Tohoku University Research Center for Policy Design Discussion Paper

TUPD-2024-005

Norms and Emotions

Hiroki Saruya
Economic and Social Research Institute, Cabinet Office, Government of Japan

Masayuki Yagasaki
Graduate School of Economics and Management, Tohoku University

June 2024

Norms and Emotions*

Hiroki Saruya[†] and Masayuki Yagasaki[‡]

June 24, 2024

Click here for the latest version

Abstract

Social norms are an important determinant of behavior, but the behavioral and welfare effects of norms are not well understood. We propose and axiomatize a decision-theoretic model in which a reference point is formed by the decision maker's perceptions of which actions are admired (prescriptive norms) and which are prevalent (descriptive norms), and utility depends on the *pride* of exceeding the reference point or the *shame* of falling below it. The model is simple, yet provides a unified explanation for previous empirical findings, and is useful for welfare analysis of norm-evoking policies with a revealed preference approach.

Keywords: norms, reference dependence, pride, shame, public recognition, norm nudge

JEL classification: D80, D81, D90, D91

*This paper substantially extends Yagasaki's previously circulated paper "Pride, Shame and Social Comparisons." We thank Leonardo Bursztyn, Takaaki Hamada, Yosuke Hashidate, Hidehiko Ichimura, Ryota Iijima, Michihiro Kandori, Akihiko Matsui, Ran Spiegler, and Norio Takeoka for their invaluable feedback. Any remaining errors are the sole responsibility of the authors.

[†]Economic and Social Research Institute, Cabinet Office, Government of Japan, `hiroki.saruya@aya.yale.edu`

[‡]Graduate School of Economics and Management, Tohoku University, `yagamasajp@gmail.com`

1 Introduction

Social norms are receiving increasing attention as a key determinant of behavior in various contexts. Norms can take effect through simple interventions such as making decisions or outcomes publicly observable[1] or providing social information.[2] As a result, policymakers have become increasingly interested in social norms as a cost-effective policy lever to induce behavioral change.

Despite the growing interest, the behavioral and welfare effects of such policies are not well understood. Norm-evoking policies may produce desired behavioral outcomes in some cases, but they may fail to do so or even backfire in others.[3] Investigations of the behavioral effects of policies are hindered by the lack of theoretical foundations on what types of payoffs or constraints are generated by norms and how they are revealed from choice data. This gap also makes it unclear how revealed preferences are useful for welfare analysis in norm-conscious decision-making.

This paper presents a novel decision-theoretic model to describe the behavior of an individual who is concerned with social norms. We consider a two-stage choice problem (Gul and Pesendorfer 2001; Noor and Takeoka 2015) adapted to decisions under social image concerns (e.g. Dillenberger and Sadowski 2012; Saito 2015; Evren and Minardi 2017; Hashidate 2021). The decision maker (hereafter DM) first *privately* chooses a menu (i.e., choice set) and then *publicly* chooses an alternative from the menu. This setting naturally expresses the behavioral effect of norms by the discrepancy between

[1] Researchers have studied the effects of publicity, for example, on educational investment (Bursztyn and Jensen 2015), career choice (Bursztyn et al. 2017), tax compliance (Perez-Truglia and Troiano 2018), charitable giving (Butera et al. 2022, and see also DellaVigna et al. 2012), blood donations (Lacetera and Macis 2010), childhood vaccination (Karing 2024), and voting (Gerber et al. 2008). See Bursztyn and Jensen (2017) for a review of empirical framework and applications.

[2] Providing information about other individuals' behavior or normative opinions affects decisions on charitable donation (Frey and Meier 2004), tax compliance (Frey and Torgler 2007; Hallsworth et al. 2017), energy conservation (Schultz et al. 2007; Allcott 2011; Allcott and Rogers 2014), and female labor participation (Bursztyn et al. 2020).

[3] Publicity of decisions may increase or decrease target behavior (Bursztyn and Jensen 2015). Providing information about the behavior of others behavior may lead to the avoidance of a choice opportunity (Klinowski 2021) or an undesirable choice (Schultz et al. 2007).

preferences in the private (norm-free) and public (norm-conscious) stages, and is also suitable for studying the avoidance of choice opportunities (e.g., Dana et al. 2006) or the welfare effects of norms.[4] We axiomatize a utility representation called a *pride-shame representation*, in which utility depends on an endogenously derived reference point (cf. Ok et al. 2015; Lleras et al. 2019; Kıbrıs et al. 2023). The reference point expresses what the DM perceives as "normal" behavior in society.

A key feature of our model is that the reference point is determined by an interaction of two types of *subjective* norms, referred to as *descriptive norms* and *prescriptive norms*. Economists typically emphasize descriptive norms, which express the DM's perception of what behavior is prevalent or common, i.e., what others *choose* to do. In contrast, social psychologists also emphasize prescriptive norms (e.g., Cialdini et al. 1991; Bicchieri 2005; Bicchieri and Dimant 2022), which express the DM's perception of what behavior is approved of or admired, i.e., what others think one *should* do.[5] Although economists have studied prescriptive norms (e.g., Akerlof and Kranton 2000), they have not extensively studied how the two notions of norm interact. We show, through a simple application to prosocial behavior, that interactions between these norms can explain a variety of previously documented behavioral patterns. Crucially, the two types of norms are the subjective beliefs of the DM and are allowed to be biased (Miller and Prentice 1994; Bursztyn et al. 2020).

An essential determinant of behavior is social emotions, such as pride and shame, which arise from comparing one's own behavior with the typical behavior of others as reference behavior. To illustrate, consider a DM who expects a donation solicitor to arrive at her home shortly (DellaVigna et al. 2012). The DM's satisfaction with donating an amount, say \$10, depends on how she perceives the behavior of others. If she believes that her neighbors donate \$0, then she gains a positive sense of pride from the \$10 donation because her behavior is perceived as normatively superior to that of her neighbors. The

[4]The two-stage framework is also useful for studying the welfare effects of a product with consumption externalities (e.g., Bursztyn et al. 2023).

[5]Prescriptive norms are also known as injunctive norms. The terminology and the relationship of our work to the social psychology literature are discussed in Section 5.1.

degree of pride depends on the perceived desirability of each action: if donating \$10 is considered normatively more desirable than donating \$0 by a small (large) amount, then the payoff gain from pride is small (large). In contrast, if she believes that her neighbors donate \$100, she suffers a negative sense of shame from donating \$10 because her behavior is considered normatively inferior. The payoff loss from shame, in turn, depends on the perceived approval of each action. As this example illustrates, descriptive norms determine which behavior the DM focuses on as a reference behavior (donating \$0 or \$100) to compare her own choice (donating \$10) to, and prescriptive norms determine the payoff from the comparison. The norms then affect the DM's behavior. For example, suppose she initially plans to donate \$10, but then thinks that her neighbors are donating \$100. If a solicitor is already at her door, she may increase her planned donation to avoid shame. Alternatively, if the solicitor has not yet arrived, she may leave the house, thereby avoiding the opportunity to donate.

Using a simple example of prosocial behavior, we illustrate that our model provides useful insights for understanding empirical findings documented by previous studies. First, our model clarifies how the choice of an action depends on descriptive and prescriptive norms, and when policies such as providing social information or publicity may be ineffective. For example, if information about others' behavior (normative opinions) mainly affects the descriptive (prescriptive) norm of the DM, then changing this norm is the main mechanism behind the effect of providing information. The effectiveness of the policy then depends on how sensitive the perceived norms are to the policy and how the DM evaluates the resulting pride or shame. Similarly, the effectiveness of public observability depends on how private preferences (preferences over singleton menus) and public preferences (choices from menus) differ.

Second, the two-stage modeling allows us to study choice avoidance and the welfare implications of policies directly. For example, if a DM strictly prefers one menu $\{\$0\}$ over another menu $\{\$0, \$10\}$, this suggests that she is avoiding an opportunity to donate \$10, and a negative welfare effect of publicly making a choice.[6] In addition, our model illustrates how policies to influence perceived norms exert differential impacts on the participation in a donation opportunity and on the donation decision conditional on participation. For example, it can explain the laboratory findings of Klinowski (2021) that informing individuals about others' high level of donation *after* participation increases the amount donated, but doing so *before* participation discourages participation.

Third, our model can explain other empirical findings that are not necessarily emphasized in economics. For example, it can rationalize previous findings that providing information about descriptive or prescriptive norms is more effective at inducing prosocial behavior when they are aligned than when they are misaligned (Cialdini 2003), and that the descriptive norm has a greater influence in the latter case (Bicchieri and Xiao 2009; Allcott 2011; Hallsworth et al. 2017). An individual is more likely to make a charitable donation when others say donations should be done *and* they do donate, than when others say that donations should be done *but* they do not donate. Intuitively, when both norms point to prosocial behavior, failure to follow them generates shame. In contrast, if the prescriptive norm points to prosocial behavior but the descriptive norm points to the opposite, acting prosocially generates pride. If avoiding shame is a stronger motivator than seeking pride, which is empirically supported (DellaVigna et al. 2017; Butera et al. 2022), then aligned norms are more likely to induce prosocial behavior. Our model can also explain why providing information about the level of prosocial behavior of others can unintentionally reduce prosocial behavior (Schultz et al. 2007), illustrating the importance of considering individuals' perceived norms when introducing a policy.

The first step toward deriving the PS representation is to characterize the DM's subjective reference. Our approach is similar in spirit to that of Masatlioglu et al. (2012) and Kıbrıs et al. (2023), who elicit the DM's consideration and reference, respectively, by observing a "choice reversal," whereby removing

[6] Without the menu-selection stage, we might draw a false welfare conclusion: e.g., if we only observe that the DM chooses a \$10 donation from the menu $\{\$0, \$10\}$, we might mistakenly infer that adding the option to donate \$10 is beneficial.

an unchosen alternative from a menu affects the choice from the menu. Instead of requiring a choice reversal, we exploit observations such that removing an unchosen alternative affects the preference over menus. Suppose we observe that the DM donates \$10 whether or not she has the option to decline donation ($\mathcal{C}(\{\$0, \$10\}) = \mathcal{C}(\{\$10\}) = \{\$10\}$), but that she strictly prefers to donate with the option to decline ($\{\$0, \$10\} \succ \{\$10\}$). This suggests that the option to decline donation improves the DM's utility from donating by generating pride, which then implies that \$0 is the reference choice at the menu $\{\$0, \$10\}$. We generalize this observation to elicit a subjective reference set, i.e., the set of reference alternatives, at each menu.

The second key step is to characterize shame and pride by describing how preferences for smaller or larger menus emerge depending on the reference set. Consider first a DM who perceives that her neighbors do not donate. Then, answering the door to meet a solicitor will never hurt ($\{\$0, \$10\} \succeq \{\$0\}$), because she can decline the donation without shame, or she can even feel pride by choosing to donate. Lemma 2(i), derived from our axioms, formalizes this idea: the DM will exhibit a preference for larger menus (cf. Evren and Minardi 2017) when the additional alternatives do not enter the reference set. Next, consider a DM who privately does not want to donate ($\{\$0\} \succeq \{\$10\}$). Suppose she notices that some neighbors are donating \$10, so \$10 enters her reference set when she chooses between donating \$10 and not donating. Then the option to donate \$10 will not improve the DM's feelings about not donating, because of the shame of falling below her neighbors' standard. She then prefers to avoid the donation option ($\{\$0\} \succeq \{\$0, \$10\}$). Lemma 2(ii) characterizes such a preference: the DM will exhibit a preference for smaller menus (cf. Gul and Pesendorfer 2001; Dillenberger and Sadowski 2012) when the extra alternatives enter the reference set.[7] Technically, this property relaxes the set-betweenness axiom of Gul and Pesendorfer (2001). These two properties are crucial for our main result.

Our contribution is to propose a simple, tractable, and axiomatically founded

[7]To match this example to Lemma 2(ii) precisely, both \$0 and \$10 donations must enter the reference set. We leave a formal discussion to Section 3.

model of norm-conscious decision-making that can be useful for applied analysis. (1) Our model is simple: it captures the operation of norms through two functions, a descriptive norm function and a prescriptive norm function. The simple model can explain a variety of previous empirical findings. Moreover, distinguishing between the two types of norms clarifies potential mechanisms behind policy effects and facilitates policy analysis. (2) Our model is tractable in that it does not require the researcher to solve for an equilibrium; instead, the model is directly disciplined by observable choices.[8] This also means that we allow the DM's perceived norms to be biased. (3) The axiomatic foundation clarifies what kind of observed behavior is a distinctive feature of our PS model, and can be used as a basis for testing it (cf. Toussaert 2018). Because of the transparent link between choice and utility representation, the model also allows the researcher to conduct empirical research in a flexible manner. For example, we illustrate how the PS model can be used to investigate whether the DM feels pride or shame using choice data alone, which allows the researcher to discuss the welfare implications of pride and shame without using a survey (cf. Butera et al. 2022).

The paper is organized as follows. In Section 2, we illustrate our model and its implications by a simple example of prosocial behavior. Section 3 presents our axioms and the main representation result. Section 4 discusses how our model can be useful for empirical research. We provide a literature review in Section 5. Section 6 concludes. Proofs and additional results are presented in the Appendices.

2 Illustration of Results

Denote a typical menu of lotteries by A. A simplified version of our pride-shame utility representation has the following form. First, preference $\succeq$ over menus is represented by

[8] Similar to Maccheroni et al. (2012), we do not investigate how norms are formed (e.g., as an equilibrium of a game), and instead consider identification of (p)reference regardless of the norm generation process.

$$V_{PS}(A) = \max_{x \in A} \Big[u(x) - \underbrace{\max\{w(\varphi_r(A)) - w(x), 0\}}_{\text{“shame”}} + \beta \underbrace{\max\{w(x) - w(\varphi_r(A)), 0\}}_{\text{“pride”}} \Big], \quad (1)$$

where $\beta > 0$, $\varphi_r(A) = \arg\max_{a \in A} r(a)$, which is assumed to be a singleton for illustrative purpose,[9] and u, w, and r are von Neumann-Morgenstern (vNM) functions.[10] Second, the ex-post choice from each menu is determined by

$$\mathcal{C}_{PS}(A) = \arg\max_{x \in A} [u(x) - \max\{w(\varphi_r(A)) - w(x), 0\} + \beta \max\{w(x) - w(\varphi_r(A)), 0\}]. \quad (2)$$

The function u represents the DM's intrinsic utility function, which describes her private preference ranking.[11] The term $w(\varphi_r(A))$ represents a social reference point, which consists of two distinct components. First, the function r is called the *descriptive norm function*, which expresses the DM's perception of the prevalence of each alternative. $\varphi_r(A)$ is then interpreted as the alternative that the DM thinks is typically chosen by other people in her society. Second, the function w is called the *prescriptive norm function*, which expresses the DM's perception of the admirability of each alternative. Together, $w(\varphi_r(A))$ represents the normative desirability of the socially prevalent choice, as perceived by the DM. Crucially, we allow the DM to have biased beliefs about others' behavior or normative opinions.

The last two terms in Eq. (1) represent the utility of social emotions. If the DM facing menu A chooses an alternative x that is normatively inferior to the reference alternative, she feels *shame*, which reduces her utility by $w(\varphi_r(A)) - w(x) > 0$. Conversely, if she chooses x that is normatively supe-

[9] Our formal theory allows $\varphi_r(A)$ to contain multiple alternatives.

[10] vNM functions may not well suit some contexts of social decision-making, where the processes as well as the consequences of stochastic events matter. The example of Rabin (1995) illustrates this point: "If you have found \$100 on the street, and are trying to decide whether to turn it in, part of your dilemma is figuring out the likelihood that the owner will be found, that the local police officials are corrupt, etc.; such cases are captured by this model. But part of the moral decision-making may simply be thinking through whether or not you have a moral obligation to turn it over." See also Fudenberg and Levine (2012).

[11] u may capture not only her self-interest, but also social preferences such as altruism or moral concerns that are not influenced by social image concerns.

rior to the reference alternative, she feels *pride*, which increases her utility by $\beta\left[w(x) - w(\varphi_r(A))\right] > 0$. By allowing $\beta \neq 1$, we allow the DM to care about a downward deviation from the reference point (shame) differently from an upward deviation (pride).[12] $\beta \neq 1$ also allows the choice from menus to be reference-dependent.

Together, $U(x; A) \equiv u(x) - \max\left\{w(\varphi_r(A)) - w(x), 0\right\} + \beta \max\left\{w(x) - w(\varphi_r(A)), 0\right\}$ is interpreted as the utility of choosing x from menu A, and the value of A, $V_{PS}(A) = \max_{x \in A}\left[U(x; A)\right]$, is the maximum utility that the DM can obtain from the menu.

2.1 A Simple Example of Prosocial Behavior

We illustrate the implications of our model by the following simple example. Let $x \in A = \{0, 1\}$ denote the DM's choice of an alternative, where $x = 1$ indicates the DM engaging in prosocial behavior, and $x = 0$ indicates non-engagement. Let $u(0) = \bar{u} > 0 = u(1)$, $w(0) = 0 < \bar{w} = w(1)$, and $\beta\bar{w} < \bar{u} < \bar{w}$. Thus, the DM privately prefers the non-prosocial choice but believes that the prosocial choice is more admired. Also, $\beta < 1$ means that the DM is more sensitive to shame than she is to pride.

Benchmark behavior. The DM chooses $x = 0$ or $x = 1$ by comparing the utility of each alternative:

$$\begin{aligned} U(0; A) &= U(0; \{0,1\}, u, w, r) = \underbrace{\bar{u}}_{\text{intrinsic}} - \underbrace{\left[w(\varphi_r(\{0,1\})) - 0\right]}_{\text{shame}} = \bar{u} - w(\varphi_r(\{0,1\})) \\ U(1; A) &= U(1; \{0,1\}, u, w, r) = \underbrace{0}_{\text{intrinsic}} + \beta \underbrace{\left[\bar{w} - w(\varphi_r(\{0,1\}))\right]}_{\text{pride}} = \beta\left[\bar{w} - w(\varphi_r(\{0,1\}))\right] \end{aligned} \tag{3}$$

These expressions are simpler than the general one in Eq. (1) because $x = 0$ never causes pride and $x = 1$ never causes shame, regardless of the reference

[12] We emphasize the case with $\beta \in (0, 1)$, which expresses shame aversion (cf. Butera et al. 2022), although our theory allows for $\beta \geq 1$. Also, it accommodates $\beta = 0$ as a limit case.

alternative $\varphi_r(\{0,1\})$.

As a benchmark, suppose $r(1) < r(0)$, i.e., the DM believes that other people in her society do not typically engage in prosocial behavior. The reference alternative is $\varphi_r(\{0,1\}) = 0$ and the reference point is $w(0) = 0$. Choosing $x = 0$ gives the DM the intrinsic utility $\bar{u}$ and no utility from social emotion, because she chooses the action dictated by the norm. On the other hand, choosing $x = 1$ gives the DM zero intrinsic utility but gives a positive utility from pride. Since $\beta\bar{w} < \bar{u}$, the DM chooses $x = 0$.

Perceived norms and behavior. The model predicts how the DM's choice depends on the descriptive and prescriptive norms. Consider the following analysis, where each type of norm shifts toward prosocial behavior relative to the above benchmark.

(i) *Higher descriptive norm.* Suppose that the descriptive norm function becomes r' such that $r'(0) < r'(1)$, shifting the reference point to $w(\varphi_{r'}(\{0,1\})) = \bar{w}$. Now, choosing $x = 0$ gives the DM utility $\bar{u} - \bar{w} < 0$, whereas choosing $x = 1$ yields zero utility. Thus, the DM chooses $x = 1$.

(ii) *Higher prescriptive norm.* Suppose that the prescriptive norm function becomes w' such that $w'(0) = 0$ and $w'(1) = \bar{w}' > \frac{\bar{u}}{\beta}$. Then, choosing $x = 1$ provides a pride benefit of $\beta\bar{w}'$, which exceeds the utility $\bar{u}$ from $x = 0$. Thus, the DM chooses $x = 1$.

The DM switches to prosocial behavior $x = 1$ in both cases, but for different reasons. In case (i), she chooses $x = 1$ because she would feel shame if she stuck to the less admirable choice $x = 0$ while perceiving that others choose $x = 1$. By contrast, in case (ii), she chooses $x = 1$ because she feels greater pride from $x = 1$ perceiving that others choose $x = 0$.

This analysis is insightful for analyzing the effect of "norm nudges," which guide people's decisions by providing social information. Plenty of economic research has explored the effect on decisions of information about how others behave (e.g., Frey and Meier 2004; Allcott 2011) or what others think is the appropriate behavior (e,g, Hallsworth et al. 2017; Bursztyn et al. 2020). Our model helps clarify the mechanisms underlying such a norm-nudging. For example, if information about others' behavior (resp. normative opinions) mainly affects an individual's perceived descriptive (resp. prescriptive) norm, then the main mechanism of the effect of such information is described by case (i) (resp. case (ii)) above.[13] This analysis also suggests under what conditions providing each type of information is ineffective for inducing prosocial behavior. Informing the DM of others' choice fails to alter behavior if it does not affect the descriptive norm r because, e.g., the DM does not believe the information or because she thinks that the "others" are dissimilar to herself and out of her reference group. Similarly, informing her of others' normative opinions will not alter her behavior if it fails to affect the prescriptive norm w sufficiently. Alternatively, the latter information is ineffective if the DM does not derive much utility from pride (β being small).

Public recognition and prosociality. The model illustrates how public observability affects the DM's prosociality. Her choice of action under a private decision environment is expressed as a choice between two menus $\{0\}$ and $\{1\}$, with the utility from each option $V_{PS}(\{x\}) = u(x)$. By contrast, her choice under a public environment is expressed as a choice between two actions 0 and 1 from the menu $\{0, 1\}$, with the utility from each option $U(x; \{0, 1\}) = u(x) - \max\{w(\varphi_r(\{0, 1\})) - w(x), 0\} + \beta \max\{w(x) - w(\varphi_r(\{0, 1\})), 0\}$.[14] Because $U(x; \{0, 1\})$ is strictly increasing in $w(x)$, the DM becomes more prosocial in the public environment than in the private environment. The above analysis also indicates when policies such as public recognition programs are ineffective for inducing prosocial behavior: they are ineffective when the descriptive and prescriptive norms do not (sufficiently) favor prosocial behavior, or when the descriptive norm points to non-prosocial behavior and the DM is relatively

[13]In reality, information about one type of norms may also affect the perception of the other.

[14]The private and public preferences represent norm-free and norm-conscious preferences, respectively. Thus, our model applies to more general settings where some environmental cue (including publicity as an example) triggers the DM to focus on norms.

insensitive to pride.

Perceived norms and avoidance. The two-stage model enables us to study how norms affect the DM's decision to participate in the opportunity for prosocial behavior, as well as her decision on prosocial behavior itself. Analysis of the participation decision is important for two reasons. First, laboratory and field experiments have documented that a large fraction of individuals avoid opportunities to engage in prosocial behavior, even if they can choose non-engagement after participation and even if avoidance is costly (e.g., Dana et al. 2006; Broberg et al. 2007; Lazear et al. 2012; DellaVigna et al. 2012; Andreoni et al. 2017; Klinowski 2021). Our model clarifies how such avoidance depends on the perceived descriptive and prescriptive norms. Second, the participation decision is informative of the DM's willingness-to-pay (WTP) for public recognition and can be used to study the welfare impacts of policies such as public recognition programs, assuming that pride and shame are welfare-relevant. For example, her valuation of the menu $\{0,1\}$ relative to that of the singleton menu $\{0\}$ is informative of her WTP for public recognition.[15]

Our model illustrates how perceived norms can differentially impact participation and choice of action. Suppose that the DM first chooses whether to participate in the opportunity for prosocial behavior. If she decides to participate, she proceeds to the binary-choice stage described above. Alternatively, she can decide not to participate and be given a singleton menu $\{0\}$, which gives her utility $V_{PS}(\{0\}) = \bar{u}$. In the benchmark case, participation gives utility $V_{PS}(\{0,1\}) = \max\{U(0;\{0,1\}), U(1;\{0,1\})\} = \bar{u}$, and it is indifferent to non-participation. Therefore, the DM can optimally participate in the opportunity and then choose not to engage in prosocial behavior.

Now, suppose that the descriptive norm shifts toward prosocial behavior

[15]Butera et al. (2022) use an incentive-compatible mechanism to elicit individuals' WTP for public recognition, in the context of charitable behavior. In our framework, the researcher can infer individuals' WTP by observing their choices of menus instead of conducting a survey, which will be useful in some empirical contexts. Although the data requirement for our approach becomes demanding as the number of alternatives increases, additional restrictions (e.g., parametric assumption) can replace part of the data requirements.

(i.e., the descriptive norm function r changes to r'). A possible interpretation is that the DM updates her perception of the norm after she is given information about others' actions. As the above analysis shows, the DM switches to prosocial behavior *conditional on participation*. On the other hand, with the descriptive norm r', we have $V_{PS}(\{0,1\}) = \max\{\bar{u} - \bar{w}, 0\} = 0 < V_{PS}(\{0\})$, so the DM avoids the opportunity for prosocial behavior. Thus, the higher (more prosocial) descriptive norm induces the DM to take a prosocial action if she has no option to avoid the choice occasion, but it induces her to avoid the occasion if she has the option.

The theoretical predictions match the empirical evidence quite well. In a laboratory experiment, Klinowski (2021) demonstrates that (1) when individuals receive information that others have made a large donation after they participate in the opportunity, they increase the amount of donation relative to the no-information benchmark, whereas (2) when they receive the same information prior to the decision to participate, the participation rate drops relative to the benchmark. Our model can rationalize these findings by the shift of the descriptive norm caused by the information treatment.[16]

Aligned vs. misaligned norms. The simple model also explains why the descriptive and prescriptive norms induce larger behavioral change when they are aligned than when they are misaligned (Cialdini 2003), and why the descriptive norm tends to trump the prescriptive norm in the latter case (Tyran and Feld 2006; Bicchieri and Xiao 2009).[17] When the prescriptive norm dic-

[16] Klinowski (2021) also examines the effect of providing subjects with information that others have made a small amount of donation, before or after subjects make the participation decision. He finds that subjects are more likely to participate in the donation opportunity than the no-information benchmark if they learn about others' small donations before the participation decision. When they receive the information after participation, they contribute slightly more than the benchmark, though the difference is statistically insignificant. Although these results can be regarded as evidence for pride seeking, explaining these results will require some extension of the above simple model, and we omit this exercise.

