilde{U}}(A)$; and (b) $y^* \notin \mathbf{C}_{\widetilde{U}}(A)$. The proof follows from the symmetric implications of condition (b). □

Proof of Proposition 2. In view of Proposition 1, it suffices to show that $\mathbf{C}_U(A) = \mathbf{C}_{\widetilde{U}}(A)$ for any $A \in \mathcal{K}_{\mathcal{P}}$ such that $y^*(A) \in \mathbf{C}_U(A) \cap \mathbf{C}_{\widetilde{U}}(A)$. Let A be such a set, and take any $\widehat{x} \in \mathbf{C}_U(A)$ that is distinct from $y^*(A)$ so that $U(\widehat{x}, y_1^*(A) - \widehat{x}_1) = U(y^*(A), 0)$. From regularity of U, it follows that there exists a sequence $\{x^n, y^n\}$ in $\mathcal{K}_{\mathcal{P}}$ that converges to $\{\widehat{x}, y^*(A)\}$ such that $U(x^n, y_1^n - x_1^n) > U(y^n, 0)$ and $y_1^n > x_1^n$ for every n. Clearly, we must also have $\lim_n x^n = \widehat{x}$. Moreover, Proposition 1(i) implies $\mathbf{C}_{\widetilde{U}}(\{x^n, y^n\}) = \{x^n\}$ for every n. So, by continuity of $\widetilde{U}$, it follows that $\widehat{x} \in \mathbf{C}_{\widetilde{U}}(A)$. Therefore, $\mathbf{C}_U(A) \subseteq \mathbf{C}_{\widetilde{U}}(A)$, and symmetrically, we also have $\mathbf{C}_U(A) \supseteq \mathbf{C}_{\widetilde{U}}(A)$. $\square$

References

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

——— (1984): "The Crowding Out Effect of Government Transfers on Private Charitable Contributions: Cross Sectional Evidence," *National Tax Journal*, 37, 563-568.

Ahn, D. S. and T. Sarver (2013): "Preference for Flexibility and Random Choice," *Econometrica*, 81, 341-361.

Andreoni, J. (1988): "Privately Provided Goods in a Large Economy: The Limits of Altruism," *Journal of Public Economics*, 35, 57-73.

——— (1989): "Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence," *Journal of Political Economy*, 97, 1447-1458.

——— (1990):"Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving," *Economic Journal*, 100, 464-477.

——— (1993): "An Experimental Test of the Public-Goods Crowding-Out Hypothesis," *American Economic Review*, 83, 1317-1327.

——— (1995): "Cooperation in Public Goods Experiments: Kindness or Confusion?" *American Economic Review*, 85, 891-904.

——— (2006): "Philanthropy," in *Handbook of the Economics of Giving, Altruism and Reciprocity*, ed. by S-C. Kolm and J. M. Ythier. Amsterdam: North Holland, 1201-1269.

Andreoni, J. and T. Bergstrom (1996): "Do Government Subsidies Increase the Private Supply of Public Goods?" *Public Choice*, 88, 295-308.

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

Arrow, K. J. (1972): "Gifts and Exchanges," *Philosophy and Public Affairs*, 1, 343-62.

Becker, G. S. (1974): "A Theory of Social Interactions," *Journal of Political Economy*, 82, 1063-1093.

Bénabou, R. and J. Tirole (2006): "Incentives and Prosocial Behavior," *American Economic Review*, 96, 1652-1678.

Bergstrom, T., L. Blume, and H. Varian (1986): "On the Private Provision of Public Goods," *Journal of Public Economics*, 29, 25-49.

Berkowitz, L. (1973): "Reactance and the Unwillingness to Help Others," *Psychological Bulletin*, 79, 310-317.

Bernheim, B. D. (1986): "On the Voluntary and Involuntary Provision of Public Goods," *American Economic Review*, 76, 789-793.

Bolton, G. E. and E. Katok (1998): "An Experimental Test of the Crowding Out Hypothesis: The nature of Beneficent Behavior," *Journal of Economic Behavior and Organization*, 37, 315-331.

Brehm, J. W. (1966): *A Theory of Psychological Reactance*, New York: Academic Press.

Cherepanov, V., T. Feddersen, and A. Sandroni (2012): "Revealed Preferences and Aspirations in Warm Glow Theory," *Economic Theory*, forthcoming.

Clee, M. A. and R. A. Wicklund (1980): "Consumer Behavior and Psychological Reactance," *Journal of Consumer Research*, 6, 389-405.

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

Coate, S. and M. Conlin (2004): "A Group Rule Utilitarian Approach to Voter Turnout: Theory and Evidence," *American Economic Review*, 94, 1476-1504.

Deci, E. L. (1975): *Intrinsic Motivation*, New York: Plenum Press.

Dekel, E., B. L. Lipman, and A. Rustichini (2001): "Representing Preferences with a Unique Subjective State Space," *Econometrica*, 69, 891-934.

Dillenberger, D. and P. Sadowski (2012): "Ashamed to be Selfish," *Theoretical Economics*, 7, 99-124.

Eckel, C. C., P. J. Grossman, and R. M. Johnston (2005): "An Experimental Test of the Crowding Out Hypothesis," *Journal of Public Economics*, 89, 1543-60.

Epstein, L. G., M. Marinacci, and K. Seo (2007): "Coarse Contingencies and Ambiguity," *Theoretical Economics*, 2, 355-394.