[17] Allcott (2011) and Hallsworth et al. (2017) find evidence that descriptive norm messages are more effective than prescriptive norm messages for inducing electricity saving and tax compliance, respectively. Heinicke et al. (2022) find that descriptive norms exhibit a stronger correlation with behavior than prescriptive norms in the context of mini-dictator games.

tates the DM to engage in prosocial behavior ($w(0) < w(1)$) but the descriptive norm dictates otherwise ($r(0) > r(1)$), the DM behaves prosocially when the pride benefit $\beta\bar{w}$ is large enough to outweigh the intrinsic benefit of non-engagement, $\bar{u}$. By contrast, if both norms point to the prosocial behavior ($w(0) < w(1)$ and $r'(0) < r'(1)$), then she behaves prosocially as long as the shame cost $\bar{w}$ of non-engagement is large enough to offset its intrinsic benefit $\bar{u}$. If the DM is shame-averse ($\beta < 1$), which finds some empirical support,[18] the aligned norms induce prosocial behavior more effectively than misaligned ones. In words, when both norms point to prosocial behavior, failing to follow them causes shame for falling below social expectations, which is a strong motivator of prosocial behavior. By contrast, when the prescriptive norms point to prosocial behavior but the descriptive norms point to the opposite, the social motivation for prosocial behavior is pride from exceeding social expectations, which may not be so strong ("People say I should behave prosocially, but they do not live up to their words, so why do I?").

Other results. In Appendix C, we show that the model can explain other phenomena observed by previous empirical studies. For example, we argue that the model can explain why providing information on others' prosocial behavior can reduce the amount of prosocial behavior (e.g., Schultz et al. 2007).

3 Model

We adopt the framework of Gul and Pesendorfer (2001) (henceforth GP). Let (Z, ρ) be a compact metric space, where Z is a finite set of all prizes, and let $\Delta \equiv \Delta(Z)$ denote the set of all probability measures on the Borel σ-algebra of Z endowed with the weak topology. Denote by $\mathcal{A}$ a set of all closed subsets

[18] Butera et al. (2022) estimate social signaling models and find evidence for shame aversion. DellaVigna et al. (2017) find that non-voters in an election sort out of a survey due to the negative feeling from admitting non-voting or lying about it, while voters do not sort in to enjoy the positive feeling from saying that they voted.

of Δ, and endow $\mathcal{A}$ with the topology generated by the Hausdorff metric.[19] A typical lottery $a \in \Delta$ is called an alternative (or choice), and a typical element $A \in \mathcal{A}$, a set of alternatives, is called a menu (or choice set). Define $\alpha A + (1-\alpha)B \equiv \{z \in \Delta : z = \alpha a + (1-\alpha)b, a \in A, b \in B\}$ for $A, B \in \mathcal{A}$ and $\alpha \in [0,1]$.

We consider a DM who has a preference $\succeq$ over menus, as in the literature (Dillenberger and Sadowski 2012; Saito 2015; Evren and Minardi 2017; Hashidate 2021), and who also makes a choice from a menu according to a choice rule $\mathcal{C}$. Specifically, $\succeq$ is a binary relation over $\mathcal{A}$ and the choice correspondence $\mathcal{C} : \mathcal{A} \twoheadrightarrow \Delta$ satisfies $\emptyset \neq \mathcal{C}(A) \subseteq A$ for all $A \in \mathcal{A}$. We assume that both $\succeq$ and $\mathcal{C}$ are observed.

Below, we consider a DM who, prior to making a choice from a menu, forms a reference point based on her subjective beliefs. Specifically, she focuses on the alternative in the menu which she believes is most commonly chosen by others, and uses it as the reference alternative. The reference alternative establishes the reference point, and the DM evaluates her choice positively (negatively) if she believes it is more (less) admirable than the reference alternative. If multiple alternatives are perceived to be most common, the DM focuses on the one that she believes is most admirable. The beliefs on the commonality and admirability of alternatives shape the descriptive norm and the prescriptive norm, respectively.

3.1 Axioms

We first introduce some basic axioms.

Axiom 1. (Order) $\succeq$ *is complete and transitive.*

Axiom 2.

(i) (Lower Semi-Continuity) *For any* $A \in \mathcal{A}$, $\{B \in \mathcal{A} : A \succeq B\}$ *is closed.*

[19] That is, $d_H(A,B) = \max\left\{ \max_{a\in A} \min_{b\in B} d(a,b), \max_{b\in B} \min_{a\in A} d(a,b)\right\}$, where d is a metric that metrizes the weak topology.

(ii) (Upper von Neumann-Morgenstern Continuity) $A \succ B \succ C$ *implies* $B \succ \alpha A + (1-\alpha) C$ *for some* $\alpha \in (0,1)$.

(iii) (Upper Singleton Continuity) $\{\{b\} \in \mathcal{A} : \{b\} \succeq \{a\}\}$ *is closed.*

Axiom 1 is standard. Axioms 2(i)-(iii), similar to axioms in GP to characterize preferences without self-control, weaken the standard continuity assumption. These axioms yield a reference point that is constrained by a vNM function,[20] which will be interpreted as a social prevalence/commonality ranking. Such a specification seems attractive given that social preferences often feature discontinuities.[21]

We proceed by introducing a "reference relation" $\succeq_r$, which elicits the DM's subjective belief about the prevalence/commonality of each alternative from observed behavior.

Definition 1. (Reference relation)

(i) $a \succ^* b$ *if there exists* $A \ni b$ *such that* $A \cup \{a\} \succ A$ *and* $a \notin \mathcal{C}(A \cup \{a\})$.

(ii) $a \succ_r b$ *if either of the following conditions holds:*

- **a.** $a \succ^* b$.
- **b.** *There exists some* $c \in \text{int}(\Delta)$ *such that* $c \not\succ^* a$ *and* $c \succ^* b$.

(iii) $a \sim_r b$ *if neither* $a \succ_r b$ *nor* $b \succ_r a$, *and* $a \succeq_r b$ *if either* $a \succ_r b$ *or* $a \sim_r b$.

To interpret $\succ_r$, suppose $a \succ^* b$. Then, for some menu $A \ni b$,[22] adding a to A makes the menu more desirable ($A \cup \{a\} \succ A$) even though a is not chosen ($a \notin \mathcal{C}(A \cup \{a\})$). This suggests that the unchosen alternative a improves the value of the menu A by lowering its reference point. We then infer that the DM perceives a to be more prevalent than the other alternatives in A,

[20] More specifically, it will have a "Strotz representation" (Strotz 1955).

[21] E.g., an equal split of dictator game endowments (cf. Andreoni and Bernheim 2009) may be a discontinuity point.

[22] In Appendix B, we discuss why we do not confine Definition 1(i) to $A = \{b\}$.

including b. Case (ii-b) deals with technical difficulties which arise when a is on the boundary of Δ.[23]

We elicit the strict reference ranking from preferences for larger menus, but not from preferences for smaller menus ($A \succ A \cup \{a\}$). Although the latter implies that an additional alternative makes the menu unattractive by raising the reference point, the alternative may be just as prevalent as another alternative. Indeed, preferences for smaller menus emerge in the GP model, without a notion of the prevalence ranking of alternatives. In Section 3.4, we show that $\succeq_r$ elicits the true reference ranking r if the choice data are generated by a PS model that satisfies some weak notion of regularity. Thus, focusing on preferences for larger menus discards virtually no information.

Next, we define another binary relation which elicits the DM's subjective belief about the normative desirability/admirability of each alternative. It is "partial" in that it elicits the ranking only among alternatives with the same reference ranking (see Theorem 2).

Definition 2. (Partial normative relation)

(i) *$a \succ_w b$ if one of the following conditions holds.*

- **a.** $\{b\} \succ \{a, b\}$
- **b.** $\{b\} \sim \{a, b\}$ *and* $\mathcal{C}(\{a, b\}) = \{a\}$.
- **c.** $a \sim_r b$ *and* $\{a\} \sim \{a, b\} \succ \{b\}$.

(ii) *$a \sim_w b$ if neither $a \succ_w b$ nor $b \succ_w a$. $a \succeq_w b$ if either $a \succ_w b$ or $a \sim_w b$.*

The elicitation of the normative ranking is similar to that of the temptation ranking in GP-style models (GP; Noor and Takeoka 2015). Consider the first case (i-a), where the presence of alternative a makes the original menu $\{b\}$ less attractive, which implies that a raises the reference point. Thus, we infer that the DM believes a is more admirable than b. To interpret case (i-b), suppose that b sets the reference point at $\{a, b\}$. Then, because the reference point is the same as $\{b\}$ but the DM does not choose b at $\{a, b\}$, she must

[23] See Figure A2 and discussions in Appendix B.

be strictly better off at $\{a, b\}$, which contradicts $\{b\} \sim \{a, b\}$. Thus, the reference point must be higher at $\{a, b\}$ than at $\{b\}$, so the DM believes a is more admirable than b. Finally, in case (i-c), because both a and b enter the reference set at $\{a, b\}$, the reference point is weakly higher at $\{a, b\}$ than at $\{b\}$. Thus, $\{a, b\} \succ \{b\}$ implies that a is the unique choice at $\{a, b\}$. Then, $\{a\} \sim \{a, b\}$ implies that a must be weakly higher in the normative ranking than b (otherwise, the DM prefers to exclude b, so $\{a\} \succ \{a, b\}$). Because a PS preference such that the reference and normative rankings are identical cannot be distinguished from GP's temptation preference, below we focus on PS preferences such that they are distinct. Then, $a \sim_r b$ means that a must be strictly higher than b in the normative ranking.

When $a \succ_w b$, a is also revealed to be at least as high as b in the reference ranking; otherwise, a does not set the reference point at $\{a, b\}$, so nothing about its normative desirability is revealed. However, we do not use the observation $a \succ_w b$ to infer the reference ranking, because it does not tell us whether a is strictly higher than b or just as high as b in the reference ranking. As Theorem 2 shows, if choice data are generated by a regular PS preference, then any strict reference relation can be elicited by $\succ_r$ alone.

A natural process of reference formation is that the DM shapes beliefs about socially prevalent actions by imagining a "typical person," and uses that person's behavior as a reference. For example, a black woman living in an urban area may refer to the behavior of another person in a similar background, than that of a white man in a rural area. Thus, we impose axioms to rationalize the reference ranking as a preference of some "typical person." For simplicity, we directly impose axioms on $\succ^*$, $\succeq_r$, and $\succeq_w$, although we can rewrite these axioms as properties of $(\succeq, \mathcal{C})$.

Axiom 3. (r-EU)

(i) *If* $a \succ_r b$ *or* $a \succ_w b$*, then neither* $b \succ_r a$ *nor* $b \succ_w a$*.*

(ii) $\succ^*$ *is transitive. Also, if* $a \sim_r b \sim_r c$*,* $a \succeq_w b$*, and* $b \succeq_w c$*, then* $a \succeq_w c$*.*

(iii) **a.** $\{\alpha \in [0, 1] : \alpha A + (1 - \alpha)C \succeq B\}$ *is closed in [0,1].*

b. *If there exists* $a^* \in A$ *such that* $a^* \succ_r a$ *for all* $a \in A \setminus \{a^*\}$*, then for any* $\{A_n\}$ *such that* $A_n \to A$*,* $a_n \in \mathcal{C}(A_n)$ *and* $a_n \to a$*, we have* $a \in \mathcal{C}(A)$*.*

(iv) *For any* $\alpha \in (0,1)$*,* $\alpha a + (1-\alpha)c \succ^* \alpha b + (1-\alpha)c$ *and* $a \in \text{int}(\Delta)$ *imply* $a \succ^* b$*.*

Axiom 3(i) imposes consistency of reference relations and partial normative relations revealed at different menus.[24] Recall $a \succ_r b$ reveals that a is perceived to be more prevalent than b. Also, $a \succ_w b$ reveals that a is perceived to be at least as prevalent as b and more admirable than b. Then, to consistently rank alternatives in prevalence and admirability, the choice data should not reveal the opposite relations. Axiom 3(ii) states that the directly revealed reference ranking $\succ^*$ is transitive and that $\succeq_w$ is transitive on the indifference set for $\succeq_r$. Axiom 3(iii) expresses Archimedeanity of $\succeq_r$. Axiom 3(iv) imposes some linearity on the reference relation. Axiom 3(iv) is only required to deal with boundary elements.[25]

We next introduce a weak version of linearity of $(\succeq, \mathcal{C})$. Because of the menu effect induced by pride and shame, the standard independence axiom does not hold. For example, suppose that the reference ranking is $\$0 \succ_r \$100 \succ_r \$10$, i.e., the DM believes that donation is uncommon but the donation amount is large conditional on donating. Consider the preference between $\{\$10, \$100\}$ and $\{\$100\}$, with a \$100 donation being the reference alternative at both menus. If the shame from donating a small amount despite the social expectation of a large donation is strong, the DM will conform to the expectation and donate \$100 at both menus, so $\{\$10, \$100\} \sim \{\$100\}$. Now, consider two mixture menus, $0.5\{\$10, \$100\} + 0.5\{\$0, \$10\}$ and $0.5\{\$100\} + 0.5\{\$0, \$10\}$, with the reference alternative $0.5\$100 + 0.5\$0$ at both menus. If adding the possibility of zero donation lowers the reference point sufficiently, then having

[24]Similar axioms appear in Dillenberger and Sadowski (2012) and Kıbrıs et al. (2023).

[25]Note also that the interiority of a is important: if a is on the boundary of Δ, it is possible that we cannot reveal $a \succ^* b$ but can reveal $\alpha a + (1-\alpha)c \succ^* \alpha b + (1-\alpha)c$, if mixing with c brings a to the interior. See the example in Figure A2 and discussions in Appendix B.

an option to donate a small amount may be valuable because it offers a convenient compromise between self-interest and pride. Thus, the former mixture may be strictly better than the latter. This phenomenon happens because the effect of changing the reference point depends on whether each alternative falls above or below the reference point, so mixing can change the relative desirability of two alternatives.

The above discussion suggests that independence will hold if we consider mixtures of menus that preserve the relative desirability of alternatives. We consider a preference that satisfies independence among menus at which the DM can never feel shame and among menus at which the DM can never feel pride. To formalize the idea, let $L_r(a) = \{b \in \Delta : a \succ_r b\}$ denote the set of alternatives which is strictly below a in the reference ranking $\succeq_r$. For an arbitrary $a \in \Delta$, any $b \preceq_r a$ belongs to exactly one of the following sets:

$$\mathcal{P}(a) = \{b \in L_r(a) : \{a, b\} \succ \{b\} \text{ and } \mathcal{C}(\{a, b\}) = \{b\}\} \tag{4}$$

$$\mathcal{S}(a) = \{b \in L_r(a) : \{a, b\} \prec \{b\} \text{ and } \mathcal{C}(\{a, b\}) = \{b\}\} \tag{5}$$

$$\mathcal{N}_1(a) = \{b \in L_r(a) : \{a, b\} \sim \{b\} \text{ and } \mathcal{C}(\{a, b\}) = \{b\}\} \tag{6}$$

$$\mathcal{N}_2(a) = \{b \in L_r(a) : a \in \mathcal{C}(\{a, b\})\} \tag{7}$$

$$\mathcal{I}(a) = \{b \in \Delta : b \sim_r a\} \tag{8}$$

$\mathcal{P}(a)$ is the set of alternatives b with the reference ranking strictly below a such that the DM feels pride by choosing b at $\{a, b\}$. We provide formal definitions of pride and shame in Section 3.2; intuitively, because the DM chooses b at both $\{a, b\}$ and $\{b\}$ but is strictly better off at the former, we infer that the unchosen alternative a gives pride by lowering the reference point. Similarly, $\mathcal{S}(a)$ is the set of alternatives $b \prec_r a$ such that the DM feels shame by choosing b at $\{a, b\}$. This interpretation follows from the observation that the DM chooses the same b at both $\{a, b\}$ and $\{b\}$ but is strictly worse off at the former, implying that a raises the reference point. $\mathcal{N}_1(a)$ is the set of alternatives $b \prec_r a$ such that the DM feels neither pride nor shame at $\{a, b\}$ because the chosen alternative b is socially as desirable as the reference alternative a. $\mathcal{N}_2(a)$ is the set of alternatives $b \prec_r a$ such that the DM feels neither pride nor shame at $\{a, b\}$

because she chooses the reference alternative.[26] Finally, $\mathcal{I}(a)$ is the set of alternatives that is indifferent in the reference ranking to a.

To state linearity axioms, define the following subsets of $\mathcal{A}$:

$$\begin{aligned}
\mathcal{B}_P &= \{\{a,b\} \in \mathcal{A} : a = b \text{ or } b \in \mathcal{P}(a) \cup \mathcal{N}_1(a)\} \\
\mathcal{B}_S &= \{\{a,b\} \in \mathcal{A} : a = b \text{ or } b \in \mathcal{S}(a) \cup \mathcal{N}_1(a) \cup \mathcal{I}(a)\} \\
\mathcal{B}_N &= \{\{a,b\} \in \mathcal{A} : a = b \text{ or } b \in \mathcal{N}_2(a)\}.
\end{aligned}$$

Here, we do not distinguish $\{a,b\}$ from $\{b,a\}$: e.g., $a \in \mathcal{P}(b)$ implies $\{a,b\} \in \mathcal{B}_P$. $\mathcal{B}_P$ is the "pride domain" and $\mathcal{B}_S$ is the "shame domain." Specifically, for any $A \in \mathcal{B}_P$, the DM may feel pride (or no social emotion) but she can never feel shame. Similarly, for any $A \in \mathcal{B}_S$, the DM may feel shame (or zero social emotion) but not pride. Finally, $\mathcal{B}_N$ is the set of binary menus at which we cannot exclude the possibility of either pride or shame. Note we have $\cup_{j=P,S,N}\mathcal{B}_j = \{\{a,b\} : a,b \in \Delta\}$ and $\cap_{j=P,S,N}\mathcal{B}_j = \{\{a\} : a \in \Delta\}$.

We now state linearity axioms on $(\succeq, \mathcal{C})$.

Axiom 4. (Weak Independence, WI) *For any $\alpha \in (0,1)$,*

- **(i)** *$A,B,C \in \mathcal{B}_P$ and $A \succ (\succeq) B$ imply $\alpha A + (1-\alpha)C \succ (\succeq) \alpha B + (1-\alpha)C$.*
- **(ii)** *$A,B,C \in \mathcal{B}_S$ and $A \succ (\succeq) B$ imply $\alpha A + (1-\alpha)C \succ (\succeq) \alpha B + (1-\alpha)C$.*
- **(iii)** *$A,B \in \mathcal{A}$, $c \in \Delta$ and $A \succ (\succeq) B$ imply $\alpha A + (1-\alpha)\{c\} \succ (\succeq) \alpha B + (1-\alpha)\{c\}$.*

Axiom 5. (Weak Linearity, WL) *For any $a,b,c,d \in \Delta$ and $\alpha \in (0,1)$, the following properties hold.*

- **(i)** *Suppose $\{a,b\},\{c,d\} \in \mathcal{B}_P$ or $\{a,b\},\{c,d\} \in \mathcal{B}_S$. Then $\mathcal{C}(\alpha\{a,b\} + (1-\alpha)\{c,d\}) = \alpha\mathcal{C}(\{a,b\}) + (1-\alpha)\mathcal{C}(\{c,d\})$.*
- **(ii)** *Let $A = \alpha\{a,b\} + (1-\alpha)\{a,c\}$ and $b \in \mathcal{N}_2(a)$.*

[26] Strictly speaking, the DM can feel pride or shame by choosing some $b \in \mathcal{N}_2(a)$ at $\{a,b\}$ because it is possible to have $\mathcal{C}(\{a,b\}) = \{a,b\}$.

a. *If* $c \in \mathcal{P}(a)$, $\{a, \alpha b + (1-\alpha)c\} \succeq \alpha\{b\} + (1-\alpha)\{a, c\}$, *and* $\mathcal{C}(\{a, \alpha b + (1-\alpha)c\}) = \{\alpha b + (1-\alpha)c\}$, *then* $\mathcal{C}(A) = \alpha\mathcal{C}(\{a, b\}) + (1-\alpha)\mathcal{C}(\{a, c\})$.

b. *If* $c \in \mathcal{S}(a)$, $\alpha\{b\} + (1-\alpha)\{a, c\} \succeq \{a, \alpha b + (1-\alpha)c\}$, *and* $\mathcal{C}(\{a, \alpha b + (1-\alpha)c\}) = \{\alpha b + (1-\alpha)c\}$, *then* $\mathcal{C}(A) = \alpha\mathcal{C}(\{a, b\}) + (1-\alpha)\mathcal{C}(\{a, c\})$.

(iii) *For any* $A \in \mathcal{A}$, $\mathcal{C}(\alpha A + (1-\alpha)\{a\}) = \alpha\mathcal{C}(A) + (1-\alpha)\{a\}$.

Axiom 4 states that the standard independence property of $\succeq$ holds within the pride domain and shame domain, and it holds with respect to mixtures with a singleton. Similarly, Axiom 5(i) states that the linearity of choice holds within the pride domain and shame domain. Axiom 5(ii) is interpreted similarly, but requires an additional condition to exclude the possibility that one of the mixed menus generates pride and the other generates shame. To interpret (ii-a), note that from $b \in \mathcal{P}(a)$ and $c \in \mathcal{N}_2(a)$, we know that a is superior to b and c in the reference ranking. Suppose we additionally know $\{a, \alpha b + (1-\alpha)c\} \succeq \alpha\{b\} + (1-\alpha)\{a, c\}$ and $\mathcal{C}(\{a, \alpha b + (1-\alpha)c\}) = \{\alpha b + (1-\alpha)c\}$. The former suggests that moving a toward b makes the menu less desirable, and the latter suggests that this is not because a is a preferred choice. We can then infer that a sets a reference point lower than b. Thus, $\{a, b\}$ can generate pride but not shame, so the linearity of ex-post choice holds if it is mixed with a menu that never generates shame. Similarly, conditions in (ii-b) suggest that a sets the reference point higher than b, so the linearity of ex-post choice holds if $\{a, b\}$ is mixed with a menu that never generates pride. Finally, Axiom 5(iii) states that linearity holds with respect to mixtures with a singleton.

The next axiom relates preferences $\succeq$ to choice $\mathcal{C}$.

Axiom 6. (Sophistication) *Suppose there exists* $a^* \in A$ *such that* $a^* \succeq_r c$ *for all* $c \in A \cup B$ *and* $a^* \succeq_w a$ *for all* $a \in A$.

(i) *Suppose* $a^* \succeq_w b$ *for all* $b \in B$. *Then,* $A \cup B \succeq A$. *Moreover,* $A \cup B \succ A$ *if and only if* $\mathcal{C}(A \cup B) \cap A = \emptyset$.

(ii) *Suppose there exists* $b^* \in B$ *such that* $b^* \succ_w a^*$. *Then,* $A \cup B \succeq A$ *implies* $\mathcal{C}(A \cup B) \cap A = \emptyset$.

Axiom 6 is analogous to the Sophistication axiom of Noor and Takeoka (2015), with an important difference that our axiom depends on $\succsim_r$ and $\succsim_w$. Axiom 6(i) is about situations in which some $a^* \in A$ sets the reference point in $A \cup B$. When the menu A is augmented by another menu B, the DM weakly prefers the larger menu $A \cup B$ because it increases options without changing the reference point. Moreover, the larger menu should be strictly more desirable if and only if the newly added menu contains an option strictly preferred to all alternatives in A. Axiom 6(ii) concerns a situation in which some newly added alternative $b^* \in B$ sets a higher reference point than the reference point at A. With a higher reference point, the DM weakly prefers the larger menu only if the newly added menu B contains a strictly better alternative to be chosen than alternatives in A.

Our next axiom captures the DM's *shame attitude*, i.e., how her social payoff depends on the size and direction of the deviation of her choice's admirability from the reference point. We consider a DM whose marginal utility from pride and that from shame are constant, respectively, but who may care about pride and shame differently. The following axiom captures such an attitude toward pride and shame. Let $e^{a,b}$ denote an alternative such that $\left\{e^{a,b}\right\} \sim \{a, b\}$. For $\{a, b\} \in \mathcal{B}_S$, such $e^{a,b}$ exists by Lemma 12.

Axiom 7. (Piecewise linear social payoffs) *There exists a unique* $\alpha \in (0, 1)$ *such that for any* $a, b, c, d \in \Delta$ *such that* $c \in \mathcal{P}(a) \cap \mathcal{P}(b)$ *and* $d \in \mathcal{S}(a) \cap \mathcal{S}(b)$, *we have* $\alpha \{a, c\} + (1 - \alpha) \left\{e^{b,d}\right\} \sim \alpha \{b, c\} + (1 - \alpha) \left\{e^{a,d}\right\}$.

To interpret Axiom 7, suppose $c \in \mathcal{P}(a) \cap \mathcal{P}(b)$ and $\{a, c\} \succ \{b, c\}$. At each menu, a or b is the reference alternative and the DM feels pride by choosing c. Therefore, the preference $\{a, c\} \succ \{b, c\}$ must be because a is considered normatively inferior to b and generates higher pride for choosing the same alternative c. Suppose also $d \in \mathcal{S}(a) \cap \mathcal{S}(b)$. Then, similarly, the DM prefers $\{a, d\}$ to $\{b, d\}$ because a generates lower shame than b does for choosing the same alternative d. Now, consider the choice between two lotteries: lottery 1 yields the payoff from the high-pride menu $\{a, c\}$ or that from the high-shame menu $\{b, d\}$ with probability α and $1 - \alpha$, respectively, and lottery 2 yields the payoff from the low-pride menu $\{b, c\}$ or that from the low-shame menu $\{a, d\}$ with the same mixing rate. As α increases, lottery 1 becomes more desirable, and the DM will be indifferent between the lotteries at some α. Such α captures the rate at which the DM trades off the gains from more pride with the loss from more shame. Axiom 7 states that this tradeoff rate is constant across alternatives involved. Moreover, the tradeoff rate measures the degree of shame aversion: the higher α, the more pride gain the DM demands to compensate for the loss from shame.

Definition 3. (Shame attitudes) *(i) The DM is α-sensitive to shame if her preference $(\succeq, \mathcal{C})$ satisfies Axiom 7 with $\alpha \in (0, 1)$. (ii) The DM who is α-sensitive to shame is shame-averse if $\alpha > \frac{1}{2}$; shame-neutral if $\alpha = \frac{1}{2}$; and shame-loving if $\alpha < \frac{1}{2}$.*

Our final axiom imposes some consistency of choices across menus. Consider the donation example above. There, the DM believes that donation is uncommon but a large donation is common conditional on donating. Then, she may choose \$100 from $\{\$10, \$100\}$ to avoid the shame of falling behind the social expectation of a large donation, whereas she may choose \$10 from $\{\$0, \$10, \$100\}$ because a small donation nicely balances self-interest with pride from exceeding the social expectation of zero donation. This choice pattern violates the WARP. By contrast, if the reference point remains at \$100 after the no-donation option is added, then the DM at menu $\{\$0, \$10, \$100\}$ will not choose \$10, because the size of her shame from doing so is unchanged. She will choose \$0 or \$100, depending on the relative importance of private vs. social payoffs. This example suggests the following axiom.

Axiom 8. (Weak WARP, WWARP) *For any $A, B \in \mathcal{A}$, suppose there exists $a^* \in A \cap B$ such that $a^* \succeq_r c$ and $a^* \succeq_w c$ for all $c \in A \cup B$. Then, $a, b \in A \cap B$, $a \in \mathcal{C}(A)$ and $b \in \mathcal{C}(B)$ imply $a \in \mathcal{C}(B)$.*

An alternative $a^* \in A$ sets the reference point at menu A if $a^* \succeq_r a$ and $a^* \succeq_w a$ hold for all $a \in A$.[27] Therefore, Axiom WWARP says that the

[27] Such a^* exists for any finite A; see Lemma 7.

standard WARP property applies to menus A and B which share a common reference-setting alternative a^*

3.2 Representation Theorem

We show that the above axioms characterize the following utility representation.

Definition 4. $(\succeq, \mathcal{C})$ *is a* pride-shame (PS) preference *if there are continuous and linear functions* u,w *and* r *and a constant* $\beta > 0$ *such that* $\succeq$ *is represented by*

$$V_{PS}(A) = \max_{x \in A} \Big[u(x) - \underbrace{\max \Big\{ \max_{y \in \varphi_r(A)} w(y) - w(x), 0 \Big\}}_{\text{“shame”}} + \beta \underbrace{\max \Big\{ w(x) - \max_{y \in \varphi_r(A)} w(y), 0 \Big\}}_{\text{“pride”}} \Big] \tag{9}$$

where $\varphi_r(A) = \arg\max_{a \in A} r(a)$*, and* $\mathcal{C}$ *coincides with*

$$\mathcal{C}_{PS}(A) = \arg\max_{x \in A} \Big[u(x) - \max \Big\{ \max_{y \in \varphi_r(A)} w(y) - w(x), 0 \Big\} + \beta \max \Big\{ w(x) - \max_{y \in \varphi_r(A)} w(y), 0 \Big\} \Big]. \tag{10}$$

The representation (9)-(10) *is called a* PS representation.

$\max_{y \in \varphi_r(A)} w(y)$ is interpreted as the normative desirability which the DM perceives is expected to achieve, which we simply call the reference point. The reference point consists of two distinct components. First, the *descriptive norm function* r represents the DM's belief about the prevalence/commonality of each alternative. The DM's reference set $\varphi(A)$ consists of alternatives which she believes is the most prevalent in A. Second, the *prescriptive norm function* w represents the DM's belief about the social desirability/approvedness of each alternative. When $\varphi(A)$ contains multiple alternatives, the DM adopts the highest value of w as the reference point. The second term in Eq. (9) then represents the social utility that the decision maker derives from social emotion.

We say that a DM with a PS preference *feels pride by choosing* $a \in A$ *at* A if $w(a) - \max_{y \in \varphi_r(A)} w(y) > 0$, and the DM *feels shame by choosing* $a \in A$ *at* A if $w(a) - \max_{y \in \varphi_r(A)} w(y) < 0$. In words, the DM feels pride (shame) if she chooses an alternative that she perceives is normatively superior (inferior) to the reference alternative. Pride (shame) gives the DM a positive (negative) payoff from social emotion. Because the DM may care about shame differently than pride (Butera et al. 2022), we allow the DM to be more or less sensitive to shame than to pride, by allowing $\beta \neq 1$.

The PS representation includes GP's temptation model as a degenerate case. For our main theorem, however, we focus on nondegenerate PS preference defined as follows.

Definition 5. (Nondegeneracy) *Preference* $(\succeq, \mathcal{C})$ *is* nondegenerate *if there exists* $x, y, y' \in \Delta$ *such that* $\{x, y\} \succ \{y\}$, $\mathcal{C}(\{x, y\}) = \{y\}$, $\{y'\} \succ \{x, y'\}$ *and* $\mathcal{C}(\{x, y'\}) = \{y'\}$.

Nondegeneracy ensures that the DM feels pride at some binary menu and shame at another. The temptation preference of GP is degenerate, because it never generates pride. In Section 3.4, we show that such a degenerate preference accommodates non-unique representations. Thus, our main theorem focuses on preferences that accommodates a unique (up to positive affine transformation) representation, and we treat degenerate cases separately. Note that nondegeneracy is testable.

We now state our main theorem.

Theorem 1. *A nondegenerate preference* $(\succeq, \mathcal{C})$ *satisfies Axioms 1-8 if and only if it admits a PS representation. Moreover, the decision maker is shame-averse if and only if* $\beta < 1$, *shame-neutral if and only if* $\beta = 1$, *and shame-loving if and only if* $\beta > 1$.

3.3 Sketch of the Proof of Theorem 1

Our proof, formally presented in Appendix A, begins by verifying that the reference relation $\succeq_r$ admits a linear representation r (i.e., $a \succeq_r b \Leftrightarrow r(a) \geq r(b)$ and $r(\alpha a + (1-\alpha)b) = \alpha r(a) + (1-\alpha) r(b)$).

Lemma 1. *Suppose Axioms 1-5 hold. Then* $\succeq_r$ *admits a linear representation* r*. The representation is unique up to positive affine transformation.*

Define the reference correspondence as $\varphi_r(A) = \{a \in A : r(a) \geq r(b)\ \forall b \in A\}$. We can then show the following important properties of φ_r.

Lemma 2. *Suppose Axioms 1-6 hold. Then* φ *satisfies the following properties.*

(i) $\varphi_r(A \cup B) = \varphi_r(A)$ *implies* $A \cup B \succeq A$.

(ii) $\varphi_r(A \cup B) = \varphi_r(A) \cup \varphi_r(B)$ *and* $A \succeq B$ *imply* $A \succeq A \cup B \succeq B$.

Lemma 2(i) states that if augmenting menu A by menu B does not affect the reference set, then the DM weakly prefers the larger menu. In this case, the DM exhibits a preference for larger menus (cf. Evren and Minardi 2017) because the addition will never worsen social emotion. On the other hand, Lemma 2(ii) states that if the addition of alternatives expands the reference set, then the set betweenness property (Gul and Pesendorfer 2001) holds.[28] In particular, the DM exhibits a preference for smaller menus (cf. Dillenberger and Sadowski 2012) because the added alternatives will just raise the bar and will never improve social emotion.

Lemma 2 implies that the preference over finite menus can be characterized by at most two elements in each menu.

Lemma 3. *Suppose Axioms 1-6 hold. Then, for any finite menu* $A \in \mathcal{A}$*, there exist* $a^* \in A$ *and* $b^* \in \varphi_r(A)$ *such that* $A \sim \{a^*, b^*\}$.

The remaining components u and w can be constructed in a way similar to that of Gul and Pesendorfer 2001, although we address several technical difficulties due to the violation of independence and WARP. We first show that there exists a function V_{PS} that represents $\succeq$ and satisfies some linearity. Let $\mathcal{A}_f$ denote the set of all finite menus in $\mathcal{A}$.

[28] The two cases analyzed in Lemma 2 are exhaustive.

Lemma 4. *Suppose Axioms 1-6 hold. Then there exists a function V_{PS} that represents $\succeq$ on $\mathcal{A}_f$ and satisfies the following property: $A, B \in \mathcal{B}_P$ or $A, B \in \mathcal{B}_S$ implies $V_{PS}(\alpha A + (1-\alpha)B) = \alpha V_{PS}(A) + (1-\alpha)V_{PS}(B)$.*

Domain-wise linearity of V_{PS} together with nondegeneracy allows us to construct a function w_P on the pride domain and w_S on the shame domain. Then, Axiom 7 implies that the two functions are proportional: $w_P(x) = \beta w_S(x) =: w(x)$ for some $\beta > 0$.[29]

We can show that the PS representation holds for all binary menus. Then, Lemma 3 allows us to extend the representation to all finite menus, and Axiom 2 further extends the result to all menus in $\mathcal{A}$. For choice $\mathcal{C}$, Axiom 8 extends the representation to all menus.

3.4 Identification of PS Representation

3.4.1 Uniqueness of Reference Ranking

Given the construction of $\succeq_r$, its representation r is unique up to positive affine transformation (see also Proposition 1). However, we can ask: is there an alternative way to construct reference (leading to a different function r') to represent the same data by another PS preference? The definition of $\succeq_r$ adopts a specific way to complete the relation, and for general (not necessarily regular) cases, there can be two PS representations with different references which both represent the same choice data.

Example 1. *Suppose choice data $(\succeq, \mathcal{C})$ are generated by the temptation preference of GP: $V_{GP}(A) = \max_{x \in A}\{u(x) + w(x) - \max_{y \in A} w(y)\}$. In this case, a strict reference relation $a \succ_r b$ never occurs, so the descriptive norm function which rationalizes $\succeq_r$ is a constant. However, the choice can also be represented by another PS preference with $r = w$, because $\max_{y \in \varphi_w(A)} w(y) = \max_{y \in A} w(y)$. Thus, choice data cannot distinguish the two models.*

[29]We construct w in a way similar to GP. Another possible approach is to construct it directly from $\succeq_w$. However, $\succeq_w$ is guaranteed to elicit the normative ranking only between alternatives with the same reference ranking, so such a proof will involve an incomplete relation. Instead, we construct w from $(\succeq, \mathcal{C})$ and then show in Theorem 2 that w indeed represents $\succeq_w$ among alternatives with the same reference ranking.

However, we show below that when data are generated by a PS preference that belongs to some nondegenerate class, $\succeq_r$ correctly elicits the true reference. Therefore, within that class, the reference function r which is consistent with observed data is unique up to positive affine transformation. It turns out that the following weak form of nondegeneracy is sufficient to ensure unique and correct elicitation of the true reference by $\succeq_r$.

Definition 6. (Weak Nondegeneracy) *Preference* $(\succeq, \mathcal{C})$ *is* weakly nondegenerate *if there exist some* $\bar{a}, \bar{b} \in \Delta$ *such that* $\bar{a} \succ^* \bar{b}$.

Theorem 2. *Suppose the data are generated by a weakly nondegenerate PS preference. Then the following statements hold. (i)* $r(a) > r(b)$ *if and only if* $a \succ_r b$. *(ii) For* $a, b \in \Delta$ *such that* $r(a) = r(b)$, $w(a) > w(b)$ *if and only if* $a \succ_w b$.

Because Weak Nondegeneracy is implied by Nondegeneracy, the representation result in Theorem 1 focuses on the case where observed data are consistent with a unique reference ranking r, which is exactly the relation elicited by $\succeq_r$. Also, under Weak Nondegeneracy, the normative relation $\succeq_w$ is consistent with a unique normative ranking of w among alternatives with the same reference ranking. Note that the PS preference that appears in Example 1 is excluded by Weak Nondegeneracy.

3.4.2 Uniqueness of (u, w, r, β)

We can also show that u and w are unique up to affine transformation and that β is unique, when $(\succeq, \mathcal{C})$ satisfies the above axioms.

Proposition 1. *Suppose a nondegenerate* $(\succeq, \mathcal{C})$ *satisfies Axioms 1-8. Then the following statements are equivalent.*

(i) *If a PS representation* (u, w, r, β) *represents* $\succeq$, *then* (u', w', r', β') *also represents* $\succeq$.

(ii) *The following properties hold.*

a. $u' = \alpha u + \gamma_u$ *and* $w' = \alpha w + \gamma_w$ *for some* $\alpha > 0$ *and* $\gamma_u, \gamma_w \in \mathbb{R}$.

b. $r' = \alpha_r r + \gamma_r$ *for some* $\alpha_r > 0$ *and* $\gamma_r \in \mathbb{R}$.

c. $\beta = \beta'$

Although r is unique up to positive affine transformation, there is no guarantee that $\alpha_r = \alpha$.

3.5 Comparing Shame Aversion

Axiom 7 and Definition 3 yield a notion of a DM being "more shame-averse" than another DM. For two DMs $i = 1, 2$, let $(\succeq_i, \mathcal{C}_i)$ denote the preference of DM i.

Definition 7. *Suppose DM* $i \in \{1, 2\}$ *is* α_i*-sensitive to shame. Then DM 1 is (weakly) more shame-averse than DM 2 if* $\alpha_1 > (\geq)\alpha_2$.

The definition captures the idea that, when the degree of shame aversion increases, the DM who faces a lottery involving pride and shame of equal size will demand a higher chance of pride to compensate for the utility cost of shame. We now state how the PS representation and observed behavior are linked in terms of shame aversion. For $i = 1, 2$, let $\mathcal{P}_i$ and $\mathcal{S}_i$ denote the set of pride-generating binary menus $\mathcal{P}$ and the set of shame-generating binary menus $\mathcal{S}$ defined in Eq. (4) and (5), respectively, for DM i. Also, let $e_i^{x,y} \in \Delta$ be such that $\{e_i^{x,y}\} \sim_i \{x, y\}$.

Proposition 2. *Suppose DM 1 and DM 2 have a PS preference, with parameters* β_1 *and* β_2*, respectively. Then the following statements are equivalent.*

(i) $\beta_1 < (\leq)\beta_2$.

(ii) *DM 1 is (weakly) more shame-averse than DM 2.*

(iii) *Take any* $\alpha \in (0, 1)$ *and any* a, b, c, d *such that* $c \in \mathcal{P}_i(a) \cap \mathcal{P}_i(b)$, $d \in \mathcal{S}_i(a) \cap \mathcal{S}_i(b)$, $\{a, c\} \succ_i \{b, c\}$ *and* $\{a, d\} \succ_i \{b, d\}$ *for* $i = 1, 2$. *Then* $\alpha \{b, c\} + (1 - \alpha)\left\{e_2^{a,d}\right\} \succeq_2 \alpha \{a, c\} + (1 - \alpha)\left\{e_2^{b,d}\right\}$ *implies* $\alpha \{b, c\} + (1 - \alpha)\left\{e_1^{a,d}\right\} \succ_1 (\succeq_1)\alpha \{a, c\} + (1 - \alpha)\left\{e_1^{b,d}\right\}$.

By the equivalence of (i) and (ii), β in the PS representation characterizes the DM's shame aversion. Also, the equivalence of (ii) and (iii) implies that we can compare the shame aversion of two DMs by the following experiment: Consider two lotteries, lottery 1 giving the payoff of a high-pride menu $\{a, c\}$ and the payoff of a high-shame menu $\{b, d\}$ with probability α and $1 - \alpha$ respectively, and lottery 2 giving the payoff of a low-pride menu $\{b, c\}$ and the payoff of a low-shame menu $\{a, d\}$ with probability α and $1 - \alpha$ respectively. Ask the DMs to choose between the two lotteries at various α. Then, DM1 is more shame-averse than DM2 if DM1 chooses the "safer" lottery 2 whenever DM2 does.

4 Empirical Perspective

Although typically available data will be insufficient to completely verify our axioms, our PS model can still be used to understand the nature of individuals' perceived norms and resulting social emotions. In this section, we discuss some implications of our model that are potentially useful for empirical analysis.

4.1 Testing for the Independence of Norms and Preferences

Researchers may be interested in testing whether the descriptive norm and prescriptive norm are distinct, or whether the choice prescribed by norms coincides with the intrinsically best alternative.

Claim 1. *Suppose the DM has a PS preference. Then, the following statements hold, where $\|$ represents equality up to positive affine transformation.*

- **(i)** $r \nparallel w$ *if Weak Nondegeneracy holds.*
- **(ii)** $u \nparallel w, r$ *if there exist some* $A \in \mathcal{A}$ *and* $a \in A$ *such that* $\{a\} \succ A$.

Proof. (i) By assumption, there exists $A \in \mathcal{A}$ and $a \in \Delta$ such that $A \cup \{a\} \succ A$ and $a \notin \mathcal{C}(A \cup \{a\})$. By the PS representation, $\max_{y \in \varphi_r(A \cup \{a\})} w(y) <$

$\max_{y\in\varphi_r(A)} w(y)$ (otherwise, we would have $A \succeq A \cup \{a\}$). Thus, for $b \in \arg\max_{y\in\varphi_r(A)} w(y)$, we have $r(a) > r(b)$ and $w(a) < w(b)$. (ii) Without loss of generality, assume $\{a\} \succeq \{b\}$ for all $b \in A$. By assumption, $\max_{y\in\varphi_r(A)} w(y) = w(c) > w(a)$ for some $c \in A$ (otherwise, we would have $A \succeq \{a\}$). Therefore, we have $u(a) \geq u(c)$, $w(a) < w(c)$, and $r(a) \leq r(c)$. If we further had $u(a) = u(c)$, then $A \succeq \{c\} \sim \{a\}$, a contradiction. Thus, $u(a) > u(c)$. □

For example, observing $\{a, b\} \succ \{b\}$ and $a \notin \mathcal{C}(\{a, b\})$ for some a, b is sufficient to conclude $r \not\parallel w$, and observing $\{a\} \succ \{a, b\}$ for some a, b is sufficient to conclude $u \not\parallel w, r$.

4.2 Testing for Pride and Shame

The PS representation allows us to infer whether the DM feels pride or shame (or no social emotion) at a given binary menu $\{a, b\}$. The following result can be used to estimate, e.g., the fractions of people who enjoy a welfare gain (pride) and those who suffer a welfare loss (shame) from choosing an alternative under a public recognition program.

Claim 2. *Suppose $(\succeq, \mathcal{C})$ is a PS preference such that $\mathcal{C}(\{a, b\}) = \{b\}$.*[30] *Then,*

(i) *the DM feels pride by choosing b at $\{a, b\}$ if and only if $\{b\} \prec \{a, b\} \succ \{a\}$.*

(ii) *the DM feels shame by choosing b at $\{a, b\}$ if and only if $\{b\} \succ \{a, b\} \succ \{a\}$.*

The proof is straightforward given the PS representation, hence omitted. Intuitively, in both cases, $\mathcal{C}(\{a, b\}) = \{b\}$ and $\{a, b\} \nsim \{b\}$ imply that the reference point is $w(a)(\neq w(b))$. Then, the preference between $\{a, b\}$ and $\{b\}$ tells whether the DM feels pride or shame at $\{a, b\}$, whereas we necessarily have $\{a, b\} \succ \{a\}$ because b is a more attractive choice than a and does not affect the reference point at $\{a, b\}$.

[30] We preclude $\mathcal{C}(\{a, b\}) = \{a, b\}$, in which case the social emotion depends on the realization of the ex-post choice.

A caveat is that a norm-evoking policy may generate welfare loss even if the DM does not feel shame of her own choice. Recall the example of prosocial behavior analyzed in Section 2.1, specifically with norms (r', w) such that $r'(0) < r'(1)$ and $w(0) < w(1)$. In this case, the norms suggest prosocial behavior $x = 1$ as a reference alternative and the DM follows the norms, so she does not feel pride or shame. However, a publicity policy that forces the DM to make a choice at $\{0, 1\}$ is welfare reducing relative to the outside option of avoiding prosocial behavior in private ($\{0\} \succ \{0, 1\}$). Incorporating preferences over menus is useful for analyzing the welfare effects of norms and norm-evoking policies with a revealed preference approach, when social pressure forces the DM to choose an action that she nonetheless would like to have removed (cf. Bursztyn et al. 2023).

4.3 Examining Belief Heterogeneity and Treatment Effect

The analysis in Section 2 derives behavioral predictions based on the knowledge of (u, w, r, β). Conversely, if we observe choice data and make some assumptions on these components, we can infer some properties of the (p)reference. For example, under a PS preference, the observations $\{0, 1\} \succ \{1\}$ and $\mathcal{C}(\{0, 1\}) = \{1\}$ imply $w(0) < w(1)$ and $r(0) > r(1)$, that is, the DM perceives that 1 is normatively superior but 0 is more common. On the other hand, if we observe $\{0, 1\} \sim \{1\}$ and $\mathcal{C}(\{0, 1\}) = \{1\}$ and additionally presume $w(0) < w(1)$, then we infer $r(0) \leq r(1)$, i.e., the DM believes that 1 is (weakly) more common. Thus, the assumption that the DM has a PS preference with $w(0) < w(1)$ enables us to examine whether $r(0) > r(1)$ or not, without collecting complete data on $(\succeq, \mathcal{C})$. Similarly, we can make an inference about w by assuming that the above observations are generated by a PS preference with $r(0) > r(1)$. We can conduct these kinds of analyses to see, e.g., whether perceptions are different across demographic groups or are affected by policies.

5 Related Literature

5.1 Social Norms and Social Image Concerns

To the best of our knowledge, this paper is the first to formalize and axiomatize the notion of descriptive and prescriptive norms in a decision-theoretic model. In the social psychology literature (Cialdini et al. 1991; Schultz et al. 2007), *descriptive norms* commonly refer to norms that dictate individuals to do what is typically done by others, and *prescriptive norms* (or *injunctive norms*) commonly refer to norms which dictate them to do what people approve of. Bicchieri and Dimant (2022) define a social norm as a behavioral rule that individuals prefer to follow because they believe that (i) others follow it *and* (ii) others think the individuals should follow it. Our model is a formalization of these ideas; social norms are shaped by two functions r and w, where r represents the belief (perception) about what others do and w represents the belief about what others think should be done.[31] This modeling will facilitate the analysis of how norms are influenced by laws (Bénabou and Tirole 2011; Lane et al. 2023) or social information (Frey and Meier 2004; Schultz et al. 2007; Allcott 2011; Hallsworth et al. 2017) by altering these perceptions. Also, the DM in our model derives social payoffs/emotions by comparing her own choice with others' choice, which closely follows the literature on social comparisons (Festinger 1954).

Our paper also contributes to the growing literature on social image concerns or social pressure (Bénabou and Tirole 2006), which are considered a key driver of behavior in various contexts (See footnote 1). Our contribution is to propose a model with social image concerns which is useful for three reasons. First, our model distinguishes the descriptive and prescriptive norms in a simple manner. It captures various patterns of social norms' influences by two functions, each representing the descriptive norm and the prescriptive norm. The distinction between the two types of norms also facilitates policy analysis.

[31]Bicchieri and Dimant (2022) refer to the former belief as *empirical expectations* and the latter as *normative expectations*, and they define a social norm as a behavioral rule governed by these expectations.

For example, we can analyze the effect of altering either the descriptive norm or the prescriptive norm on the DM's prosocial behavior, holding the other norm constant. Second, our model is tractable because it does not require the researcher to compute an equilibrium, in contrast to commonly used social signaling models. Instead of restricting the DM's beliefs, our model is directly disciplined by observable choice data. This also allows the perceived norms to be biased from the objective distribution of choices or normative opinions. Third, the axioms clarify what types of choice behavior is a distinct feature of the PS model, and can be used for testing the model. The transparent link between choice and utility representation also allows the researcher to study the behavioral and welfare effects of norms and norm-evoking policies in a flexible manner. For example, as we show in Section 4, we can use the model to examine whether the DM feels pride or shame using choice data alone, without using a survey.

5.2 Axiomatic Decision Theory

Our model relates to the axiomatic two-stage models of choices with temptation (Gul and Pesendorfer 2001; Noor and Takeoka 2015) or social emotions (Dillenberger and Sadowski 2012; Saito 2015; Evren and Minardi 2017; Hashidate 2021), and axiomatic models of endogenous reference points (Ok et al. 2015; Lleras et al. 2019; Kıbrıs et al. 2023). The former papers consider decision problems in which the DM chooses a menu of alternatives in the first stage and then chooses an alternative from the menu in the second stage. In particular, those with social emotions assume that the first stage is private whereas the second stage is publicly observed, and that the DM anticipates social emotions due to public observability when making the first-stage choice. We adopt the same observability assumptions, but their domains are different from ours. Regarding the latter papers on endogenous reference, their models do not include the menu-choosing stage, and domains of ex-post choice are also different. Below, we discuss each paper in more detail.

Both Gul and Pesendorfer (2001) and Noor and Takeoka (2015), as we do, consider preferences over menus of lotteries. Our PS representation includes the GP representation as a degenerate case. Noor and Takeoka (2015) use observations of both preferences $\succeq$ over menus and choice correspondence $\mathcal{C}$ describing choices from each menu, as we do, to derive a *Menu-Dependent Self-Control (MDSC)* representation. In the MDSC representation, the self-control cost (similar to the social emotions in our model) of choosing x from A is specified by $\psi\left(\max_{y\in A} w(y)\right)\left(\max_{y\in A} w(y) - w(x)\right)$, where $\psi(\cdot) \geq 0$ is increasing. Both their and our representations generalize the GP representation by relaxing the WARP on ex-post choice. Their specification allows for a flexible scaling of the self-control cost, in contrast to our piecewise linear scaling of the "cost" of social emotions. On the other hand, their self-control cost is restricted to be non-negative, whereas our emotional cost can be negative, which is essential for generating pride.

Dillenberger and Sadowski (2012) and Evren and Minardi (2017) study preferences over menus consisting of social allocations of non-stochastic objects (e.g., dictator games). Dillenberger and Sadowski (2012) characterize shame, which involves a preference for smaller menus. By contrast, Evren and Minardi (2017) characterizes warm-glow, which involves a preference for larger menus. Our axioms capture both types of preferences, depending on whether adding alternatives to a menu expands the reference set (see Lemma 2).