Evren, O. (2012): "Altruism and Voting: A Large-Turnout Result That Does not Rely on Civic Duty or Cooperative Behavior," *Journal of Economic Theory*, 147, 2124-2157.

Faravelli, M. and R. Walsh (2011): "Mandate and Paternalism: A Theory of Large Elections," *mimeo*, University of Queensland.

Falk, A. and M. Kosfeld (2006): "The Hidden Costs of Control," *American Economic Review*, 96, 1611-1630.

Feddersen, T. and A. Sandroni (2006): "A Theory of Participation in Elections," *American Economic Review*, 96, 1271-1282.

Fehr, E. and B. Rockenbach (2003): "Detrimental Effects of Sanctions on Human Altruism," *Nature*, 422, 137-140.

Frey, B. S. and R. Jegen (2001): "Motivation Crowding Theory," *Journal of Economic Surveys*, 15, 589-611.

Funk, P. (2010): "Social Incentives and Voter Turnout: Evidence from the Swiss Mail Ballot System," *Journal of the European Economic Association*, 8, 1077-1103.

Gilboa, I., F. Maccheroni, M. Marinacci, and D. Schmeidler (2010): "Objective and Subjective Rationality in a Multiple Prior Model," *Econometrica*, 78, 755-770.

Glazer, A. and K. A. Konrad (1996): "A Signaling Explanation for Charity," *American Economic Review*, 86, 1019-1028.

Gneezy, U. and A. Rustichini (2000): "Pay Enough or Don't Pay At All," *Quarterly Journal of Economics*, 115, 791-810.

Gul, F. and W. Pesendorfer (2001): "Temptation and Self Control," *Econometrica*, 69, 1403-1435.

Harsanyi, J. C. (1953): "Cardinal Utility in Welfare Economics and in the Theory of Risk-Taking," *Journal of Political Economy*, 61, 434-435.

——— (1955): "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparison of Utility: Comment," *Journal of Political Economy*, 63, 309-321.

Hochman, H. M. and J. D. Rodgers (1973): "Utility Interdependence and Income Transfers through Charity," in *Transfers in an Urbanized Economy*, ed. by K. E. Boulding, M. Pfaff, and A. B. Pfaff. Belmont: Wadsworth, 63-77.

Horowitz, I. A. (1968): "Effect of Choice and Locus of Dependence on Helping Behavior," *Journal of Personality and Social Psychology*, 8, 373-376.

Houser, D., E. Xiao, K. McCabe, and V. Smith (2008): "When Punishment Fails: Research on Sanctions, Intentions and Non-Cooperation," *Games and Economic Behavior*, 62, 509-532.

Kreps, D. M. (1979): "A Representation Theorem for 'Preference for Flexibility'," *Econometrica*, 47, 565-577.

Mak, K-T. (1984): "Notes on Separable Preferences," *Journal of Economic Theory*, 33, 309-321.

McFadden, D. L. (2005): "Revealed Stochastic Preference: A Synthesis," *Economic Theory*, 26, 245-264.

McFadden, D. L. and M. K. Richter (1990): "Stochastic Rationality and Revealed Stochastic Preference," in *Preferences, Uncertainty, and Optimality*, ed. by J. Chipman, D. L. McFadden, and M. K. Richter. Boulder, CO: Westview Press, 166-186.

Mellström, C. and M. Johannesson (2008): "Crowding Out in Blood Donation: Was Titmuss Right?" *Journal of the European Economic Association*, 6, 845-863.

Myatt, D. P. (2012): "A Rational Choice Theory of Voter Turnout," *mimeo*, London Business School.

Noor, J. and L. Ren (2011): "Temptation and Social Preference," *mimeo*, Boston University.

Palfrey, T. R. and J. E. Prisbrey (1997): "Anomalous Behavior in Public Goods Experiments: How Much and Why?" *American Economic Review*, 87, 829-46.

Puppe, C. (1996): "An Axiomatic Approach to 'Preference for Freedom of Choice'," *Journal of Economic Theory*, 68, 174-199.

Riker, W. and P. Ordeshook (1968): "A Theory of the Calculus of Voting," *American Political Science Review*, 62, 25-42.

Roberts, R. D. (1984): "A Positive Model of Private Charity and Public Transfers," *Journal of Political Economy*, 92, 136-148.

Saito, K. (2013): "Pride, Shame, and Temptation to Act Selfishly," *mimeo*, California Institute of Technology.

Sarver, T. (2008): "Anticipating Regret: Why Fewer Options May Be Better," *Econometrica*, 76, 263-305.

Sen, A. (1985): *Commodities and Capabilities*, Amsterdam: North Holland.

——— (1988): "Freedom of Choice: Concept and Content," *European Economic Review*, 32, 269-294.

Steinberg, R. (1989): "The Theory of Crowding Out: Donations, Local Government Spending, and the 'New Federalism'," in *Philanthropic Giving*, ed. by R. Magat. Oxford: Oxford University Press, 143-56.

Sugden, R. (1998): "The Metric of Opportunity," *Economics and Philosophy*, 14, 307-337.

Titmuss, R. M. (1970): *The Gift Relationship: From Human Blood to Social Policy*, London: Allen and Unwin.

Wakker, P. P. (1989): *Additive Representations of Preferences*, Berlin: Springer.

Warr, P. G. (1982): "Pareto Optimal Redistribution and Private Charity," *Journal of Public Economics*, 13, 207-211.

——— (1983): "The Private Provision of a Public Good is Independent of the Distribution of Income," *Economics Letters*, 13, 207-211.