Saito (2015) and Hashidate (2021) study preferences over menus consisting of social allocations $p = (p_i)_{i\in\{1\}\cup S}$ of lotteries, where 1 denotes the DM and S denotes the set of other agents. Saito (2015) derives the *generalized utilitarian* (GU) representation, which generates the pride $\beta_1 \max_{q\in A} \alpha_1(u(q_1) - u(p_1)) > 0$ of acting altruistically if the DM compromises her private payoff, and the shame $-\beta_S \max_{r\in A}(\sum_{i\in S} \alpha_i u(r_i) - \sum_{i\in S} \alpha_i u(p_i)) < 0$ of acting selfishly if the DM compromises other agents' private payoffs.[32] Hashidate (2021) generalizes the GU representation by replacing the reference point with a convex combination of the maximum and minimum (instead of maximum), which allows the signs of the social emotion terms to depend on menus and allows various social

[32] Saito (2015) allows for $\beta_1 < 0$, in which case the term captures the temptation to act selfishly.

emotions to emerge. Their models generate social emotions by the comparison of the DM's own private payoff or other agents' private payoffs relative to reference points. By contrast, pride and shame in our model arise from the comparison of the perceived normative desirability of own action against the reference alternative's desirability. We adopt our formulation to separate these emotions from private payoffs. We believe this separation is vital for two reasons. First, our model formalizes the concept of social norms discussed in Section 5.1 closely. Moreover, we argue that pride and shame should depend on the degree to which the DM can live up to social expectations (norms), not the degree to which the DM's or other agents' private payoffs are sacrificed.[33] Second, our formulation explains empirical findings discussed in Section 2 straightforwardly. For example, when information about others' behavior alters the DM's choice, it is plausibly due to changes in perceived social expectations, rather than changes in private payoffs.

We conclude this section by discussing the differences between our model and the aforementioned axiomatic models of endogenous reference dependence.[34] Ok et al. (2015) characterize choice behavior which exhibits the "attraction effect." Specifically, by relaxing the WARP, they derive a model in which a dominated alternative serves as a reference alternative and restricts the choice set to alternatives that dominate it. Therefore, in their model, the reference point is an unchosen alternative, and imposes a mental constraint on choice sets. By contrast, in our model, the reference alternative may be chosen, and it affects the preference but not choice sets. Lleras et al. (2019) consider a preference over state-contingent contracts (acts) and derive a representation that evaluates an act based on its expected value and expected gain/loss relative to the expected value. Their representation allows the DM to derive payoffs from either expected gain or loss, but not both. By contrast, the DM in our model may feel pride from an alternative and shame from an-

[33]Scheff (1988) discusses how perceived social expectations set a system in which conformity to norms is sustained by the reward of pride and punishment of shame (together with deference).

[34]Köszegi and Rabin (2006) develop a non-axiomatic model of endogenous reference formation. In their model, the agent's reference point is constrained by rational expectations.

other. Kıbrıs et al. (2023) consider choices from finite choice sets, and derive a reference-dependent choice model in which the reference point is determined by an endogenously derived conspicuity ranking, just as an endogenously derived r defines reference in our model. Their representation is quite general, but they do not characterize a specific structure of the reference point in our model, so our axiomatization result is not implied by their work.

6 Conclusion

Despite the growing interest in using social norms for behavioral change, their behavioral and welfare effects are not well understood. We propose and axiomatize a model of reference-dependent decision-making in which the decision maker's perception of others' choice ("descriptive norm") and her perception of others' normative opinions ("prescriptive norm") together shape a reference point that indicates the approvedness of the socially prevalent behavior. The key drivers of the decision maker's behavior are social emotions, such as a positive payoff from pride, which she enjoys if her choice exceeds the reference point, and a negative payoff from shame, which she suffers if her choice falls short of it. A simple model of prosocial behavior can provide various useful implications, such as when policies such as public recognition programs or norm nudges likely induce prosocial behavior, or how policies differentially influence the choice of an action conditional on participating in an opportunity for prosocial behavior and the decision on participation. It also rationalizes empirical findings that aligned descriptive and prescriptive norms are more effective for inducing prosocial behavior compared to misaligned ones, and that the descriptive norms have a larger impact than the prescriptive norms when they are misaligned. Moreover, the axiomatic model is simple enough to clarify the mechanisms behind policy effects and facilitate policy analysis, tractable because it does not impose an equilibrium assumption, and transparent in its relation to observed choice, which may usefully guide empirical analysis.

There are two limitations to this paper. First, we do not model how individuals' perceptions are shaped. For example, they may arise as equilibrium objects of a game (cf. Bénabou and Tirole 2006), and such an equilibrium restriction may be necessary to study how norms evolve over time. Also, individuals may form perceptions in a self-serving manner (Heinicke et al. 2022; Bicchieri et al. 2023), and accounting for this possibility may be crucial for predicting the effect of an informational intervention. Second, norms may be specified in a more flexible manner. For example, an individual may compare her behavior with the behavior of a group of individuals rather than that of a single "typical person." This can be modeled by specifying the reference point to depend on the distribution of descriptive norms of other individuals.[35] Eliciting information on individuals' reference groups based on their choice is an interesting topic for empirical research.

[35]Technically, this will yield a random Strotz representation (Dekel and Lipman 2012).

References

Akerlof, George A. and Rachel E. Kranton, "Economics and Identity," *The Quarterly Journal of Economics*, 08 2000, *115* (3), 715–753.

Allcott, Hunt, "Social Norms and Energy Conservation," *Journal of Public Economics*, 2011, *95* (9-10), 1082–1095.

______ **and Todd Rogers**, "The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation," *American Economic Review*, 2014, *104* (10), 3003–3037.

Andreoni, James and B. Douglas Bernheim, "Social Image and the 5050 Norm: A Theoretical and Experimental Analysis of Audience Effects," *Econometrica*, 2009, *77* (5), 1607–1636.

______ **, Justin M Rao, and Hannah Trachtman**, "Avoiding the Ask: A Field Experiment on Altruism, Empathy, and Charitable giving," *Journal of Political Economy*, 2017, *125* (3), 625–653.

Bénabou, Roland and Jean Tirole, "Incentives and Prosocial Behavior," *American Economic Review*, December 2006, *96* (5), 1652–1678.

______ **and** ______ , "Laws and Norms," Working Paper 17579, National Bureau of Economic Research November 2011.

Bicchieri, Cristina, *The Grammar of Society: The Nature and Dynamics of Social norms*, Cambridge University Press, 2005.

______ **and Erte Xiao**, "Do the Right Thing: But Only if Others Do So," *Journal of Behavioral Decision Making*, 2009, *22* (2), 191–208.

______ **and Eugen Dimant**, "Nudging with Care: The Risks and Benefits of Social Information," *Public Choice*, 2022, *191* (3-4), 443–464.

______ **,** ______ **, and Silvia Sonderegger**, "It's not a lie if you believe the norm does not apply: Conditional norm-following and belief distortion," *Games and Economic Behavior*, 2023, *138*, 321–354.

Birke, David J, "Anti-Bunching: A New Test for Signaling Motives in Prosocial Behavior," 2020.

Broberg, Tomas, Tore Ellingsen, and Magnus Johannesson, "Is Generosity Involuntary?," *Economics Letters*, 2007, *94* (1), 32–37.

Bursztyn, Leonardo, Alessandra L González, and David Yanagizawa-Drott, "Misperceived Social Norms: Women Working Outside the Home in Saudi Arabia," *American Economic Review*, 2020, *110* (10), 2997–3029.

——— **and Robert Jensen**, " How Does Peer Pressure Affect Educational Investments? ," *The Quarterly Journal of Economics*, 05 2015, *130* (3), 1329–1367.

——— **and** ——— , "Social Image and Economic Behavior in the Field: Identifying, Understanding, and Shaping Social Pressure," *Annual Review of Economics*, 2017, *9* (1), 131–153.

——— **, Benjamin R Handel, Rafael Jimenez, and Christopher Roth**, "When Product Markets Become Collective Traps: The Case of Social Media," Working Paper 31771, National Bureau of Economic Research October 2023.

——— **, Thomas Fujiwara, and Amanda Pallais**, "'Acting Wife': Marriage Market Incentives and Labor Market Investments," *American Economic Review*, November 2017, *107* (11), 3288–3319.

Butera, Luigi, Robert Metcalfe, William Morrison, and Dmitry Taubinsky, "Measuring the Welfare Effects of Shame and Pride," *American Economic Review*, 2022, *112* (1), 122–168.

Cialdini, Robert B., "Crafting Normative Messages to Protect the Environment," *Current Directions in Psychological Science*, 2003, *12* (4), 105–109.

Cialdini, Robert B, Carl A Kallgren, and Raymond R Reno, "A Focus Theory of Normative Conduct: A Theoretical Refinement and Reevaluation of the Role of Norms in Human Behavior," in "Advances in Experimental Social Psychology," Vol. 24, Elsevier, 1991, pp. 201–234.

Dana, Jason, Daylian M Cain, and Robyn M Dawes, "What You Dont Know Wont Hurt Me: Costly (but Quiet) Exit in Dictator Games," *Organizational Behavior and Human Decision Processes*, 2006, *100* (2), 193–201.

Dekel, Eddie and Barton L. Lipman, "Costly Self-Control and Random Self-Indulgence," *Econometrica*, 2012, *80* (3), 1271–1302.

DellaVigna, Stefano, John A List, and Ulrike Malmendier, "Testing for Altruism and Social Pressure in Charitable Giving," *The Quarterly Journal of Economics*, 2012, *127* (1), 1–56.

______ **, John A. List, Ulrike Malmendier, and Gautam Rao**, "Voting to Tell Others," *The Review of Economic Studies*, 10 2017, *84* (1), 143–181.

Dillenberger, David and Philipp Sadowski, "Ashamed to be Selfish," *Theoretical Economics*, 2012, *7* (1), 99–124.

Evren, Özgür and Stefania Minardi, "WarmGlow Giving and Freedom to be Selfish," *The Economic Journal*, 02 2017, *127* (603), 1381–1409.

Festinger, Leon, "A Theory of Social Comparison Processes," *Human Relations*, 1954, *7* (2), 117–140.

Frey, Bruno S and Benno Torgler, "Tax Morale and Conditional Cooperation," *Journal of Comparative Economics*, 2007, *35* (1), 136–159.

______ **and Stephan Meier**, "Social Comparisons and Pro-social Behavior: Testing Conditional Cooperation in a Field Experiment," *American Economic Review*, 2004, *94* (5), 1717–1722.

Fudenberg, Drew and David K. Levine, "Fairness, Risk preferences and Independence: Impossibility Theorems," *Journal of Economic Behavior & Organization*, 2012, *81* (2), 606–612.

Gerber, Alan S., Donald P. Green, and Christopher W. Larimer, "Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment," *American Political Science Review*, 2008, *102* (1), 3348.

Gul, Faruk and Wolfgang Pesendorfer, "Temptation and Self-Control," *Econometrica*, 2001, *69* (6), 1403–1435.

Hallsworth, Michael, John A List, Robert D Metcalfe, and Ivo Vlaev, "The Behavioralist as Tax Collector: Using Natural Field Experiments to enhance tax compliance," *Journal of Public Economics*, 2017, *148*, 14–31.

Hashidate, Yosuke, "Social Image Concern and Reference Point Formation," *The B.E. Journal of Theoretical Economics*, 2021, *21* (1), 19–84.

Heinicke, Franziska, Christian König-Kersting, and Robert Schmidt, "Injunctive vs. Descriptive Social norms and Reference Group Dependence," *Journal of Economic Behavior & Organization*, 2022, *195*, 199–218.

Karing, Anne, "Social Signaling and Childhood Immunization: A Field Experiment in Sierra Leone," 2024.

Kıbrıs, Özgür, Yusufcan Masatlioglu, and Elchin Suleymanov, "A Theory of Reference Point Formation," *Economic Theory*, 2023, *75* (1), 137–166.

Klinowski, David, "Reluctant Donors and Their Reactions to Social Information," *Experimental Economics*, 2021, *24* (2), 515–535.

Köszegi, Botond and Matthew Rabin, "A Model of Reference-Dependent Preferences," *The Quarterly Journal of Economics*, 11 2006, *121* (4), 1133–1165.

Lacetera, Nicola and Mario Macis, "Social image concerns and prosocial behavior: Field evidence from a nonlinear incentive scheme," *Journal of Economic Behavior & Organization*, 2010, *76* (2), 225 – 237.

Lane, Tom, Daniele Nosenzo, and Silvia Sonderegger, "Law and Norms: Empirical Evidence," *American Economic Review*, May 2023, *113* (5), 1255–93.

Lazear, Edward P, Ulrike Malmendier, and Roberto A Weber, "Sorting in experiments with application to social preferences," *American Economic Journal: Applied Economics*, 2012, *4* (1), 136–163.

Lleras, Juan Sebastián, Evan Piermont, and Richard Svoboda, "Asymmetric Gain–Loss Reference Dependence and Attitudes toward Uncertainty," *Economic Theory*, 2019, *68* (3), 669–699.

Maccheroni, Fabio, Massimo Marinacci, and Aldo Rustichini, "Social Decision Theory: Choosing within and between Groups," *The Review of Economic Studies*, 01 2012, *79* (4), 1591–1636.

Masatlioglu, Yusufcan, Daisuke Nakajima, and Erkut Y. Ozbay, "Revealed Attention," *American Economic Review*, May 2012, *102* (5), 2183–2205.

Miller, Dale T. and Deborah A. Prentice, "Collective Errors and Errors about the Collective," *Personality and Social Psychology Bulletin*, 1994, *20* (5), 541–550.

Noor, Jawwad and Norio Takeoka, "Menu-dependent Self-control," *Journal of Mathematical Economics*, 2015, *61*, 1–20.

Ok, Efe A., Pietro Ortoleva, and Gil Riella, "Revealed (P)Reference Theory," *American Economic Review*, January 2015, *105* (1), 299–321.

Perez-Truglia, Ricardo and Ugo Troiano, "Shaming Tax Delinquents," *Journal of Public Economics*, 2018, *167*, 120–137.

Rabin, Matthew, "Moral Preferences, Moral Constraints, and Self-serving Biases," 1995.

Saito, Kota, "Impure Altruism and Impure Selfishness," *Journal of Economic Theory*, 2015, *158*, 336–370.

Scheff, Thomas J., "Shame and Conformity: The Deference-Emotion System," *American Sociological Review*, 1988, *53* (3), 395–406.

Schultz, P. Wesley, Jessica M. Nolan, Robert B. Cialdini, Noah J. Goldstein, and Vladas Griskevicius, "The Constructive, Destructive, and Reconstructive Power of Social Norms," *Psychological Science*, 2007, *18* (5), 429–434.

Strotz, Robert H., "Myopia and Inconsistency in Dynamic Utility Maximization," *Review of Economic Studies*, 1955, *23* (3), 165–180.

Toussaert, Séverine, "Eliciting Temptation and Self-control through Menu Choices: A Lab Experiment," *Econometrica*, 2018, *86* (3), 859–889.

Tyran, Jean-Robert and Lars P Feld, "Achieving Compliance when Legal Sanctions are Non-deterrent," *Scandinavian Journal of Economics*, 2006, *108* (1), 135–156.

For Online Publication

A Proofs

Below, we denote a mixed lottery $\alpha a + (1-\alpha)b \in \Delta$ by $a\alpha b$ and a mixed menu $\alpha A + (1-\alpha)B \in \mathcal{A}$ by $A\alpha B$, for any $a, b \in \Delta$, $A, B \in \mathcal{A}$, and $\alpha \in [0,1]$.

A.1 Proof of Theorem 1

A.1.1 Proof of Lemma 1

We first introduce two lemmas.

Lemma 5. *Suppose Axioms 4 and 5 hold. Then, for any $a, b, c \in \Delta$ and $\alpha \in (0,1)$, $a \succ^* b$ implies $a\alpha c \succ^* b\alpha c$.*

Proof. Suppose $a \succ^* b$. By definition, there exists $A \ni b$ such that $A \cup \{a\} \succ A$ and $a \notin \mathcal{C}(A \cup \{a\})$. Take any $c \in \Delta$ and any $\alpha \in (0,1)$. By Axiom 4(iii), $[A \cup \{a\}]\alpha\{c\} \succ A\alpha\{c\}$. Also, by Axiom 5(iii), $\mathcal{C}([A \cup \{a\}]\alpha\{c\}) = \mathcal{C}(A \cup \{a\})\alpha\{c\}$, so $a\alpha c \notin \mathcal{C}([A \cup \{a\}]\alpha\{c\})$. Thus, $a\alpha c \succ^* b\alpha c$. □

Lemma 6. *Suppose Axioms 1-3 hold. Then, the following statements hold.*

(i) If $a \succ^* b$, then for any $c \in \Delta$, there exists $\alpha \in (0,1)$ such that $a\alpha c \succ^* b$.

(ii) If $b \succ^* c$, then for any $a \in \Delta$, there exists $\beta \in (0,1)$ such that $b \succ^* a\beta c$.

Proof. (i) By $a \succ^* b$, we have some $A \ni b$ such that $A \cup \{a\} \succ A$ and $a \notin \mathcal{C}(A \cup \{a\})$. By Axiom 2(i), there exists $\alpha_1 \in (0,1)$ such that $A \cup \{a\alpha c\} \succ A$ for all $\alpha \in (\alpha_1, 1)$. Also, by Axiom 3(iii-b), we have some $\alpha_2 \in (0,1)$ such that $a\alpha c \notin \mathcal{C}(A \cup \{a\alpha c\})$ for all $\alpha \in (\alpha_2, 1)$.[36] Therefore, taking $\alpha \in (\max\{\alpha_1, \alpha_2\}, 1)$, we have $a\alpha c \succ^* b$.

(ii) By $b \succ^* c$, there exists some $B \ni c$ such that $B \cup \{b\} \succ B$ and $b \notin \mathcal{C}(B \cup \{b\})$. By Axiom 3(iii), there exists some $\beta \in (0,1)$ such that $[\{a\}\beta B] \cup \{b\} \succ \{a\}\beta B$ and $b \notin \mathcal{C}([\{a\}\beta B] \cup \{b\})$.[37] Thus, $b \succ^* a\beta c$. □

---

[36] Suppose not. Then, there is a sequence $\{\alpha^n\}_n$ such that $\alpha^n \to 1$ and $a\alpha^n c \in \mathcal{C}(A \cup \{a\alpha^n c\})$. Since $a \succ^* a'$ for all $a' \in A$, Axiom 3(iii-b) implies $a \in \mathcal{C}(A \cup \{a\})$, a contradiction.

[37] Following the argument in footnote 36, the latter property holds for all sufficiently small $\beta > 0$. To show that the former property holds for all sufficiently small $\beta > 0$, note first that Axioms 2(i) and 3(iii-a) ensure $A \cup \{a\} \succ \tilde{A} \succ A$ where $\tilde{A} = \tilde{A}(\gamma) = [A \cup \{a\}]\gamma A$ for some $\gamma \in (0,1)$. (Otherwise, $\Gamma^L = \{\gamma \in [0,1] : A \succeq \tilde{A}(\gamma)\}$ and $\Gamma^U = \{\gamma \in [0,1] : \tilde{A}(\gamma) \succeq A \cup \{a\}\}$ are nonempty closed sets such that $\Gamma^L \cup \Gamma^U = [0,1]$, so $\tilde{A}(\gamma) \succ \tilde{A}(\gamma)$ for $\gamma \in \Gamma^L \cap \Gamma^U$, a contradiction.) Then, for all sufficiently small β, we must have $[\{a\}\beta B] \cup \{b\} \succ \tilde{A} \succ \{a\}\beta B$.

Proof of Lemma 1. We prove the following properties.
Completeness. Immediate from the definition of $\succeq_r$.

Transitivity. We first show that $\succ_r$ is transitive. Suppose $a \succ_r b \succ_r c$. If $a \succ^* b \succ^* c$, then $a \succ^* c$ by Axiom 3(ii), so $a \succ_r c$. Next, suppose $a \succ^* b$, $d \not\succ^* b$, and $d \succ^* c$ for some $d \in \text{int}(\Delta)$. If $d \succ^* a$, then $d \succ^* b$ by Axiom 3(ii), a contradiction. Therefore, $d \not\succ^* a$, hence $a \succ_r c$. Next, suppose $d \not\succ^* a$, $d \succ^* b$, and $b \succ^* c$ for some $d \in \text{int}(\Delta)$. Then $d \succ^* c$ by Axiom 3(ii), so $a \succ_r c$. Finally, suppose $d \not\succ^* a$, $d \succ^* b$, $e \not\succ^* b$, and $e \succ^* c$ for some $d, e \in \text{int}(\Delta)$. If $d \not\succ^* c$, then $c \succ_r b$, contradicting Axiom 3(i). Therefore, $d \succ^* c$, hence $a \succ_r c$.

Now, suppose $a \succeq_r b \succeq_r c$. If $a \succ_r b \succ_r c$, we have $a \succ_r c$ by transitivity. Next, suppose $a \succ_r b \sim_r c$. Then, for any $d \in \text{int}(\Delta)$, either $d \not\succ^* b, c$ or $d \succ^* b, c$ must hold (otherwise, $b \nsim_r c$). Now, by $a \succ_r b$, there exists $d \in \text{int}(\Delta)$ such that $d \not\succ^* a$ and $d \succ^* b$.[38] Therefore, $d \succ^* c$, hence $a \succ_r c$. Similarly, if $a \sim_r b \succ_r c$, then there exists $d \in \text{int}(\Delta)$ such that $d \not\succ^* b$ and $d \succ^* c$, and we must have $d \not\succ^* a$ to maintain $a \sim_r b$. Thus, $a \succ_r c$. Finally, suppose $a \sim_r b \sim_r c$. If $a \not\succeq_r c$, then we have $c \succ_r a \sim_r b$, so the above argument yields $c \succ_r b$, a contradiction. Thus, $a \succeq_r c$.

Independence. Suppose $a \succ_r b$. If $a \succ^* b$, then we have $a\alpha c \succ^* b\alpha c$ by Lemma 5, hence $a\alpha c \succ_r b\alpha c$. Next, suppose there exists some $d \in \text{int}(\Delta)$ such that $d \not\succ^* a$ and $d \succ^* b$. Then Lemma 5 yields $d\alpha c \succ^* b\alpha c$ and Axiom 3(iv) yields $d\alpha c \not\succ^* a\alpha c$, so $a\alpha c \succ_r b\alpha c$.

Archimedeanity. Suppose $a \succ_r b \succ_r c$. If $a \succ^* b \succ^* c$, Lemma 6 yields $a\alpha c \succ_r b \succ_r a\beta c$ for some $\alpha, \beta \in (0,1)$. Next, suppose $a \succ^* b$, $d \not\succ^* b$, and $d \succ^* c$ for some $d \in \text{int}(\Delta)$. By Lemma 6, we have $a\alpha c \succ_r b$ and $d \succ^* a\beta c$, hence $b \succ_r a\beta c$, for some $\alpha, \beta \in (0,1)$. Next, suppose $b \succ^* c$, $d \not\succ^* a$, and $d \succ^* b$ for some $d \in \text{int}(\Delta)$. We then have $b \succ_r a\beta c$ and $d\alpha c \succ^* b$ for some $\alpha, \beta \in (0,1)$. Also, by Axiom 3(iv), we have $d\alpha c \not\succ^* a\alpha c$. Because $d\alpha c \in \text{int}(\Delta)$, we have $a\alpha c \succ_r b$. Finally, suppose there exist some $d, e \in \text{int}(\Delta)$ such that $d \not\succ^* a$, $d \succ^* b$, $e \not\succ^* b$, and $e \succ^* c$. By Axiom 3(iv) and Lemma 6, we have $d\alpha c \not\succ^* a\alpha c$, $d\alpha c \succ^* b$ and $e \succ^* a\beta c$ for some $\alpha, \beta \in (0,1)$. Therefore, $a\alpha c \succ_r b$ and $b \succ_r a\beta c$.

[38] If $a \succ^* b$, then $a \succ^* a0.5b \succ^* b$ by Lemma5. Then, taking an arbitrary $e \in \text{int}(\Delta)$ and $\alpha \in (0,1)$ sufficiently close to one, Lemma 6 implies $d \equiv (a0.5b)\alpha e$ satisfies the condition.

Thus, by the Mixture Space Theorem, $\succeq_r$ admits a linear representation r, and the representation is unique up to positive affine transformation. $\square$

A.1.2 Proof of Lemma 2

The proof is trivial when $A \supset B$, so suppose otherwise. For (ii), we also assume $A \not\subset B$; otherwise the proof is trivial. We first introduce two lemmas on $\succeq_r$ and $\succeq_w$. Note we impose Axioms 1-5 in the lemmas, so that $\succeq_r$ admits a linear representation r.

Lemma 7. *Suppose Axioms 1-5 hold. Then, for any finite $A \in \mathcal{A}$, there exists $a^* \in A$ such that $a^* \succeq_r a$ and $a^* \succeq_w a$ for all $a \in A$.*

Proof. Because $\succeq_w$ is complete by definition and transitive on $\varphi_r(A)$ by Axiom 3(ii), there exists a^* which maximizes $\succeq_w$ on $\varphi_r(A)$. By $a^* \in \varphi_r(A)$, we must have $a^* \succeq_r a$ for all $a \in A$. Also, for any $a \in A \setminus \varphi_r(A)$, we have $a^* \succ_r a$, so Axiom 3(i) implies $a^* \succeq_w a$. Thus, $a^* \succeq_r a$ and $a^* \succeq_w a$ for all $a \in A$. $\square$

Lemma 8. *Suppose that Axioms 1-6 hold and that A and B are finite.*

(i) If $A \succ A \cup B$, then there exists $b \in B \setminus A$ such that $b \succ_r a$ or $b \succ_w a$ for all $a \in A$.

(ii) If $A \cup B \succ A$ and $\mathcal{C}(A \cup B) \cap A \neq \emptyset$, then there exists $b \in B \setminus A$ such that $b \succ_r a$ for all $a \in A$.

Proof. (i) By Lemma 7, there exists $a^* \in A$ such that $a^* \succeq_r a$ and $a^* \succeq_w a$ for all $a \in A$. To prove the contrapositive of the statement, suppose that for any $b \in B \setminus A$, there exists $a \in A$ such that $a \succeq_r b$ and $a \succeq_w b$. By the transitivity of $\succeq_r$ and Axiom 3(i)(ii), $a^* \succeq_r c$ and $a^* \succeq_w c$ for all $c \in A \cup B$. Therefore, Axiom 6(i) yields $A \cup B \succeq A$.

(ii) Similarly, if the conclusion is false, then we have some $a^* \in A$ such that $a^* \succeq_r c$ for all $c \in A \cup B$ and $a^* \succeq_w c$ for all $c \in A$. If $a^* \succeq_w b$ for all $b \in B$, then by Axiom 6(i), it is impossible to have $A \cup B \succ A$ and $\mathcal{C}(A \cup B) \cap A \neq \emptyset$ simultaneously. If there exists $b^* \in B$ such that $b \succ_w a^*$, then Axiom 6(ii) implies we cannot have both $A \cup B \succ A$ and $\mathcal{C}(A \cup B) \cap A \neq \emptyset$. Thus, the claim holds. $\square$

Proof of Lemma 2 (Continued).

(i) If $A \succ A \cup B$, then by Lemma 8(i), there exists $b \in B \setminus A$ such that $b \succ_r a$ or $b \succ_w a$ for all $a \in A$. By Axiom 3(i), we have $b \succeq_r a$, hence $r(b) \geq r(a)$, for all $a \in A$. Thus, $\varphi_r(A \cup B) \neq \varphi_r(A)$.

(ii) Suppose $\varphi_r(A \cup B) = \varphi_r(A) \cup \varphi_r(B)$ and $A \succeq B$. If $A \succeq B \succ A \cup B$, then by Lemma 8(i) there exists some $a^* \in A \setminus B$ such that $a^* \succ_r b$ or $a^* \succ_w b$ for all $b \in B$ and some $b^* \in B \setminus A$ such that $b^* \succ_r a$ or $b^* \succ_w a$ for all $a \in A$, which contradicts Axiom 3(i). Next, suppose $A \cup B \succ A \succeq B$. Note at least one of $\mathcal{C}(A \cup B) \cap A$ or $\mathcal{C}(A \cup B) \cap B$ is nonempty. If $\mathcal{C}(A \cup B) \cap A \neq \emptyset$, then by Lemma 8(ii) there exists $b \in B \setminus A$ such that $b \succ_r a$, hence $r(b) > r(a)$, for all $a \in A$, which contradicts $\varphi_r(A \cup B) = \varphi_r(A) \cup \varphi_r(B)$. A similar contradiction results if $\mathcal{C}(A \cup B) \cap B \neq \emptyset$. Thus, $A \succeq A \cup B \succeq B$. $\square$

A.1.3 Proof of Lemma 3.

For any $a \in A$, pick $b_a \in \varphi_r(A)$ such that $\{a, b\} \succeq \{a, b_a\}$ for all $b \in \varphi_r(A)$. Let $a^* \in A$ be such that $\{a^*, b_{a^*}\} \succeq \{a, b_a\}$ for all $a \in A$ and let $b^* \equiv b_{a^*}$. Then, iteratively applying Lemma 2(ii), $\{a^*\} \cup \varphi_r(A) = \cup_{b \in \varphi_r(A)} \{a^*, b\} \succeq \{a^*, b^*\}$. Further, because $A = (\{a^*\} \cup \varphi_r(A)) \cup (A \setminus (\{a^*\} \cup \varphi_r(A)))$ and $\varphi_r(A) = \varphi_r(\{a^*\} \cup \varphi_r(A))$, applying Lemma 2(i) yields $A \succeq \{a^*\} \cup \varphi_r(A) \succeq \{a^*, b^*\}$. Secondly, by construction, $\{a^*, b^*\} \succeq \{a, b_a\}$ for all $a \in A$. Then iteratively applying Lemma 2(ii) yields $\{a^*, b^*\} \succeq \cup_{a \in A} \{a, b_a\} = A$. $\square$

A.1.4 Proof of Lemma 4.

We first introduce a lemma.

Lemma 9. *Suppose Axioms 3 and 6 hold. Then, $b \in \mathcal{N}_2(a)$ implies $\{a, b\} \sim \{a\}$.*

Proof. Suppose $a \succ_r b$ and $a \in \mathcal{C}(\{a, b\})$. By Axiom 3(i), we have $a \succeq_w b$. Thus, by Axiom 6(i), we have $\{a, b\} \sim \{a\}$. $\square$

Now, let

$$\mathcal{A}_P = \left\{ A \in \mathcal{A} : A = \sum_{m=1}^{M} \alpha_m A_m, A_m \in \mathcal{B}_P, M < \infty \right\}$$

and

$$\mathcal{A}_S = \left\{ A \in \mathcal{A} : A = \sum_{m=1}^{M} \alpha_m A_m, A_m \in \mathcal{B}_S, M < \infty \right\}$$

denote the set of finite mixtures over $\mathcal{B}_P$ and that of finite mixtures over $\mathcal{B}_S$, respectively. Also, for notational simplicity, define $\mathcal{A}_N = \mathcal{B}_N$. We show that $\succeq$ admits a linear representation on each domain $\mathcal{A}_j$, $j = P, S, N$.

Lemma 10. *Suppose Axioms 1-6 hold. Then, for each $j \in \{P, S, N\}$, $\succeq$ restricted to $\mathcal{A}_j$ has a representation V^j. Moreover, (i) V^P and V^S are linear and (ii) the restrictions of V^P, V^S and V^N to singleton sets are continuous.*

Proof. The statements regarding V^P and V^S follow from Lemma 1 of GP. To construct V^N, note that by Lemma 9, for each $A \in \mathcal{A}_N$, we have $A \sim \left\{a^A\right\}$ for some known $a^A \in A$. By letting $V^N(A) = V^P(\left\{a^A\right\})$, V^N represents $\succeq$ on $\mathcal{A}_N$ and its restriction to singleton sets is continuous. □

By construction, for any $j, k \in \{P, S, N\}$, $V^j(\{a\}) \geq V^j(\{b\}) \Leftrightarrow \{a\} \succeq \{b\} \Leftrightarrow V^k(\{a\}) \geq V^k(\{b\})$. Because linear representation of $\succeq$ on singleton sets is unique up to positive affine transformation, we can normalize them so that $V^j(\{a\}) = V^k(\{a\}) \equiv V^{\text{singleton}}(\{a\})$ for all a and all j, k.

To obtain a desired representation of $\succeq$ across different domains, we introduce two lemmas. The second lemma relates preferences over $\mathcal{A}_S$ and $\mathcal{A}_N$ to those for singleton menus.

Lemma 11. *Suppose Axioms 1-6 hold. If $\{c, d\} \in \mathcal{B}_S$, then $\{c\} \succeq \{c, d\}$ or $\{d\} \succeq \{c, d\}$ holds.*

Proof. The conclusion trivially holds if $c = d$, so we assume $c \neq d$. Without loss of generality, let $c \succ_r d$. If $d \in \mathcal{S}(c) \cup \mathcal{N}_1(c)$, then $\{d\} \succeq \{c, d\}$ by definition. If $d \in \mathcal{I}(c)$, then Lemma 2(ii) implies either $\{c\} \succeq \{c, d\}$ or $\{d\} \succeq \{c, d\}$. □

Lemma 12. *Suppose Axioms 1-6 hold. If $A \in \mathcal{A}_S \cup \mathcal{A}_N$, then there exists some $e \in \Delta$ such that $A \sim \{e\}$.*

Proof. If $A \in \mathcal{A}_N$, then the conclusion follows from Lemma 9. Suppose $A \in \mathcal{A}_S$. We first note that iteratively applying Lemma 2 yields $A \succeq \{a\}$ for some $a \in A$.[39] If $A \sim \{a\}$ for some $a \in A$, the conclusion holds. Next, suppose $\{a'\} \succ A \succ \{a\}$ for some $a, a' \in A$. Then, because $\{a\}, \{a'\}, A \in \mathcal{A}_S$, we have $V^S(\{a'\}) > V^S(A) > V^S(\{a\})$. By linearity, there exists $\alpha \in (0, 1)$ such that

[39] Denote $A = \left\{a_1, \cdots, a_{|A|}\right\}$ where $\{a_1\} \succeq \{a_2\} \succeq \cdots \succeq \left\{a_{|A|}\right\}$. If $a_{|A|-1} \succ_r a_{|A|}$, then Lemma 2(i) implies $\left\{a_{|A|-1}, a_{|A|}\right\} \succeq \left\{a_{|A|-1}\right\} \succeq \left\{a_{|A|}\right\}$. If $a_{|A|} \succ_r a_{|A|-1}$, then Lemma 2(i) implies $\left\{a_{|A|-1}, a_{|A|}\right\} \succeq \left\{a_{|A|}\right\}$. If $a_{|A|-1} \sim_r a_{|A|}$, then Lemma 2(ii) implies $\left\{a_{|A|-1}, a_{|A|}\right\} \succeq \left\{a_{|A|}\right\}$. Next, if $a_{|A|-2} \succ_r a_{|A|-1}, a_{|A|}$, then Lemma 2(i) implies $\left\{a_{|A|-2}, a_{|A|-1}, a_{|A|}\right\} \succeq \left\{a_{|A|-2}\right\} \succeq \left\{a_{|A|}\right\}$. If $a_{|A|-1} \succ_r a_{|A|-2}$ or $a_{|A|} \succ_r a_{|A|-2}$, then Lemma 2(i) implies $\left\{a_{|A|-2}, a_{|A|-1}, a_{|A|}\right\} \succeq \left\{a_{|A|-1}, a_{|A|}\right\} \succeq \left\{a_{|A|}\right\}$. If $a_{|A|-2} \sim_r a_{|A|-1} \succeq a_{|A|}$ or $a_{|A|-2} \sim_r a_{|A|} \succeq a_{|A|-1}$, then Lemma 2(ii) implies either $\left\{a_{|A|-2}, a_{|A|-1}, a_{|A|}\right\} \succeq \left\{a_{|A|-1}, a_{|A|}\right\} \succeq \left\{a_{|A|}\right\}$ or $\left\{a_{|A|-2}, a_{|A|-1}, a_{|A|}\right\} \succeq \left\{a_{|A|-2}\right\} \succeq \left\{a_{|A|}\right\}$. Repeating this finite times, we obtain $A \succeq \left\{a_{|A|}\right\}$.

$V^S(\{a'\alpha a\}) = \alpha V^S(\{a'\}) + (1-\alpha)V^S(\{a\}) = V^S(A)$. Because V^S represents $\succeq$ on $\mathcal{A}_S$ and $\{a'\alpha a\} \in \mathcal{A}_S$, we have $\{a'\alpha a\} \sim A$. Finally, to see that $A \succ \{a\}$ for all $a \in A$ does not occur, recall we can write $A = \sum_{m=1}^{M_A} \alpha_m \{a_{1m}, a_{2m}\}$ where $\{a_{1m}, a_{2m}\} \in \mathcal{B}_S$ and $\sum_{m=1}^{M_A} \alpha_m = 1$. By Lemma 11, there exist $(e_m)_{m=1}^{M_A}$, with $e_m \in \{a_{1m}, a_{2m}\}$ for each m, such that $\{e_m\} \succeq \{a_{1m}, a_{2m}\}$. By Axiom 4, we have $\{\sum_{m=1}^{M_A} \alpha_m e_m\} \succeq A$. □

We now obtain the desired representation of $\succeq$ on $\mathcal{A}_P \cup \mathcal{A}_S \cup \mathcal{A}_N$, which contains all binary menus. For notational simplicity, in Lemma 13 below we eliminate from $\mathcal{A}_P$ menus which are contained in $\mathcal{A}_S$ (i.e., $\mathcal{A}_P$ denotes $\mathcal{A}_P \setminus \mathcal{A}_S$).[40]

Lemma 13. *Suppose Axioms 1-6 hold. Define $V : \mathcal{A}_P \cup \mathcal{A}_S \cup \mathcal{A}_N \to \mathbb{R}$ as follows:*

$$V(A) = \sum_{j \in \{P,S,N\}} V^j(A) \times I\{A \in \mathcal{A}_j \text{ and } |A| > 1\} + V^{\text{singleton}}(A) \times I\{|A| = 1\}.$$

Then, V represents $\succeq$ on $\mathcal{A}_P \cup \mathcal{A}_S \cup \mathcal{A}_N$. Moreover, $A, B \in \mathcal{B}_P$ or $A, B \in \mathcal{B}_S$ implies $V(A\alpha B) = \alpha V(A) + (1-\alpha)V(B)$.

Proof. Note that any $A \in \mathcal{A}_P \cup \mathcal{A}_S \cup \mathcal{A}_N$ belongs to exactly one of $\{B \in \mathcal{A}_j : |B| > 1\}$, $j = P, S, N$, or $\{B : |B| = 1\}$ (recall the re-definition of $\mathcal{A}_P$). Suppose $A \succ B$ and let $A \in \mathcal{A}_j$ and $B \in \mathcal{A}_k$. If $|A| = 1$ and $|B| = 1$, then $V(A) = V^{\text{singleton}}(A) > V^{\text{singleton}}(B) = V(B)$ where the inequality follows from $V^{\text{singleton}}$ representing $\succeq$ over singletons. If $|A| = 1$ and $|B| > 1$, then $V(A) = V^{\text{singleton}}(A) = V^k(A) > V^k(B) = V(B)$ where the inequality follows from V^k representing $\succeq$ on $\mathcal{A}_k$ and $A, B \in \mathcal{A}_k$. Similarly, we obtain $V(A) > V(B)$ if $|A| > 1$ and $|B| = 1$. Finally, suppose $|A| > 1$ and $|B| > 1$. If $j = k = P$, then $V(A) = V^P(A) > V^P(B) = V(B)$. Otherwise, by Lemma 12, there exists some $e \in \Delta$ such that $A \sim \{e\}$ or $B \sim \{e\}$. For the former case,

$$V(A) = V^j(A) = V^j(\{e\}) = V^k(\{e\}) > V^k(B) = V(B),$$

thus $V(A) > V(B)$. Proof for the latter case is analogous. Thus, $A \succ B \Rightarrow V(A) > V(B)$ holds. Similarly, $B \succeq A \Rightarrow V(B) \geq V(A)$ holds.

To prove the last statement, note that for $A \in \mathcal{A}_j$, $j \in \{P, S, N\}$,

$$V(A) = V^j(A) \times I\{|A| > 1\} + V^{\text{singleton}}(A) \times I\{|A| = 1\} = V^j(A).$$

[40] Such duplicates arise from menus $\{a, b\}$ such that $b \in \mathcal{N}_1(a)$, which belongs to $\mathcal{B}_P \cap \mathcal{B}_S$.

Because $A\alpha B \in \mathcal{A}_j$ for any $A, B \in \mathcal{B}_j$, we obtain $V(A\alpha B) = V^j(A\alpha B) = \alpha V^j(A) + (1-\alpha)V^j(B) = \alpha V(A) + (1-\alpha)V(B)$. □

We now obtain the desired representation as follows. Take V from Lemma 13 and define, for each finite $A \in \mathcal{A}$, $V_{PS}(A) = V(\{a^*, b^*\})$ where $a^* = a^{A*} \in A$ and $b^* = b^{A*} \in \varphi(A)$ are constructed as in Lemma 3. Then

$$\begin{aligned} A \succeq B &\Leftrightarrow \left\{a^{A*}, b^{A*}\right\} \succeq \left\{a^{B*}, b^{B*}\right\} \\ &\Leftrightarrow V\left(\left\{a^{A*}, b^{A*}\right\}\right) \geq V\left(\left\{a^{B*}, b^{B*}\right\}\right) \\ &\Leftrightarrow V_{PS}(A) \geq V_{PS}(A). \end{aligned}$$

Thus, we have obtained the desired function V_{PS}. □

With V_{PS} that represents $\succeq$ over finite menus, we obtain the following result, analogous to GP.

$$V_{PS}(A) = \max_{a \in A} \min_{b \in \varphi_r(B)} V_{PS}(\{a, b\}) = \min_{b \in \varphi_r(B)} \max_{a \in A} V_{PS}(\{a, b\}). \tag{11}$$

Proof is straightforward given the proof of Lemma 3, hence omitted.

Proof of Theorem 1 (Continued).
For some two-component mixtures A of binary menus, Axioms 5 and 6 identify a binary subset of A to which A is indifferent.

Lemma 14. *Suppose Axioms 1-6 hold. Let $A = \{a, b\}\,\alpha\,\{c, d\} \in \mathcal{A}$. Then, the following statements hold.*

(i) If $b \in \mathcal{P}(a) \cup \mathcal{N}_1(a)$ and $d \in \mathcal{P}(c) \cup \mathcal{N}_1(c)$, then $A \sim \{a\alpha c, b\alpha d\}$.

(ii) If $b \in \mathcal{S}(a) \cup \mathcal{N}_1(a)$ and $d \in \mathcal{S}(c) \cup \mathcal{N}_1(c)$, then $A \sim \{a\alpha c, b\alpha d\}$.

(iii) If $b \in \mathcal{I}(a)$, $\{a, b\} \succ \{a\}$ and $d \in \mathcal{S}(c)$, then, $\{a\alpha c, b\alpha d\} \succeq A$. If, in addition, $\{b\} \succ \{a, b\}$, then $\{a\alpha c, b\alpha d\} \sim A$.

(iv) Suppose $b \in \mathcal{I}(a)$, $\{b\} \succ \{a, b\} \sim \{a\}$, $a \in \mathcal{C}(\{a, b\})$ and $d \in \mathcal{S}(c)$. Then, $A \sim \{a\}\,\alpha\,\{c, d\} \succeq \{a\alpha c, b\alpha d\}$, and the latter relation is strict if and only if $\mathcal{C}(\{a, b\}) = \{a\}$.

(v) Suppose $b \in \mathcal{I}(a)$, $\{a\} \sim \{a, b\} \sim \{b\}$ and $d \in \mathcal{S}(c)$. Then,

(v-a) $\mathcal{C}(\{a, b\}) = \{a\}$ implies $\{b\alpha c, a\alpha d\} \succ A \sim \{a\}\,\alpha\,\{c, d\} \succ \{a\alpha c, b\alpha d\}$ and

(v-b) $\mathcal{C}(\{a, b\}) = \{a, b\}$ implies $A \sim \{a\}\,\alpha\,\{c, d\} \sim \{a\alpha c, b\alpha d\}$.

Proof. (i) By Axiom 5(i), $\mathcal{C}(\{a,b\}\,\alpha\,\{c,d\}) = \{b\alpha d\}$. Also, by the linearity of r and Axiom 3(i), we have $r(a\alpha c) > r(z)$, hence $a\alpha c \succeq_w z$, for all $z \in A \backslash \{a\alpha c\}$. Therefore, Axiom 6(i) implies $A \sim \{a\alpha c, b\alpha d\}$.

(ii) The same argument as (i) yields the result.

(iii) Let V_{PS} be a function that represents $\succeq$ over finite menus in $\mathcal{A}$. By Eq.(11), there exists $z \in A$ such that $V_{PS}(A) = \min_{z' \in \varphi_r(A)} V_{PS}(\{z, z'\})$. If $z = a\alpha c$, then by $a\alpha c \in \varphi_r(A)$ and Axioms 3(i) and 6(i), we have $V_{PS}(A) \leq V_{PS}(\{a\alpha c\}) < \alpha V_{PS}(\{a,b\}) + (1-\alpha)V_{PS}(\{c,d\}) = V_{PS}(A)$. Thus, we reach a contradiction. If $z = a\alpha d$, then $V_{PS}(A) \leq V_{PS}(\{a\}\,\alpha\,\{c,d\}) < \alpha V_{PS}(\{a,b\}) + (1-\alpha)V_{PS}(\{c,d\}) = V_{PS}(A)$, a contradiction. If $z = b\alpha c$, then $V_{PS}(A) \leq V_{PS}(\{a,b\}\,\alpha\,\{c\}) < \alpha V_{PS}(\{a,b\}) + (1-\alpha)V_{PS}(\{c,d\}) = V_{PS}(A)$, a contradiction. Thus, $V_{PS}(A) = \min_{z' \in \varphi_r(A)} V_{PS}(\{b\alpha d, z'\}) \leq V_{PS}(\{a\alpha c, b\alpha d\})$. Now, suppose $V_{PS}(b) > V_{PS}(\{a,b\}) > V_{PS}(a)$. Recall $V_{PS}(A) = V_{PS}(\{b\alpha d, z'\})$ for some $z' \in \varphi_r(A)$. By the linearity of r, we have $\varphi_r(A) = \{a\alpha c, b\alpha c\}$. If $z' = b\alpha c$, then $V_{PS}(A) = \alpha V_{PS}(\{b\}) + (1-\alpha)V_{PS}(\{c,d\}) > \alpha V_{PS}(\{a,b\}) + (1-\alpha)V_{PS}(\{c,d\}) = V_{PS}(A)$, a contradiction. Thus, $A \sim \{a\alpha c, b\alpha d\}$.

(iv) By the linearity of r, we have $\varphi_r(A) = \{a\alpha c, b\alpha c\}$. Also, by Axiom 4(i), we have $\{b\alpha c\} \succ \{a\alpha c, b\alpha c\}$, so $a\alpha c \succ_w b\alpha c$. By Axiom 3(i), $a\alpha c \succeq_w z$ for all $z \in A$. Also, by Axiom 5(i), $\mathcal{C}(A) = \mathcal{C}(\{a,b\})\,\alpha\,\{d\}$. Therefore, Axiom 6(i) yields the desired conclusion.

(v) Let $\mathcal{C}(\{a,b\}) = \{a\}$. We first prove the last two relations in (v-a). By Axiom 5(i), $\mathcal{C}(A) = \{a\alpha d\}$ and $\mathcal{C}(\{a,b\}\,\alpha\,\{c\}) = \{a\alpha c\}$. Also, Axiom 4 implies $\{b\alpha c\} \sim \{a\alpha c, b\alpha c\} \sim \{a\alpha c\}$, so $a\alpha c \succ_w b\alpha c$ by case (i-b) of Definition 2. Also, by the linearity of r, $r(a\alpha c) > r(a\alpha d), r(b\alpha d)$. Thus, by Axiom 3(i), $a\alpha c \succeq_r z$ and $a\alpha c \succeq_w z$ for all $z \in A$. By Axiom 6(i), $A \sim \{a\alpha c, a\alpha d\} \succ \{a\alpha c, b\alpha d\}$. Next, to show the first relation in (v-a), note that $b\alpha c \succeq_r z$ for all $z \in A$, $b\alpha c \succeq_w a\alpha d$ (by Axiom 3(i)), and $a\alpha c \succ_w b\alpha c$. Thus, applying Axiom 6(ii) to $\tilde{A} = \{b\alpha c, a\alpha d\}$ and $\tilde{B} = \{a\alpha c, b\alpha d\}$, we obtain $\{b\alpha c, a\alpha d\} \succ \tilde{A} \cup \tilde{B} = A$, establishing the conclusion. Finally, to show (v-b), suppose $\mathcal{C}(\{a,b\}) = \{a,b\}$. Then we have $\mathcal{C}(A) = \{a\alpha d, b\alpha d\}$, and $a\alpha c, b\alpha c \succeq_r z$ and $a\alpha c, b\alpha c \succeq_w z$ for all $z \in A$. Thus, applying Axiom 6(i) to $\tilde{A} = \{a\alpha c, b\alpha d\}$ and $\tilde{B} = \{b\alpha c, a\alpha d\}$ yields $A \sim \{a\alpha c, b\alpha d\}$ and applying it to $\tilde{A} = \{a\alpha c, a\alpha d\}$ and $\tilde{B} = \{b\alpha c, b\alpha d\}$ yields $A \sim \{a\alpha c, a\alpha d\}$. □

Intuitively, previous Lemma 3 says that the value of any finite menu A is determined by the chosen alternative and reference alternative. Lemma 14(i) considers mixing two binary menus in the pride domain, $\{a,b\}$ and $\{c,d\}$, with a and c setting the reference and b and d being chosen. Then, at the mixture menu $A = \{a,b\}\,\alpha\,\{c,d\}$, the reference alternative is the mixture of originally referenced alternatives a and c and the chosen alternative is the mixture of originally chosen alternatives b and d. (ii) states an analogous result for the shame domain. Next, conditions in (iii) imply that b is the chosen alternative at $\{a,b\}$, because the reference point at $\{a,b\}$ is at least as high as that at $\{a\}$. If a sets a lower reference point than b, then $\{a\alpha c, b\alpha d\}$ is strictly better than A, as the two menus share the common alternative to be chosen (i.e., $b\alpha d$) and the former yields a lower reference point (set by $a\alpha c$) than the latter (by $b\alpha c$). Otherwise, the two sets feature the same reference point (set by $a\alpha c$) and are indifferent. Condition $a \sim_r b$ and $\{b\} \succ \{a,b\}$ in particular implies a sets a strictly higher reference point than b does. Case (iv) considers the situation where a sets a higher reference point than b. Then, because $a\alpha c$ sets the reference at A, a binary subset $\{a\alpha c, z\}$ of A is indifferent to A if z is chosen at A and inferior to A otherwise. Finally, when $a \sim_r b$ and $\{a\} \sim \{a,b\} \sim \{b\}$, choice $\mathcal{C}(\{a,b\})$ is useful for identifying the reference setter at $\{a,b\}$. Specifically, if $\mathcal{C}(\{a,b\}) = \{a\}$, then a must set a reference strictly higher than b does; otherwise, $\{a,b\}$ would be strictly better than $\{b\}$. The last two relations in (v-a) follows from $a\alpha c$ setting the reference and $a\alpha d$ being chosen ($b\alpha d$ not being chosen) at A. Also, the first relation in (v-a) holds because $a\alpha d$ is chosen at A and $b\alpha c$ sets a reference point strictly lower than the reference point at A (set by $a\alpha c$). On the other hand, if $\mathcal{C}(\{a,b\}) = \{a,b\}$, then a and b set the same reference point and are equally desirable choices. Thus, for both $z = a, b$, $z\alpha c$ is the reference setter and $z\alpha d$ is the choice at A.

The next lemma is used to construct the normative utility functions separately on the pride domain and shame domain.

Lemma 15. *Suppose Axioms 1-5 hold, $y \in \mathcal{P}(x)$, and $y' \in \mathcal{S}(x)$.*

(i) There exists $\delta \in (0,1)$ such that $y(1-\delta)c \in \mathcal{P}(x)$ and $y'(1-\delta)c \in \mathcal{S}(x)$ for all $c \in \Delta$.

(ii) $y(1-\delta)c \in \mathcal{P}(x(1-\delta)c)$ and $y'(1-\delta)c \in \mathcal{S}(x(1-\delta)c)$ for all $c \in \Delta$ and $\delta \in (0,1)$.

Proof. (i) By definition, $\{x,y\} \succ \{y\}$ and $\mathcal{C}(\{x,y\}) = \{y\}$. Because the restriction of $\succeq$ to singleton sets is continuous, and because Δ is compact, there exists $\delta_1 \in (0,1)$ such that $\{x, y(1-\delta)c\} \succ \{y(1-\delta)c\}$ for all $c \in \Delta$ and $\delta \in (0,\delta_1)$.[41] Also, by Axiom 3(iii-b) and compactness, we have some $\delta_2 \in (0,1)$ such that $\mathcal{C}(\{x, y(1-\delta)c\}) = \{y(1-\delta)c\}$ for all $c \in \Delta$ and $\delta \in (0,\delta_2)$. Therefore, by taking $\underline{\delta}^P = \min\{\delta_1,\delta_2\}$, the first half of the statement holds for

[41] Let $A = \{x,y\}\, 0.5 \,\{y\}$. By Axiom 4, we have $\{x,y\} \succ A \succ \{y\}$. By lower semi-continuity and singleton continuity, we have $\{x, y(1-\delta)c\} \succ A$ and $A \succ \{y(1-\delta)c\}$ for all sufficiently small δ.

all $\delta < \underline{\delta}^P$. An analogous argument yields $\underline{\delta}^S$ such that the second half of the statement holds for all $\delta < \underline{\delta}^S$. Thus, $\delta < \min\left\{\underline{\delta}^P, \underline{\delta}^S\right\}$ satisfies the desired property.

(ii) By Axiom 4(iii), $\{x(1-\delta)c, y(1-\delta)c\} \succ \{y(1-\delta)c\}$ for any $c \in \Delta$ and $\delta \in (0,1)$. Also, by Axiom 5(iii), $\mathcal{C}(\{x(1-\delta)c, y(1-\delta)c\}) = \{y(1-\delta)c\}$. Thus, the first half of the statement holds. The second half can be proved analogously. □

We now construct u and w. For any $a \in \Delta$, let $u(a) = V_{PS}(\{a\})$. By Nondegeneracy, there exist $x, y, y' \in \Delta$ such that $y \in \mathcal{P}(x)$ and $y' \in \mathcal{S}(x)$. Below, we fix such x, y, y'. By Lemma 15, there exists $\delta \in (0,1)$ such that, for all $c \in \Delta$, we have $y(1-\delta)c \in \mathcal{P}(x)$ and $y'(1-\delta)c \in \mathcal{S}(x)$. Note that by Axioms 3(i) and 6(i), $b \in \mathcal{P}(a) \cup \mathcal{S}(a)$ implies $\{a,b\} \succ \{a\}$ for any $a, b \in \Delta$, so we in particular have $\{x, y(1-\delta)c\} \succ \{x\}$ and $\{x, y'(1-\delta)c\} \succ \{x\}$. Then, define w_P and w_S by

$$w_P(c; x, y, \delta) = \frac{1}{\delta} V_{PS}(\{x, y(1-\delta)c\}) - \frac{1-\delta}{\delta} V_{PS}(\{x, y\}) - V_{PS}(\{c\})$$
$$w_S(c; x, y', \delta) = \frac{1}{\delta} V_{PS}(\{x, y'(1-\delta)c\}) - \frac{1-\delta}{\delta} V_{PS}(\{x, y'\}) - V_{PS}(\{c\}).$$

$w_P(c; x, y, \delta)$ measures how the utility changes as the reference alternative x is moved slightly toward c, keeping the ex post choice constant.$w_S(c; x, y', \delta)$ is interpreted analogously. The next two lemmas show some properties of w_P and w_S, including its linearity and independence of the specific choice of $x, y, y' \in \Delta$.

Lemma 16. *Suppose Axioms 1-6 hold. If $y(1-\delta)c \in \mathcal{P}(x)$ for all $c \in \Delta$, then the following statements hold.*

(i) If $c \in \mathcal{P}(x)$, then $w_P(c; x, y, \delta) = V_{PS}(\{x, c\}) - V_{PS}(\{c\})$.

(ii) $w_P(x; x, y, \delta) = 0$.

(iii) $w_P(c\alpha c'; x, y, \delta) = \alpha w_P(c; x, y, \delta) + (1-\alpha) w_P(c'; x, y, \delta)$ for any $\alpha \in (0,1)$.

(iv) $w_P(c; x, y, \delta') = w_P(c; x, y, \delta)$ for any $\delta' \in (0, \delta)$.

(v) Suppose $b(1-\delta)c \in \mathcal{P}(a)$ for all $c \in \Delta$. Then $w_P(c; x, y, \delta) = w_P(c; a, b, \delta) + w_P(a; x, y, \delta)$.

Proof. (i) Because $y, c \in \mathcal{P}(x)$, $V_{PS}(\{x, y(1-\delta)c\}) = V_{PS}(\{x,y\}(1-\delta)\{x,c\})$ by Lemma 14(i). Therefore,

$$\begin{aligned} w_P(c; x, y, \delta) &= \frac{1}{\delta} V_{PS}(\{x, y(1-\delta)c\}) - \frac{1-\delta}{\delta} V_{PS}(\{x,y\}) - V_{PS}(\{c\}) \\ &= \frac{1}{\delta}\left[(1-\delta)V_{PS}(\{x,y\}) + \delta V_{PS}(\{x,c\}) - (1-\delta)V_{PS}(\{x,y\}) - \delta V_{PS}(\{c\})\right] \\ &= V_{PS}(\{x,c\}) - V_{PS}(\{c\}) \end{aligned}$$

where the second equality follows from $\{x,y\}, \{x,c\} \in \mathcal{B}_P$.

(ii) By $\{x,y\}, \{x\} \in \mathcal{B}_P$, $V_{PS}(\{x, y(1-\delta)x\}) = (1-\delta)V_{PS}(\{x,y\}) + \delta V_{PS}(\{x\})$, so

$$\begin{aligned} w_P(x; x, y, \delta) &= \frac{1}{\delta} V_{PS}(\{x, y(1-\delta)x\}) - \frac{1-\delta}{\delta} V_{PS}(\{x,y\}) - V_{PS}(\{x\}) \\ &= 0. \end{aligned}$$

(iii) Because $y(1-\delta)c, y(1-\delta)c' \in \mathcal{P}(x)$, Lemma 14(i) implies

$$V_{PS}(\{x, [y(1-\delta)c]\,\alpha\,[y(1-\delta)c']\}) = V_{PS}(\{x, y(1-\delta)c\}\,\alpha\,\{x, y(1-\delta)c'\}).$$

Therefore,

$$\begin{aligned} w_P(c\alpha c'; x, y, \delta) =& \frac{1}{\delta} V_{PS}(\{x, [y(1-\delta)c]\,\alpha\,[y(1-\delta)c']\}) \\ &- \frac{1-\delta}{\delta} V_{PS}(\{x,y\}) - V_{PS}(\{c\alpha c'\}) \\ =& \frac{\alpha}{\delta} V_{PS}(\{x, y(1-\delta)c\}) + \frac{1-\alpha}{\delta} V_{PS}(\{x, y(1-\delta)c'\}) \\ &- \frac{1-\delta}{\delta} V_{PS}(\{x,y\}) - \alpha V_{PS}(\{c\}) - (1-\alpha)V_{PS}(\{c'\}) \\ =& \alpha w_P(c; x, y, \delta) + (1-\alpha) w_P(c'; x, y, \delta). \end{aligned}$$

(iv) Let $\delta' \in (0, \delta)$. Note that $y(1-\delta')c = \frac{\delta-\delta'}{\delta} y + \left(1 - \frac{\delta-\delta'}{\delta}\right)[y(1-\delta)c]$. Because $y, y(1-\delta)c \in \mathcal{P}(x)$, Lemma 14(i) implies

$$\begin{aligned} & V_{PS}\left(\left\{x, \frac{\delta-\delta'}{\delta} y + \left(1 - \frac{\delta-\delta'}{\delta}\right)[y(1-\delta)c]\right\}\right) \\ =& V_{PS}\left(\frac{\delta-\delta'}{\delta}\{x,y\} + \left(1 - \frac{\delta-\delta'}{\delta}\right)\{x, y(1-\delta)c\}\right). \end{aligned}$$

Therefore, $V_{PS}(\{x, y(1-\delta)c\}) = \frac{\delta}{\delta'} V_{PS}(\{x, y(1-\delta')c\}) - \frac{\delta-\delta'}{\delta'} V_{PS}(\{x, y\})$. Substituting this into the definition, we have

$$\begin{aligned} w_P(c; x, y, \delta) =& \frac{1}{\delta} V_{PS}(\{x, y(1-\delta)c\}) - \frac{1-\delta}{\delta} V_{PS}(\{x, y\}) - V_{PS}(\{c\}) \\ =& \frac{1}{\delta'} V_{PS}(\{x, y(1-\delta')c\}) - \frac{\delta-\delta'}{\delta\delta'} V_{PS}(\{x, y\}) \\ &- \frac{1-\delta}{\delta} V_{PS}(\{x, y\}) - V_{PS}(\{c\}) \\ =& w_P(c; x, y, \delta'). \end{aligned}$$

(v) What to be shown is

$$\begin{aligned} & w_P(c; x, y, \delta) = w_P(c; a, b, \delta) + w_P(a; x, y, \delta). \\ \Leftrightarrow & \frac{1}{\delta} V_{PS}(\{x, y(1-\delta)c\}) = \frac{1}{\delta} V_{PS}(\{a, b(1-\delta)c\}) - \frac{1-\delta}{\delta} V_{PS}(\{a, b\}) \\ & - V_{PS}(\{a\}) + \frac{1}{\delta} V_{PS}(\{x, y(1-\delta)a\}). \end{aligned}$$

By (ii), we have $V_{PS}(\{a\}) = \frac{1}{\delta} V_{PS}(\{a, b(1-\delta)a\}) - \frac{1-\delta}{\delta} V_{PS}(\{a, b\})$. Substituting this into the above expression, our goal is to show

$$V_{PS}(\{x, y(1-\delta)c\}\, 0.5\, \{a, b(1-\delta)a\}) = V_{PS}(\{a, b(1-\delta)c\}\, 0.5\, \{x, y(1-\delta)a\}).$$

Because $y(1-\delta)c, y(1-\delta)a \in \mathcal{P}(x)$ and $b(1-\delta)a, b(1-\delta)c \in \mathcal{P}(a)$, Lemma 14(i) implies that both sides of this equation equal to

$$V_{PS}(x0.5a, [(1-\delta)(y+b)]\, 0.5\, [\delta(a+c)]).$$

□

Lemma 17. *Suppose Axioms 1-6 hold. If $y'(1-\delta)c \in \mathcal{S}(x)$ for all $c \in \Delta$, then the following properties hold.*

(i) If $c \in \mathcal{S}(x)$, then $w_S(c; x, y', \delta) = V_{PS}(\{x, c\}) - V_{PS}(\{c\})$.

(ii) $w_S(x; x, y', \delta) = 0$.

(iii) $w_S(c\alpha c'; x, y', \delta) = \alpha w_S(c; x, y', \delta) + (1-\alpha) w_S(c'; x, y', \delta)$ for any $\alpha \in (0, 1)$.

(iv) $w_S(c; x, y', \delta') = w_S(c; x, y', \delta)$ for all $\delta' \in (0, \delta)$.

(v) *Suppose* $b' \in \mathcal{S}(a')$ *for all* $c \in \Delta$*. Then* $w_S(c; x, y', \delta) = w_S(c; a', b', \delta) + w_S(a'; x, y', \delta)$*.*

Proof is analogous to that of Lemma 16 and therefore omitted.

Now, we relate w_P to w_S using Axiom 7.

Lemma 18. *Suppose Axioms 1-7 hold, and suppose* $y(1-\delta)c \in \mathcal{P}(x)$ *and* $y'(1-\delta)c \in \mathcal{S}(x)$ *for all* $c \in \Delta$*. Then, there exists* $\beta > 0$ *such that* $w_P(c; x, y, \delta) = \beta w_S(c; x, y', \delta)$ *for all* $c \in \Delta$*.*

Proof. Let $\beta = \frac{1}{\alpha} - 1 > 0$, where α is as defined in Axiom 7. We then have

$$\begin{aligned}
&\delta \left[w_P\left(c; x, y, \delta\right) - \beta w_S\left(c; x, y', \delta\right) \right] \\
&= V_{PS}\left(\{x, y(1-\delta)c\}\right) - (1-\delta) V_{PS}\left(\{x, y\}\right) - \delta V_{PS}\left(\{c\}\right) \\
&\quad - \beta \left[V_{PS}\left(\{x, y'(1-\delta)c\}\right) - (1-\delta) V_{PS}\left(\{x, y'\}\right) - \delta V_{PS}\left(\{c\}\right) \right] \\
&= \frac{1}{\alpha} \left[\alpha V_{PS}\left(\{x, y(1-\delta)c\}\right) + (1-\alpha) V_{PS}\left(\{x(1-\delta)c, y'(1-\delta)c\}\right) \right] \\
&\quad - \frac{1}{\alpha} \left[\alpha V_{PS}\left(\{x(1-\delta)c, y(1-\delta)c\}\right) + (1-\alpha) V_{PS}\left(\{x, y'(1-\delta)c\}\right) \right] \\
&= \frac{1}{\alpha} \left[\alpha V_{PS}\left(\{x, y(1-\delta)c\}\right) + (1-\alpha) V_{PS}\left(\left\{ e^{x(1-\delta)c, y'(1-\delta)c} \right\} \right) \right] \\
&\quad - \frac{1}{\alpha} \left[\alpha V_{PS}\left(\{x(1-\delta)c, y(1-\delta)c\}\right) + (1-\alpha) V_{PS}\left(\left\{ e^{x, y'(1-\delta)c} \right\} \right) \right] \\
&= 0
\end{aligned}$$

where the last equality holds because $y(1-\delta)c \in \mathcal{P}(x) \cap \mathcal{P}(x(1-\delta)c)$ and $y'(1-\delta)c \in \mathcal{S}(x) \cap \mathcal{S}(x(1-\delta)c)$ hold by Lemma 15, so that Axiom 7 applies. Thus, $w_P\left(c; x, y, \delta\right) = \beta w_S\left(c; x, y', \delta\right)$ where $\beta > 0$. □

We now show that the representation holds for binary menus which include x. Let $w(c; x, y, y', \delta) = \frac{1}{\beta} w_P(c; x, y, \delta) = w_S(c; x, y', \delta)$, which is well-defined by Lemma 18.

Lemma 19. *Suppose Axioms 1-8 hold. Consider* $x, b \in \Delta$ *such that* $r(x) \geq r(b)$ *and* $V_{PS}\left(\{x, b\}\right) \geq V_{PS}\left(\{x\}\right)$*. Suppose* $y(1-\delta)c \in \mathcal{P}(x)$ *and* $y'(1-\delta)c \in \mathcal{S}(x)$ *for all* $c \in \Delta$*. Then* V_{PS} *is expressed as*

$$V_{PS}\left(\{x, b\}\right) = \max_{c \in \{x, b\}} \left[g\left(c, \max_{c' \in \varphi_r(\{x, b\})} w(c'; x, y, y', \delta) \right) \right]$$

and $\mathcal{C}(\{x,b\})$ *coincides with*

$$\mathcal{C}_{PS}(\{x,b\}) = \arg\max_{c\in\{x,b\}}\left[g\left(c, \max_{c'\in\varphi_r(\{x,b\})} w(c';x,y,y',\delta)\right)\right]$$

where

$$\begin{aligned} g(c,R) &= g(c,R;x,y,y',\delta) \\ &= u(c) - \max\{R - w(c;x,y,y',\delta), 0\} + \beta\max\{w(c;x,y,y',\delta) - R, 0\} \end{aligned}$$

and $\varphi_r(A) = \arg\max_A r$.

Proof. Note first that Axioms 3(i) and 6(i) imply $V_{PS}(\{x, y(1-\delta)c\}) > V_{PS}(\{x\})$ and $V_{PS}(\{x, y'(1-\delta)c\}) > V_{PS}(\{x\})$. Consider the following exhaustive cases.

Case 1. Suppose $b \in \mathcal{P}(x)$. By Axioms 3(i) and 6(i), $V_{PS}(\{x,b\}) > V_{PS}(x)$. By definition, $\mathcal{C}(\{x,b\}) = \{b\}$. By Lemma 16(i)(ii) and $b \in \mathcal{P}(x)$, $w(b;x,y,y',\delta) - w(x;x,y,y',\delta) = \frac{1}{\beta}w_P(b;x,y,\delta) = \frac{1}{\beta}[V_{PS}(\{x,b\}) - V_{PS}(\{b\})] > 0$. Therefore,

$$\begin{aligned} g\left(b, \max_{c'\in\varphi_r(\{x,b\})} w(c';x,y,y',\delta)\right) &= u(b) + \beta[w(b;x,y,y',\delta) - w(x;x,y,y',\delta)] \\ &= V_{PS}(\{x,b\}) \\ &> V_{PS}(\{x\}) \\ &= g\left(x, \max_{c'\in\varphi_r(\{x,b\})} w(c';x,y,y',\delta)\right). \end{aligned}$$

Thus, the conclusion holds.

Case 2. Suppose $b \in \mathcal{S}(x)$ (and therefore $V_{PS}(\{x,b\}) > V_{PS}(x)$). By definition, $\mathcal{C}(\{x,b\}) = \{b\}$. By Lemma 17(i)(ii) and $b \in \mathcal{S}(x)$, $w(b;x,y,y',\delta) - w(x;x,y,y',\delta) = w_S(b;x,y',\delta) = V_{PS}(\{x,b\}) - V_{PS}(\{b\}) < 0$. Therefore,

$$\begin{aligned} g\left(b, \max_{c'\in\varphi_r(\{x,b\})} w(c';x,y,y',\delta)\right) &= u(b) + w(b;x,y,y',\delta) - w(x;x,y,y',\delta) \\ &= V_{PS}(\{x,b\}) \\ &> V_{PS}(\{x\}) \\ &= g\left(x, \max_{c'\in\varphi_r(\{x,b\})} w(c';x,y,y',\delta)\right). \end{aligned}$$

Therefore, the conclusion holds.

Case 3. Suppose $b \in \mathcal{N}_1(x)$ (so $V_{PS}(\{x,b\}) > V_{PS}(x)$). By definition, $\mathcal{C}(\{x,b\}) = \{b\}$. Letting $A' = \{x,y'\}(1-\delta)\{x,b\}$, we also have

$$\begin{aligned} w(b;x,y,y',\delta) - w(x;x,y,y',\delta) &= w_S(b;x,y',\delta) \\ &= \frac{1}{\delta}V_{PS}(\{x,y'(1-\delta)b\}) - \frac{1-\delta}{\delta}V_{PS}(\{x,y'\}) - V_{PS}(\{b\}) \\ &= \frac{1}{\delta}\left[V_{PS}(A') - (1-\delta)V_{PS}(\{x,y'\}) - \delta V_{PS}(\{b\})\right] \\ &= 0 \end{aligned}$$

where the third equality follows from Lemma 14(ii) and the last equality follows from $\{x,b\} \sim \{b\}$. Therefore,

$$\begin{aligned} g\left(b, \max_{c' \in \varphi_r(\{x,b\})} w(c';x,y,y',\delta)\right) &= V_{PS}(\{b\}) \\ &= V_{PS}(\{x,b\}) \\ &> V_{PS}(\{x\}) \\ &= g\left(x, \max_{c' \in \varphi_r(\{x,b\})} w(c';x,y,y',\delta)\right). \end{aligned}$$

Thus, the conclusion holds.

Case 4. Suppose $b \in \mathcal{N}_2(x)$. By Lemma 9, $V_{PS}(\{x,b\}) = V_{PS}(\{x\})$. By definition, $x \in \mathcal{C}(\{x,b\})$. Consider first the case where $w(b;x,y,y',\delta) > 0$. By Axiom 4(iii),

$$\begin{aligned} &\frac{1}{\delta}V_{PS}(\{x,y(1-\delta)b\}) - \frac{1-\delta}{\delta}V_{PS}(\{x,y\}) - V_{PS}(\{b\}) \\ &= w_P(b;x,y,\delta) \\ &> 0 \\ &= \frac{1}{\delta}V_{PS}(\{x,y\}(1-\delta)\{b\}) - \frac{1-\delta}{\delta}V_{PS}(\{x,y\}) - V_{PS}(\{b\}), \end{aligned}$$

so $V_{PS}(\{x,y(1-\delta)b\}) > V_{PS}(\{x,y\}(1-\delta)\{b\})$, which together with Axiom 5(ii-a) implies $\mathcal{C}(A) = \{y\}(1-\delta)\mathcal{C}(\{x,b\})$ where $A = \{x,y\}(1-\delta)\{x,b\}$.[42] Then, by Axiom 6(i), $\{x,y(1-\delta)x\} \sim A \succeq \{x,(1-\delta)y+\delta b\}$ and the latter

[42] To apply Axiom 5(ii), recall $\mathcal{C}(\{x,y(1-\delta)c\}) = \{y(1-\delta)c\}$ for all $c \in \Delta$.

relation is strict if and only if $\mathcal{C}(\{x,b\}) = \{x\}$. Therefore,

$$\begin{aligned}
g\left(b, \max_{c' \in \varphi_r(\{x,b\})} w(c'; x, y, y', \delta)\right) &= u(b) + \beta \cdot \frac{1}{\beta} w_P(b; x, y, \delta) \\
&\leq \frac{1}{\delta} V_{PS}\left(\{x, y(1-\delta)x\}\right) - \frac{1-\delta}{\delta} V_{PS}\left(\{x, y\}\right) \\
&= V_{PS}\left(\{x\}\right) \\
&= V_{PS}\left(\{x, b\}\right) \\
&= g\left(x, \max_{c' \in \varphi_r(\{x,b\})} w(c'; x, y, y', \delta)\right)
\end{aligned}$$

where the inequality is strict if and only if $\mathcal{C}(\{x,b\}) = \{x\}$. Next, consider the case with $w(b; x, y, y', \delta) \leq 0$. By Axiom 4(iii),

$$\begin{aligned}
&\frac{1}{\delta} V_{PS}\left(\{x, y'(1-\delta)b\}\right) - \frac{1-\delta}{\delta} V_{PS}\left(\{x, y'\}\right) - V_{PS}\left(\{b\}\right) \\
&= w_S(b; x, y', \delta) \\
&\leq 0 \\
&= \frac{1}{\delta} V_{PS}\left(\{x, y'\}(1-\delta)\{b\}\right) - \frac{1-\delta}{\delta} V_{PS}\left(\{x, y'\}\right) - V_{PS}\left(\{b\}\right),
\end{aligned}$$

so $V_{PS}\left(\{x, y'(1-\delta)b\}\right) \leq V_{PS}\left(\{x, y'\}(1-\delta)\{b\}\right)$, which together with Axiom 5(ii-b) implies $\mathcal{C}(A') = \{y'\}(1-\delta)\mathcal{C}(\{x,b\})$ where $A' = \{x, y'\}(1-\delta)\{x, b\}$. Then, by Axiom 6(i), $\{x, y'(1-\delta)x\} \sim A' \succeq \{x, y'(1-\delta)b\}$ and the latter relation is strict if and only if $\mathcal{C}(\{x,b\}) = \{x\}$. Therefore,

$$\begin{aligned}
g\left(b, \max_{c' \in \varphi_r(\{x,b\})} w(c'; x, y, y', \delta)\right) &= u(b) + w_S(b; x, y', \delta) \\
&\leq \frac{1}{\delta} V_{PS}\left(\{x, y'(1-\delta)x\}\right) - \frac{1-\delta}{\delta} V_{PS}\left(\{x, y'\}\right) \\
&= V_{PS}\left(\{x\}\right) \\
&= V_{PS}\left(\{x, b\}\right) \\
&= g\left(x, \max_{c' \in \varphi_r(\{x,b\})} w(c'; x, y, y', \delta)\right)
\end{aligned}$$

where the inequality is strict if and only if $\mathcal{C}(\{x,b\}) = \{x\}$.

Case 5. Suppose $b \in \mathcal{I}(x)$ and $V_{PS}\left(\{x,b\}\right) > V_{PS}(x)$. By Lemma 2, we have

$V_{PS}(b) \geq V_{PS}(\{x,b\}) > V_{PS}(x)$. Also, Axiom 6 implies $\mathcal{C}(\{x,b\}) = \{b\}$.[43] By Lemma 14(iii), $V_{PS}(\{x, y'(1-\delta)b\}) \geq V_{PS}(A')$ where A' is defined above. Now, consider first the case with $V_{PS}(b) = V_{PS}(\{x,b\}) > V_{PS}(x)$. Then,

$$\begin{aligned} V_{PS}(\{x, y'(1-\delta)b\}) &\geq (1-\delta)V_{PS}(\{x,y'\}) + \delta V_{PS}(\{x,b\}) \\ &= (1-\delta)V_{PS}(\{x,y'\}) + \delta V_{PS}(\{b\}), \end{aligned}$$

so $w(b;x,y,y',\delta) \geq 0 = w(x;x,y,y',\delta)$. Therefore,

$$\begin{aligned} g\left(b, \max_{c' \in \varphi_r(\{x,b\})} w(c';x,y,y',\delta)\right) &= V_{PS}(\{b\}) \\ &= V_{PS}(\{x,b\}) \\ &> V_{PS}(\{x\}) \\ &\geq g\left(x, \max_{c' \in \varphi_r(\{x,b\})} w(c';x,y,y',\delta)\right). \end{aligned}$$

Next, consider the case with $V_{PS}(b) > V_{PS}(\{x,b\}) > V_{PS}(x)$. By Lemma 14(iii), we have $V_{PS}(\{x, y'(1-\delta)b\}) = V_{PS}(A')$. Therefore,

$$\begin{aligned} w(b;x,y,y',\delta) &= \frac{1-\delta}{\delta} V_{PS}(\{x,y'\}) + V_{PS}(\{x,b\}) - \frac{1-\delta}{\delta} V_{PS}(\{x,y'\}) - V_{PS}(\{b\}) \\ &< 0 = w(x;x,y,y',\delta). \end{aligned}$$

Thus,

$$\begin{aligned} g\left(b, \max_{c' \in \varphi_r(\{x,b\})} w(c';x,y,y',\delta)\right) &= u(b) + w(b;x,y,y',\delta) \\ &= V_{PS}(\{x,b\}) \\ &> V_{PS}(\{x\}) \\ &= g\left(x, \max_{c' \in \varphi_r(\{x,b\})} w(c';x,y,y',\delta)\right). \end{aligned}$$

Case 6. Suppose $b \in \mathcal{I}(x)$ and $V_{PS}(\{b\}) > V_{PS}(\{x,b\}) = V_{PS}(x)$. Note we have $x \in \mathcal{C}(\{x,b\})$.[44] Then, by Lemma 14(iv), $V_{PS}(A') \geq V_{PS}(\{x, y'(1-\delta)b\})$.

---

[43] If $x \succeq_w b$, then $\mathcal{C}(\{x,b\}) = \{b\}$ by Axiom 6(i). Suppose $b \succ_w x$. Then, by Axiom 6(ii), $x \in \mathcal{C}(\{x,b\})$ implies $\{x\} \succ \{x,b\}$, a contradiction. Thus, $\mathcal{C}(\{x,b\}) = \{b\}$.

[44] Otherwise, $\{b\} \succ \{x,b\}$ implies $x \succ_w b$ whereas $\{x\} \sim \{x,b\}$ and $\mathcal{C}(\{x,b\}) = \{b\}$ imply $b \succ_w x$, contradicting Axiom 3(i).

Therefore,

$$\begin{aligned} w(b;x,y,y',\delta) &\le \frac{1}{\delta}V_{PS}\left(\{x,y'\}\left(1-\delta\right)\{x,b\}\right) - \frac{1-\delta}{\delta}V_{PS}\left(\{x,y'\}\right) - V_{PS}\left(\{b\}\right) \\ &< 0 = w(x;x,y,y',\delta), \end{aligned}$$

so

$$\begin{aligned} g\left(b, \max_{c'\in\varphi_r(\{x,b\})} w(c';x,y,y',\delta)\right) &= u(b) + w(b;x,y,y',\delta) \\ &\le \frac{1}{\delta}V_{PS}\left(\{x,y'\}\left(1-\delta\right)\{x,b\}\right) - \frac{1-\delta}{\delta}V_{PS}\left(\{x,y'\}\right) \\ &= V_{PS}\left(\{x,b\}\right) \\ &= V_{PS}\left(\{x\}\right) \\ &= g\left(x, \max_{c'\in\varphi_r(\{x,b\})} w(c';x,y,y',\delta)\right). \end{aligned}$$

By Lemma 14(iv) the inequality is strict if and only if $\mathcal{C}\left(\{x,b\}\right) = \{x\}$.

Case 7. Suppose $b \in \mathcal{I}(x)$ and $V_{PS}\left(\{b\}\right) = V_{PS}\left(\{x,b\}\right) = V_{PS}\left(x\right)$. If $\mathcal{C}\left(\{x,b\}\right) = \{x\}$, then by Lemma 14(v-a), we have

$$\begin{aligned} w(b;x,y,y',\delta) &< \frac{1}{\delta}V_{PS}\left(\{x,y'(1-\delta)x\}\right) - \frac{1-\delta}{\delta}V_{PS}\left(\{x,y'\}\right) - V_{PS}\left(\{b\}\right) \\ &= 0 = w(x;x,y,y',\delta). \end{aligned}$$

Therefore,

$$\begin{aligned} g\left(b, \max_{c'\in\varphi_r(\{x,b\})} w(c';x,y,y',\delta)\right) &= V_{PS}\left(\{b\}\right) + w(b;x,y,y',\delta) \\ &< V_{PS}\left(\{b\}\right) \\ &= V_{PS}\left(\{x,b\}\right) \\ &= V_{PS}\left(\{x\}\right) \\ &= g\left(x, \max_{c'\in\varphi_r(\{x,b\})} w(c';x,y,y',\delta)\right). \end{aligned}$$

If $\mathcal{C}(\{x,b\}) = \{b\}$, then Lemma 14(v-a) implies

$$\begin{aligned} w(b;x,y,y',\delta) &> \frac{1}{\delta}V_{PS}\left(\{x(1-\delta)b, y'(1-\delta)b\}\right) - \frac{1-\delta}{\delta}V_{PS}\left(\{x,y'\}\right) - V_{PS}\left(\{b\}\right) \\ &= 0 = w(x;x,y,y',\delta), \end{aligned}$$

so

$$\begin{aligned} g\left(b, \max_{c'\in\varphi_r(\{x,b\})} w(c';x,y,y',\delta)\right) &= V_{PS}\left(\{b\}\right) \\ &= V_{PS}\left(\{x,b\}\right) \\ &= V_{PS}\left(\{x\}\right) \\ &> g\left(x, \max_{c'\in\varphi_r(\{x,b\})} w(c';x,y,y',\delta)\right). \end{aligned}$$

Finally, if $\mathcal{C}(\{x,b\}) = \{x,b\}$, then Lemma 14(v-b) implies

$$w(b;x,y,y',\delta) = \frac{1}{\delta}V_{PS}\left(\{x, y'(1-\delta)x\}\right) - \frac{1-\delta}{\delta}V_{PS}\left(\{x,y'\}\right) - V_{PS}\left(\{b\}\right) = 0,$$

so the desired representation holds.

Case 8. Suppose $b \in \mathcal{I}(x)$ and $V_{PS}(\{x,b\}) > V_{PS}(b)$. Note we must have $\mathcal{C}(\{x,b\}) = \{x\}$ and $V_{PS}(\{x\}) = V_{PS}(\{x,b\})$.[45] By Axiom 4, $\{x\} \sim \{x,b\}\,\delta\,\{x\} \succ \{b\}\,\delta\,\{x\}$. Therefore, $x \succ_w x(1-\delta)b$ by case (i-c) of Definition 2. Thus, Axioms 3(i) and 6(i)imply $\{x, x(1-\delta)b, y'(1-\delta)b\} \succeq \{x, y'(1-\delta)b\}$. Also, $\mathcal{C}(\{x, y'(1-\delta)b\}) = \{y'(1-\delta)b\}$ by construction, and Axiom 8 imply $x \notin \mathcal{C}(\{x, x(1-\delta)b, y'(1-\delta)b\})$. Therefore, by Axiom 6(ii), $\{x(1-\delta)b, y'(1-\delta)b\} \succ \{x, x(1-\delta)b, y'(1-\delta)b\}$. Combining these results, $\{x(1-\delta)b, y'(1-\delta)b\} \succ \{x, y'(1-\delta)b\}$. Then,

$$\begin{aligned} w(b;x,y,y',\delta) &< \frac{1}{\delta}V_{PS}\left(\{x,y'\}(1-\delta)\{b\}\right) - \frac{1-\delta}{\delta}V_{PS}\left(\{x,y'\}\right) - V_{PS}\left(\{b\}\right) \\ &= 0 = w(x;x,y,y',\delta), \end{aligned}$$

[45] The former can be shown by applying Axiom 6 to each of the cases $b \succeq_w x$ and $x \succ_w b$. The latter follows from Lemma 2(ii) and the assumption that $V_{PS}(\{x,b\}) \geq V_{PS}(\{x\})$.

so

$$g\left(b, \max_{c' \in \varphi_r(\{x,b\})} w(c'; x, y, y', \delta)\right) = V_{PS}(\{b\}) + w(b; x, y, y', \delta)$$
$$< V_{PS}(\{b\})$$
$$< V_{PS}(\{x, b\})$$
$$= V_{PS}(\{x\})$$
$$= g\left(x, \max_{c' \in \varphi_r(\{x,b\})} w(c'; x, y, y', \delta)\right).$$

□

Now, we can prove that the representation holds for an arbitrary binary menu $\{a, b\}$.

Lemma 20. *Suppose Axioms 1-8 and Nondegeneracy hold. Then there exists continuous and linear functions u, w and r such that V_{PS} is expressed as*

$$V_{PS}(\{a, b\}) = \max_{c \in \{a,b\}} g\left(c, \max_{c' \in \varphi_r(\{a,b\})} w(c')\right)$$

and $\mathcal{C}(\{a, b\})$ coincides with

$$\mathcal{C}_{PS}(\{a, b\}) = \arg\max_{c \in \{a,b\}} g\left(c, \max_{c' \in \varphi_r(\{a,b\})} w(c')\right)$$

where g and φ are defined above.

Proof. Take $x, y, y' \in \Delta$ such that $y \in \mathcal{P}(x)$ and $y' \in \mathcal{S}(x)$, which exist by Nondegeneracy. Take an arbitrary $\{a, b\}$. First, consider the case with $a, b \in \text{int}(\Delta)$. By Lemma 15 and Axioms 3(i) and 6(i), we can choose a sufficiently small $\delta \in (0, 1)$ such that $y(1-\delta)b \in \mathcal{P}(x)$, $V_{PS}(\{x, y(1-\delta)b\}) > V_{PS}(\{x\})$, $y'(1-\delta)b \in \mathcal{S}(x)$, and $V_{PS}(\{x, y'(1-\delta)b\}) > V_{PS}(\{x\})$. Without loss of generality, suppose $r(a) \geq r(b)$ and $V_{PS}(\{a, b\}) \geq V_{PS}(\{a\})$.[46] Because $a \in \text{int}(\Delta)$, there exist $\bar{a} \in \Delta$ and $\alpha \in (0, 1)$ such that $a = \bar{a}\alpha x$. Define $z = \bar{a}\alpha y$ and $z' = \bar{a}\alpha y'$. Then, by Lemma 15 and Axioms 3(i), 4, and 6(i), we have $z \in \mathcal{P}(a)$, $V_{PS}(\{a, z\}) = \alpha V_{PS}(\{\bar{a}\}) + (1-\alpha)V_{PS}(\{x, y\}) > V_{PS}(\{a\})$, $z' \in \mathcal{S}(a)$, and $V_{PS}(\{a, z'\}) = \alpha V_{PS}(\{\bar{a}\}) + (1-\alpha)V_{PS}(\{x, y'\}) > V_{PS}(\{a\})$.

[46] If $r(a) > r(b)$, then Lemma 2(i) implies $V_{PS}(\{a, b\}) \geq V_{PS}(\{a\})$. If $r(a) = r(b)$, then Lemma 2(ii) implies $V_{PS}(\{a, b\}) \geq V_{PS}(\{a\})$ or $V_{PS}(\{a, b\}) \geq V_{PS}(\{b\})$, so the assumption is without loss of generality.

By Lemma 15, Axiom 2(i), and compactness, for sufficiently small $\delta' \in (0,1)$, we have $z(1-\delta')c \in \mathcal{P}(a)$, $V_{PS}(\{a, z(1-\delta')c\}) > V_{PS}(\{a\})$, $z'(1-\delta')c \in \mathcal{S}(a)$, and $V_{PS}(\{a, z'(1-\delta')c\}) > V_{PS}(\{a\})$ for all $c \in \Delta$. Defining $g(c,R) = g(c,R;a,z,z',\delta')$ as in Lemma 19, V_{PS} can be expressed as

$$V_{PS}(\{a,b\}) = \max_{c\in\{a,b\}} g\left(c, \max_{c'\in\varphi_r(\{a,b\})} w(c';a,z,z',\delta')\right)$$

and $\mathcal{C}(\{a,b\})$ coincides with

$$\mathcal{C}_{PS}(\{a,b\}) = \arg\max_{c\in\{a,b\}} g\left(c, \max_{c'\in\varphi_r(\{a,b\})} w(c';a,z,z',\delta')\right).$$

Now, let $\delta^* = \min\{\delta, \delta'\}$. Then, by Lemmas 16 and 17, we have $w(\cdot;x,y,y',\delta) = w(\cdot;x,y,y',\delta^*)$, $w(\cdot;a,z,z',\delta') = w(\cdot;a,z,z',\delta^*)$ and $w(\cdot;a,z,z',\delta^*) = w(\cdot;x,y,y',\delta^*) +$ Constant. Therefore, defining $w(\cdot) = w(\cdot;x,y,y',\delta)$ yields the conclusion.[47]

Next, suppose $a \in \Delta$ and $b \in \text{int}(\Delta)$. Because $a\alpha b \in \text{int}(\Delta)$ for $\alpha \in (0,1)$,

$$\begin{aligned}
&V_{PS}(\{a\alpha b, b\}) = \max_{c\in\{a\alpha b, b\}} g\left(c, \max_{c'\in\varphi_r(\{a\alpha b,b\})} w(c')\right)\\
\Leftrightarrow\ &\alpha V_{PS}(\{a,b\}) + (1-\alpha)V_{PS}(\{b\}) = \alpha \max_{c\in\{a,b\}} g\left(c, \max_{c'\in\varphi_r(\{a,b\})} w(c')\right) + (1-\alpha)u(b).
\end{aligned}$$

where the right-hand side follows from the linearity of u, w and r.[48] Letting $\alpha \to 1$ yields the conclusion for V_{PS}. The conclusion for $\mathcal{C}_{PS}$ is obtained analogously using Axiom 5. Proof for the general case with $a, b \in \Delta$ is now straightforward. □

[47] Note δ does not depend on (a,b).

[48] Note that for $x \in \{a,b\}$,

$$\max_{c'\in\varphi_r(\{a,b\}\alpha\{b\})} w(c') - w(x\alpha b) = \alpha\left[\max_{c'\in\varphi_r(\{a,b\})} w(c') - w(x)\right].$$

Therefore,

$$\begin{aligned}
&g\left(x\alpha b, \max_{c'\in\varphi_r(\{a\alpha b,b\})} w(c')\right)\\
&= \alpha u(x) + (1-\alpha)u(b) - \alpha\max\left\{\max_{c'\in\varphi_r(\{a,b\})} w(c') - w(x), 0\right\} + \alpha\beta\max\left\{w(x) - \max_{c'\in\varphi_r(\{a,b\})} w(c'), 0\right\}\\
&= \alpha g\left(x, \max_{c'\in\varphi_r(\{a,b\})} w(c')\right) + (1-\alpha)u(b).
\end{aligned}$$

We now extend the representation to any finite menus.

Lemma 21. *Suppose Axioms 1-8 and Nondegeneracy hold. Then there exists continuous and linear functions u, w and r such that V_{PS} is expressed as*

$$V_{PS}(A) = \max_{c\in A} g\left(c, \max_{c'\in\varphi_r(A)} w(c')\right)$$

and $\mathcal{C}(A)$ coincides with

$$\mathcal{C}_{PS}(A) = \arg\max_{c\in A} g\left(c, \max_{c'\in\varphi_r(A)} w(c')\right)$$

where g and φ are defined above.

Proof. Take any finite set A. By Lemma 3, Lemma 20, and Eq.(11), and because $g(c,\cdot)$ is strictly decreasing at each $c \in \Delta$,

$$\begin{aligned}
V_{PS}(A) &= \min_{b\in\varphi_r(A)} \max_{a\in A} V_{PS}(\{a,b\}) \\
&= \min_{b\in\varphi_r(A)} \max_{a\in A} \max_{c\in\{a,b\}} g\left(c, \max_{c'\in\varphi_r(\{a,b\})} w(c')\right) \\
&= \min_{b\in\varphi_r(A)} \max_{a\in A} \max\left\{ g\left(a, \max_{c'\in\varphi_r(\{a,b\})} w(c')\right), g\left(b, \max_{c'\in\varphi_r(\{a,b\})} w(c')\right)\right\} \\
&= \min_{b\in\varphi_r(A)} \max_{a\in A} g\left(a, \max_{c'\in\varphi_r(\{a,b\})} w(c')\right) \\
&= \max_{a\in A} g\left(a, \max_{c'\in\varphi_r(A)} w(c')\right).
\end{aligned}$$

where the fourth equality holds because $b \in \varphi_r(A)$ implies

$$\max_{a\in A} g\left(b, \max_{c'\in\varphi_r(\{a,b\})} w(c')\right) \leq g(b, w(b)) \leq \max_{a\in A} g\left(a, \max_{c'\in\varphi_r(\{a,b\})} w(c')\right).$$

To prove the second result, we first introduce some lemmas. Lemmas 22 and 23 establish Lemma 24, which then establishes Lemma 21.

Lemma 22. *Suppose Axioms 1-8 hold, so that the representation in Lemma 20 holds for binary menus. If $r(a) = r(b)$ and $w(a) > w(b)$, then $a \succ_w b$.*

Proof. Consider the following exhaustive cases.

Case 1. Suppose $u(a) + w(a) \leq u(b) + w(b)$. Then $V_{PS}(\{a,b\}) = u(b) + w(b) - w(a) < V_{PS}(\{b\})$, so $a \succ_w b$.

Case 2. Suppose $u(a) + w(a) > u(b) + w(b)$ and $u(a) < u(b)$. Then $V_{PS}(\{a,b\}) = u(a) < V_{PS}(\{b\})$, so $a \succ_w b$.

Case 3. Suppose $u(a) + w(a) > u(b) + w(b)$ and $u(a) = u(b)$. Then $V_{PS}(\{a,b\}) = u(a) = V_{PS}(\{b\})$ and $\mathcal{C}(\{a,b\}) = \{a\}$, so $a \succ_w b$.

Case 4. Suppose $u(a) + w(a) > u(b) + w(b)$ and $u(a) > u(b)$. Then $V_{PS}(\{a\}) = V_{PS}(\{a,b\}) > V_{PS}(\{b\})$. Also, since r represents $\succ_r$, $a \sim_r b$. Thus, $a \succ_w b$. □

Lemma 23. *Suppose Axioms 1-8 and Weak Nondegeneracy hold, so that the representation in Lemma 20 holds for binary menus. If $r(a) = r(b)$ and $a \neq b$, then $w(a) \neq w(b)$.*

Proof. By the proof of Claim 1(i), Weak Nondegeneracy implies $r(x) > r(y)$ and $w(x) < w(y)$ for some $x, y \in \Delta$. Now, suppose $r(a) = r(b)$ and $w(a) = w(b)$ for some $a \neq b$. Since the indifference curves for r and w are parallel straight lines, this implies that the indifference curves for r and w are also parallel, which contradicts $r(x) > r(y)$ and $w(x) < w(y)$. □

Lemma 24. *Suppose Axioms 1-8 and Weak Nondegeneracy hold, so that the representation in Lemma 20 holds for binary menus. Then, for any $A \in \mathcal{A}$ and any $b \in \arg\max_{c' \in \varphi_r(A)} w(c')$, we have $b \succeq_r a$ and $b \succeq_w a$ for all $a \in A$.*

Proof. Take any $a \in A \setminus \{b\}$. Because $b \in \varphi_r(A)$, $b \succeq_r a$. If $a \notin \varphi_r(A)$, then $b \succ_r a$, so Axiom 3(i) implies $b \succeq_w a$. If $a \in \varphi_r(A)$, then $r(a) = r(b)$ and $w(a) \leq w(b)$ by definition. By Lemma 23, $w(a) < w(b)$. Thus, $b \succ_w a$ by Lemma 22. □

Proof of Lemma 21 (Continued). We show the following equality in two steps:

$$\mathcal{C}(A) = \arg\max_{c \in A} g\left(c, \max_{c' \in \varphi_r(A)} w(c')\right).$$

Step 1. Take $a \in \arg\max_{c \in A} g\left(c, \max_{c' \in \varphi_r(A)} w(c')\right)$, $b \in \arg\max_{c' \in \varphi_r(A)} w(c')$ and $d \in \mathcal{C}(A)$. By Lemma 24, we have $b \succeq_r a'$ and $b \succeq_w a'$ for all $a' \in A$. By the representation of V_{PS}, $\{a,b,d\} \sim A$. By Axiom 8, $d \in \mathcal{C}(\{a,b,d\})$. Also, by the representation and Lemma 20, we have $\{a,b\} \sim \{a,b,d\}$ and $a \in \mathcal{C}(\{a,b\})$. By Axiom 6(i), we have $a \in \mathcal{C}(\{a,b,d\})$ or $b \in \mathcal{C}(\{a,b,d\})$, so Axiom 8 implies $a \in \mathcal{C}(\{a,b,d\})$. Thus, again by Axiom 8, $a \in \mathcal{C}(A)$.

Step 2. Suppose $d \in \mathcal{C}(A)$. Take $a \in \arg\max_{c \in A} g\left(c, \max_{c' \in \varphi_r(A)} w(c')\right)$ and $b \in \arg\max_{c' \in \varphi_r(A)} w(c')$. If $a = d$, we are done, so suppose not. By step 1, we have $a \in \mathcal{C}(A)$. We also have $b \succeq_r a'$ and $b \succeq_w a'$ for all $a' \in A$, so Axiom 8 implies $d \in \mathcal{C}(\{a, b, d\})$. By Axiom 6(i), $\{b, d\} \sim \{a, b, d\}$. By the representation and definition of a and b,

$$\max_{c \in \{b,d\}} g\left(c, \max_{c' \in \varphi_r(\{b,d\})} w(c')\right) = \max_{c \in \{a,b,d\}} g\left(c, \max_{c' \in \varphi_r(\{a,b,d\})} w(c')\right) = \max_{c \in A} g\left(c, \max_{c' \in \varphi_r(A)} w(c')\right).$$

By Axiom 8, $d \in \mathcal{C}(\{b, d\})$, so Lemma 20 implies $d \in \arg\max_{c \in \{b,d\}} g\left(c, \max_{c' \in \varphi_r(\{b,d\})} w(c')\right)$. Thus $d \in \arg\max_{c \in A} g\left(c, \max_{c' \in \varphi_r(A)} w(c')\right)$. □

To complete the proof of the "only if" part of Theorem 1, we use the following result.

Lemma 25. *Suppose Axioms 1-8 hold. Suppose that for any $A \in \mathcal{A}$, there exists a finite subset A' of A such that (i) $\max_A r = \max_{A'} r$ and (ii) for any finite A'' such that $A' \subset A'' \subset A$, $A'' \sim A'$. Then we have $A \sim A'$.*

Proof. Note that there exists a sequence of finite subsets $\{A_n\}_{n=1}^{\infty}$ of A such that $d_H(A_n, A) \to 0$ as $n \to \infty$ by Lemma 0 in GP. By (ii), $A' \sim A_n \cup A'$ for all n, so $A' \succsim A \cup A' = A$ by Axiom 2(i). To show the opposite relation, note that since A is compact, for every $\epsilon > 0$, there are finite $x_1, \cdots, x_n \in A$ such that $A \subset \cup_{i=1}^{n} N(x_i, \epsilon)$. If $(A' \cup \overline{N(x_i, \epsilon)}) \cap A \succ A$ for all $i = 1, \cdots, n$, then iteratively applying Lemma 2(ii) yields $A = \cup_{i=1}^{n} \{(A' \cup \overline{N(x_i, \epsilon)}) \cap A\} \succ A$, which is a contradiction. Therefore, $A \succsim (A' \cup \overline{N(x_i, \epsilon)}) \cap A$ for some i. Thus, we can take a sequence $\{x_n\}_{n=1}^{\infty}$ in A such that $A \succsim (A' \cup \overline{N(x_n, \frac{1}{n})}) \cap A$ for all $n = 1, 2, \cdots$. Since A is compact, there exists a subsequence $\{x_{n_k}\}_{k=1}^{\infty}$ such that $x_{n_k} \to x^* \in A$. Then letting $k \to \infty$ yields $A \succsim A' \cup \{x^*\} \sim A'$ by Axiom 2(i). □

Proof of the Theorem 1 (Continued).
Take any closed set $A \in \mathcal{A}$, and take $a^* \in \arg\max_{c \in A} g\left(c, \max_{c' \in \varphi_r(A)} w(c')\right)$ and $b^* \in \arg\max_{c' \in \varphi_r(A)} w(c')$. By construction, $\max_{\{a^*, b^*\}} r = \max_A r$. By Lemma 21, $\{a^*, b^*\} \sim A''$ for any finite A'' such that $\{a^*, b^*\} \subset A'' \subset A$. Therefore, by Lemma 25, we have $A \sim \{a^*, b^*\}$. Thus, defining

$$V_{PS}(A) = V_{PS}(\{a^*, b^*\}) = \max_{c \in \{a^*, b^*\}} g\left(c, \max_{c' \in \varphi_r(\{a^*, b^*\})} w(c')\right) = \max_{c \in A} g\left(c, \max_{c' \in \varphi_r(A)} w(c')\right),$$

V_{PS} represents $\succeq$ on $\mathcal{A}$. Also, following the argument for the proof of Lemma 21, we obtain $\mathcal{C}\left(A\right)=\mathcal{C}_{PS}\left(A\right)=\arg\max_{c\in A} g\left(c,\max_{c'\in\varphi_r(A)} w(c')\right)$.

Finally, by construction, $\beta=\frac{1}{\alpha}-1$ if the DM is α-sensitive to shame. Thus, the DM is shame-averse ($\alpha>\frac{1}{2}$) if and only if $\beta<1$, shame-neutral ($\alpha=\frac{1}{2}$) if and only if $\beta=1$ and shame-loving ($\alpha<\frac{1}{2}$) if and only if $\beta>1$.

Proof of "if" part.
Below, we show that a nondegenerate PS representation implies each of the axioms. Proofs of Axiom 1 and Axiom 2c are straightforward and omitted.

To proceed to other axioms, we first note that

$$g(c,R)=u(c)-\max\left\{R-w(c),0\right\}+\beta\max\left\{w(c)-R,0\right\}$$

and

$$G(A,R)=\max_{c\in A} g(c,R)$$

are strictly decreasing and continuous in R and continuous in A.

Axiom 2(i) (Lower Semi-Continuity). Suppose $A\succeq B_n$ and $B_n\to B$. Then, because $\max_{c'\in\varphi_r(B)} w(c')\geq\lim_{n\to\infty}\max_{c'\in\varphi_r(B_n)} w(c')$,

$$\begin{aligned}
V_{PS}(A)&=G\left(A,\max_{c'\in\varphi_r(A)} w(c')\right)\\
&\geq\lim_{n\to\infty} G\left(B_n,\max_{c'\in\varphi_r(B_n)} w(c')\right)\\
&=G\left(B,\lim_{n\to\infty}\max_{c'\in\varphi_r(B_n)} w(c')\right)\\
&\geq G\left(B,\max_{c'\in\varphi_r(B)} w(c')\right)\\
&=V_{PS}\left(B\right)
\end{aligned}$$

where the first inequality follows from $A\succeq B_n$ for all n.

Axiom 2(ii) (Upper vNM continuity). Suppose $A \succ B \succ C$. Note

$$\begin{aligned} V_{PS}\left(A\alpha C\right) &= G\left(A\alpha C, \max_{c' \in \varphi_r(A\alpha C)} w(c')\right) \\ &= G\left(A\alpha C, \alpha \max_{c' \in \varphi_r(A)} w(c') + (1-\alpha) \max_{c' \in \varphi_r(C)} w(c')\right) \end{aligned}$$

By the continuity of G, $V_{PS}(A\alpha C) \approx V_{PS}(C) < V_{PS}(B)$ for sufficiently small $\alpha \in (0,1)$.

We now introduce some lemmas.

Lemma 26. *Suppose the data are generated by a nondegenerate PS preference. If $a \succ_w b$, then $r(a) \geq r(b)$ and $w(a) > w(b)$.*

Proof. Consider the following exhaustive cases.

Case 1. Suppose $\{b\} \succ \{a,b\}$. Then

$$u(b) > \max_{c \in \{a,b\}} g\left(c, \max_{c' \in \varphi_r(\{a,b\})} w(c')\right) \geq g\left(b, \max_{c' \in \varphi_r(\{a,b\})} w(c')\right),$$

so $\max_{c' \in \varphi_r(\{a,b\})} w(c') > w(b)$, yielding the conclusion.

Case 2. Suppose $\{b\} \sim \{a,b\}$ and $\mathcal{C}(\{a,b\}) = \{a\}$. Then

$$u(b) = g\left(a, \max_{c' \in \varphi_r(\{a,b\})} w(c')\right) > g\left(b, \max_{c' \in \varphi_r(\{a,b\})} w(c')\right),$$

so the conclusion holds as in Case 1.

Case 3. Suppose $a \sim_r b$ and $\{a\} \sim \{a,b\} \succ \{b\}$. Then $\max_{c' \in \varphi_r(\{a,b\})} w(c') = \max_{c' \in \{a,b\}} w(c') \geq w(b)$, so we must have $\mathcal{C}(\{a,b\}) = \{a\}$. Therefore,

$$u(a) = \max_{c \in \{a,b\}} g\left(c, \max_{c' \in \{a,b\}} w(c')\right).$$

This in turn implies $\max_{c' \in \{a,b\}} w(c') = w(a)$, so $w(a) \geq w(b)$. By Nondegeneracy and $r(a) = r(b)$, we have $w(a) > w(b)$. □

Lemma 27. *Suppose the data are generated by a PS preference. If $A \cup \{a\} \succ A$ and $a \notin \mathcal{C}\left(A \cup \{a\}\right)$, then $w(a) = \max_{c' \in \varphi_r(A \cup \{a\})} w(c') < \max_{c' \in \varphi_r(A)} w(c')$.*

Proof. Suppose $A \cup \{a\} \succ A$ and $a \notin \mathcal{C}\left(A \cup \{a\}\right)$. If $\max_{c' \in \varphi_r(A \cup \{a\})} w(c') \geq$

$\max_{c' \in \varphi_r(A)} w(c')$, then

$$\begin{aligned} V_{PS}(A \cup \{a\}) &= G\left(A \cup \{a\}, \max_{c' \in \varphi_r(A\cup\{a\})} w(c')\right) \\ &= G\left(A, \max_{c' \in \varphi_r(A\cup\{a\})} w(c')\right) \\ &\leq G\left(A, \max_{c' \in \varphi_r(A)} w(c')\right) \\ &= V_{PS}(A) \end{aligned}$$

where the second equality follows from $a \notin \mathcal{C}(A \cup \{a\})$. This is a contradiction. Thus $\max_{c' \in \varphi_r(A\cup\{a\})} w(c') < \max_{c' \in \varphi_r(A)} w(c')$, and we must have $\max_{c' \in \varphi_r(A\cup\{a\})} w(c') = w(a)$.[49] □

Proof of "if" part (Continued).
Axiom 3(i). If $a \succ_r b$, then Theorem 2 (to be proven below) implies $r(a) > r(b)$. Thus, Theorem 2 again implies $b \not\succ_r a$, and Lemma 26 implies $b \not\succ_w a$. Similarly, if $a \succ_w b$, then Lemma 26 implies $r(a) \geq r(b)$ and $w(a) > w(b)$, so we cannot have $b \succ_r a$ or $b \succ_w a$.

Axiom 3(ii). Suppose $a \succ^* b \succ^* c$. By definition, there exist $A \ni b$ and $B \ni c$ such that $A \cup \{a\} \succ A$, $a \notin \mathcal{C}(A \cup \{a\})$, $B \cup \{b\} \succ B$, and $b \notin \mathcal{C}(B \cup \{b\})$. Now, let $C = A \cup B$. Because $r(a) > r(a')$ for all $a' \in A$ and $r(b) > r(b')$ for all $b' \in B$, we have $r(a) > r(c')$ for all $c' \in C$. Then, together with Lemma 27, we have $w(a) < \max_{c' \in \varphi_r(A)} w(c') = \max_{c' \in \varphi_r(C)} w(c')$. Therefore, the representation implies $a \notin \arg\max_{d \in C \cup \{a\}} g(d, w(a)) = \mathcal{C}(C \cup \{a\})$ and $C \cup \{a\} \succ C$. Thus, $a \succ^* c$.

Next, if $a \sim_r b \sim_r c$, Theorem 2 (to be shown below) implies $[a \succeq_w b] \wedge [b \succeq_w c] \Leftrightarrow [w(a) \geq w(b)] \wedge [w(b) \geq w(c)] \Rightarrow w(a) \geq w(c) \Leftrightarrow a \succeq_w c$.

Axiom 3(iii-a). Suppose $A\alpha_n C \succeq B$ and $\alpha_n \to \alpha$. Because

$$\begin{aligned} \lim_{n\to\infty} \max_{c' \in \varphi_r(A\alpha_n C)} w(c') &= \lim_{n\to\infty} \left[\alpha_n \max_{c' \in \varphi_r(A)} w(c') + (1-\alpha_n) \max_{c' \in \varphi_r(C)} w(c')\right] \\ &= \max_{c' \in \varphi_r(A\alpha C)} w(c'), \end{aligned}$$

[49]Note that for any $A, B \in \mathcal{A}$, $\varphi_r(A \cup B) \in \{\varphi_r(A), \varphi_r(B), \varphi_r(A) \cup \varphi_r(B)\}$.

we have

$$
\begin{aligned}
V_{PS}\left(A\alpha C\right) &= G\left(A\alpha C, \max_{c'\in\varphi_r(A\alpha C)} w(c')\right) \\
&= \lim_{n\to\infty} G\left(A\alpha_n C, \max_{c'\in\varphi_r(A\alpha_n C)} w(c')\right) \\
&\geq G\left(B, \max_{c'\in\varphi_r(B)} w(c')\right) \\
&= V_{PS}(B)
\end{aligned}
$$

where the inequality follows from $V_{PS}(B) \leq V_{PS}\left(A\alpha_n C\right)$ for all n.

Axiom 3(iii-b). Suppose $a^* \in A$ is such that $a^* \succ_r a$ for all $a \in A \setminus \{a^*\}$. Take any $(A_n)_n$ and any $(a_n)_n$ such that $A_n \to A$, $a_n \in \mathcal{C}(A_n)$ and $a_n \to a$. By Theorem 2, $r(a^*) > r(a)$ for all $a \in A\setminus\{a^*\}$, so $\lim_{n\to\infty} \max_{c'\in\varphi_r(A_n)} w(c') = \max_{c'\in\varphi_r(A)} w(c') = w(a^*)$. By continuity, $g\left(c, \max_{c'\in\varphi_r(A_n)} w(c')\right) \to g\left(c, w(a^*)\right)$ for all c. By $a_n \in \mathcal{C}(A_n)$, we have

$$
g\left(a_n, \max_{c'\in\varphi_r(A_n)} w(c')\right) = \max_{c\in A_n} g\left(c, \max_{c'\in\varphi_r(A_n)} w(c')\right),
$$

so letting $n \to \infty$ yields $g\left(a, \max_{c'\in\varphi_r(A)} w(c')\right) = \max_{c\in A} g\left(c, \max_{c'\in\varphi_r(A)} w(c')\right)$. Thus, $a \in \mathcal{C}(A)$.

Axiom 3(iv). By Lemma 30 below, $a\alpha c \succ^* b\alpha c \Rightarrow r(a) > r(b) \Rightarrow a \succ^* b$.

To prove some of the remaining axioms, we need a lemma.

Lemma 28. *Suppose the choice data are generated by a PS preference. (i) If $b \in \mathcal{P}(a)$, then $w(a) < w(b)$. (ii) If $b \in \mathcal{S}(a)$, then $w(a) > w(b)$. (iii) If $b \in \mathcal{N}_1(a)$, then $w(a) = w(b)$.*

Proof. (i) By the representation and the definition of $\mathcal{P}(a)$, $g(b, w(a)) > u(b) = g(b, w(b))$, so $w(a) < w(b)$. (ii) If $b \in \mathcal{S}(a)$, $g(b, w(a)) < g(b, w(b))$, so $w(a) > w(b)$. (iii) If $b \in \mathcal{N}_1(a)$, $g(b, w(a)) = g(b, w(b))$, so $w(a) = w(b)$. □

Proof of Theorem 1 (Continued).

Axiom 4(i). We only provide the proof for the case where all menus are binary; proof is analogous and simpler if a singleton is involved. Suppose

$b \in \mathcal{P}(a) \cup \mathcal{N}_1(a)$, $d \in \mathcal{P}(c) \cup \mathcal{N}_1(c)$ and $f \in \mathcal{P}(e) \cup \mathcal{N}_1(e)$. By Lemma 28(i), $r(a) > r(b)$, $w(a) \leq w(b)$, $r(e) > r(f)$ and $w(e) \leq w(f)$. Therefore,

$$\begin{aligned}
&V_{PS}(\alpha \{a,b\} \alpha \{e,f\}) \\
&= \max_{x \in \{a,b\}\alpha\{e,f\}} \Bigg[u(x) - \max\left\{\max_{y \in \varphi(\{a,b\}\alpha\{e,f\})} w(y) - w(x), 0\right\} \\
&\qquad\qquad + \beta \max\left\{w(x) - \max_{y \in \varphi(\{a,b\}\alpha\{e,f\})} w(y), 0\right\}\Bigg] \\
&= \max_{x \in \{a,b\}\alpha\{e,f\}} \left[u(x) + \beta\left(w(x) - w(a\alpha e)\right)\right] \\
&= \alpha \max_{x \in \{a,b\}} \left[u(x) + \beta\left(w(x) - w(a)\right)\right] + (1-\alpha) \max_{x \in \{e,f\}} \left[u(x) + \beta\left(w(x) - w(e)\right)\right] \\
&= \alpha V_{PS}(\{a,b\}) + (1-\alpha) V_{PS}(\{e,f\}).
\end{aligned}$$

Similarly, $V_{PS}(\{c,d\} \alpha \{e,f\}) = \alpha V_{PS}(\{c,d\}) + (1-\alpha)V_{PS}(\{e,f\})$. Therefore, $\{a,b\} \succ (\succeq) \{c,d\}$ implies $\{a,b\} \alpha \{e,f\} \succ (\succeq) \{c,d\} \alpha \{e,f\}$.

Axiom 4(ii). Take any $\{a,b\}, \{e,f\} \in \mathcal{B}_S$ such that $a \neq b$ and $e \neq f$. By Lemma 28(ii) we can assume without loss of generality that $r(a) \geq r(b)$, $w(a) \geq w(b)$, $r(e) \geq r(f)$ and $w(e) \geq w(f)$. Then, similarly to above,

$$\begin{aligned}
V_{PS}(\{a,b\} \alpha \{e,f\}) &= \max_{x \in \{a,b\}\alpha\{e,f\}} \left[u(x) + w(x) - w(a\alpha e)\right] \\
&= \alpha V_{PS}(\{a,b\}) + (1-\alpha) V_{PS}(\{e,f\}).
\end{aligned}$$

Therefore, the conclusion of Axiom 4(ii) holds.

Axiom 4(iii). Note that for any $x \in A$,

$$\max_{c' \in \varphi_r(A\alpha\{c\})} w(c') - w(x\alpha c) = \alpha \left[\max_{c' \in \varphi_r(A)} w(c') - w(x)\right].$$

Therefore,

$$\begin{aligned}
&g\left(x\alpha c, \max_{c' \in \varphi_r(A\alpha\{c\})} w(c')\right) \\
&= \alpha u(x) + (1-\alpha)u(c) - \alpha \max\left\{\max_{c' \in \varphi_r(A)} w(c') - w(x), 0\right\} + \alpha\beta \max\left\{w(x) - \max_{c' \in \varphi_r(A)} w(c'), 0\right\} \\
&= \alpha g\left(x, \max_{c' \in \varphi_r(A)} w(c')\right) + (1-\alpha)u(c).
\end{aligned}$$

Thus,

$$\begin{aligned}V_{PS}\left(A\alpha\left\{c\right\}\right) &= \max_{x\in A} g\left(x\alpha c, \max_{c'\in\varphi_r(A\alpha\{c\})} w(c')\right)\\ &= \alpha \max_{x\in A} g\left(x, \max_{c'\in\varphi_r(A)} w(c')\right) + (1-\alpha)u(c)\\ &= \alpha V_{PS}\left(A\right) + (1-\alpha)V_{PS}\left(\left\{c\right\}\right),\end{aligned}$$

which implies the conclusion.

Axiom 5(i). Consider first the case where $b \in \mathcal{P}(a) \cup \mathcal{N}_1(a)$ and $d \in \mathcal{P}(c) \cup \mathcal{N}_1(c)$. Following the argument in the proof of Axiom 4(i),

$$\begin{aligned}&\mathcal{C}(\{a,b\}\,\alpha\,\{c,d\})\\ &= \underset{x\in\{a,b\}\alpha\{c,d\}}{\arg\max}\left[u(x) + \beta\left(w(x) - w(a\alpha c)\right)\right]\\ &= \alpha \underset{x\in\{a,b\}}{\arg\max}\left[u(x) + \beta\left(w(x) - w(a)\right)\right] + (1-\alpha)\underset{x\in\{c,d\}}{\arg\max}\left[u(x) + \beta\left(w(x) - w(c)\right)\right]\\ &= \mathcal{C}(\{a,b\})\alpha\mathcal{C}(\{c,d\}).\end{aligned}$$

Proof for the case $b \in \mathcal{S}(a) \cup \mathcal{N}_1(a) \cup \mathcal{I}(a)$ and $d \in \mathcal{S}(c) \cup \mathcal{N}_1(c) \cup \mathcal{I}(a)$ is analogous: letting $w(a) \geq w(b)$ and $w(c) \geq w(d)$ without loss of generality and following the proof of Axiom 4(ii),

$$\begin{aligned}\mathcal{C}(\{a,b\}\,\alpha\,\{c,d\}) &= \underset{x\in\{a,b\}\alpha\{c,d\}}{\arg\max}\left[u(x) + w(x) - w(a\alpha c)\right]\\ &= \mathcal{C}(\{a,b\})\alpha\mathcal{C}(\{c,d\}).\end{aligned}$$

Axiom 5(ii). For (ii-a), suppose $A = \{a,b\}\,\alpha\,\{a,c\}$, $b \in \mathcal{N}_2(a)$, $c \in \mathcal{P}(a)$, $\{a, b\alpha c\} \succeq \{b\}\,\alpha\,\{a,c\}$, and $b\alpha c \in \mathcal{C}\left(\{a, b\alpha c\}\right)$. By the PS representation,

$$\begin{aligned}&u(b\alpha c) - \max\{w(a) - w(b\alpha c), 0\} + \beta\max\{w(b\alpha c) - w(a), 0\}\\ &= V_{PS}(\{a, b\alpha c\})\\ &\geq V_{PS}(\{b\}\,\alpha\,\{a,c\})\\ &\geq u(b\alpha c) + \beta\left(w(b\alpha c) - w(b\alpha a)\right).\end{aligned}$$

Therefore, we must have $w(b\alpha a) \geq w(a)$, so $w(b) \geq w(a)$. Because $r(a) > r(b), r(c)$ and $w(a) \leq w(b), w(c)$, the same argument as the proof of Axiom 5(i) yields $\mathcal{C}\left(A\right) = \mathcal{C}\left(\{a,b\}\right)\alpha\mathcal{C}\left(\{a,c\}\right)$. Proof for (ii-b) is analogous: the assumptions imply

$$u(b\alpha c) + w(b\alpha c) - w(b\alpha a) \geq u(b\alpha c) + w(b\alpha c) - w(a)$$

and therefore $w(a) \geq w(b), w(c)$, so the argument in the proof of Axiom 4(ii) yields the result.

Axiom 5(iii). Proof for is analogous to that of Axiom 4(iii), so omitted.

Axiom 6(i). Suppose there exists $a^* \in A$ such that $a^* \succeq_r c$ and $a^* \succeq_w c$ for all $c \in A \cup B$. Then, by Theorem 2, $\max_{y \in \varphi_r(A \cup B)} w(y) = \max_{y \in \varphi_r(A)} w(y) = w(a^*)$. Therefore,

$$V_{PS}(A \cup B) = \max_{x \in A \cup B} g\left(x, w(a^*)\right) \geq \max_{x \in A} g\left(x, w(a^*)\right) = V_{PS}(A)$$

and the inequality is strict if and only if $\arg\max\limits_{x \in A \cup B} g\left(x, w(a^*)\right) \cap A = \emptyset$.

Axiom 6(ii). Suppose there exists $a^* \in A$ such that $a^* \succeq_r c$ for all $c \in A \cup B$ and $a^* \succeq_w a$ for all $a \in A$, and there exists $b^* \in B$ such that $b^* \succ_w a^*$. By Lemma 26 and Theorem 2, $r(b^*) = r(a^*) \geq r(b)$ for all $b \in B$ and $w(b^*) > w(a^*)$. Without loss of generality, let b^* be a maximizer of $\succeq_w$ on $\varphi_r(B)$. Then by Themrem 2, $\max_{y \in \varphi_r(A \cup B)} w(y) = w(b^*) > w(a^*) = \max_{y \in \varphi_r(A)} w(y)$. Therefore, if there exists $c \in \mathcal{C}(A \cup B) \cap A$, then

$$V_{PS}(A \cup B) = g\left(c, w(b^*)\right) < g\left(c, w(a^*)\right) \leq V_{PS}(A).$$

Axiom 7. Take any $a, b, c, d \in \Delta$ such that $c \in \mathcal{P}(a) \cap \mathcal{P}(b)$ and $d \in \mathcal{S}(a) \cap \mathcal{S}(b)$. Then,

$$\begin{aligned} V_{PS}(\{a, c\}) + \beta V_{PS}(\{b, d\}) &= u(c) + \beta\left(w(c) - w(a)\right) + \beta\left[u(d) + w(d) - w(b)\right] \\ &= u(c) + \beta\left(w(c) - w(b)\right) + \beta\left[u(d) + w(d) - w(a)\right] \\ &= V_{PS}(\{b, c\}) + \beta V_{PS}(\{a, d\}). \end{aligned}$$

Therefore, by letting $\alpha = \frac{1}{1+\beta} \in (0,1)$,

$$\begin{aligned} V_{PS}\left(\{a,c\}\,\alpha\left\{e^{b,d}\right\}\right) &= \alpha V_{PS}\left(\{a,c\}\right) + (1-\alpha)V_{PS}\left(\left\{e^{b,d}\right\}\right) \\ &= \alpha V_{PS}\left(\{a,c\}\right) + (1-\alpha)V_{PS}\left(\{b,d\}\right) \\ &= \alpha V_{PS}\left(\{b,c\}\right) + (1-\alpha)V_{PS}\left(\{a,d\}\right) \\ &= \alpha V_{PS}\left(\{b,c\}\right) + (1-\alpha)V_{PS}\left(\left\{e^{a,d}\right\}\right) \\ &= V_{PS}\left(\{b,c\}\,\alpha\left\{e^{a,d}\right\}\right). \end{aligned}$$

Axiom 8. Suppose there exists $a^* \in A \cap B$ such that $a^* \succeq_r c$ and $a^* \succeq_w c$ for all $c \in A \cup B$. By Theorem 2, $\max_{y \in \varphi_r(A)} w(y) = \max_{y \in \varphi_r(B)} w(y) = w(a^*)$. Now, suppose $a, b \in A \cap B$, $a \in \mathcal{C}(A)$ and $b \in \mathcal{C}(B)$. By the PS representation, $g(a, w(a^*)) = g(b, w(a^*)) \geq g(c, w(a^*))$ for all $c \in A \cup B$. Therefore, $a \in \mathcal{C}(B)$.

A.2 Proof of Theorem 2

Lemma 29. *Suppose the data are generated by a nondegenerate PS preference. Then,*

$$P_1(a) = \{c \in \Delta : r(a) > r(c) \text{ and } u(a) < g(c, w(a))\}$$

and

$$P_2(a) = \{c \in \Delta : r(a) > r(c) \text{ and } w(a) < w(c)\}$$

are nonempty for all $a \in \text{int}(\Delta)$.

Proof. Take $\bar{a}, \bar{b} \in \Delta$ such that $\bar{a} \succ^* \bar{b}$. Then, there exist $A \ni \bar{b}$ and $c \neq \bar{a}$ such that $A \cup \{\bar{a}\} \succ A$, $\bar{a} \notin \mathcal{C}(A \cup \{\bar{a}\})$, and $c \in \mathcal{C}(A \cup \{\bar{a}\})$. By Lemma 27, $w(a) = \max_{y \in \varphi_r(A \cup \{\bar{a}\})} w(y) < \max_{y \in \varphi_r(A)} w(y)$, so we must have $\varphi_r(A \cup \{\bar{a}\}) = \{\bar{a}\}$, i.e., $r(\bar{a}) > r(a)$ for all $a \in A$. Therefore, there exists some $d \in A$ such that $w(d) = \max_{y \in \varphi_r(A)} w(y) > w(\bar{a})$ and $r(d) < r(\bar{a})$. Thus, $d \in P_2(\bar{a})$. Also, the PS representation and $\bar{a} \notin \mathcal{C}(A \cup \{\bar{a}\})$ imply $g(c, w(\bar{a})) > g(\bar{a}, w(\bar{a})) = u(\bar{a})$, so $c \in P_1(\bar{a})$.

Now, take any $a \in \text{int}(\Delta)$. There exist some $\alpha \in (0,1]$ and $e \in \Delta$ such that $a = \bar{a}\alpha e$.[50] By the linearity of u, w, and r, we have $r(a) > \max\{r(c\alpha e), r(d\alpha e)\}$, $g(c\alpha e, w(a)) > g(a, w(a)) = u(a)$, and $w(a) < w(d\alpha e)$. Therefore, $c\alpha e \in P_1(a)$ and $d\alpha e \in P_2(a)$. □

[50] To see this, note first that by taking $e = a - \varepsilon(\bar{a} - a)$ where $\varepsilon > 0$ is an arbitrary scalar, we obtain $a = \bar{a}\frac{\varepsilon}{1+\varepsilon}e$. By construction, we have $\sum_{z \in Z} e(z) = 1$. Because $a \in \text{int}(\Delta)$, we have $a(z) > 0$ for all $z \in Z$. Because $|Z| < \infty$, $e \in \text{int}(\Delta)$ for sufficiently small ε.

$c \in P_1(a)$ is an alternative which is below a in the descriptive norm ranking r but which is a choice preferred to a. $d \in P_2(a)$ is an alternatives which is below a in the descriptive norm ranking but above a in the prescriptive norm ranking w. Therefore, if $d \in A \cap P_2(a)$ sets the reference point at menu A, then a sets the reference point at $A \cup \{a\}$ which is lower than the reference point at A. Moreover, if $c \in A \cap P_1(a)$, then a is not chosen from $A \cup \{a\}$. Such c, d are key to establishing $a \succ^* b$ for any $b \in A$. Lemma 30 formalizes the idea. See also the graphical illustration in Figure A1 and discussions in Appendix B.

Lemma 30. *Suppose the data are generated by a nondegenerate PS preference. Then, for any $a, b \in \Delta$, $a \succ^* b$ implies $r(a) > r(b)$. Moreover, if $a \in \text{int}(\Delta)$, then $r(a) > r(b)$ implies $a \succ^* b$.*

Proof. Suppose $a \succ^* b$. Then there exists $A \ni b$ such that $A \cup \{a\} \succ A$ and $a \notin \mathcal{C}(A \cup \{a\})$. By Lemma 27, we must have $\varphi_r(A \cup \{a\}) = \{a\}$, hence $r(a) > r(b)$. Next, suppose $a \in \text{int}(\Delta)$ and $r(a) > r(b)$. By Lemma 29, there exists $c \in P_1(a)$ and $d \in P_2(a)$. Note that $d \in P_2(a)$ implies $\alpha d + (1-\alpha)a \in P_2(a)$ for all $\alpha \in (0,1)$. Therefore, we can assume without loss of generality that $r(a) > r(d) > \max\{r(b), r(c)\}$. Then,

$$\begin{aligned} V_{PS}(\{a,b,c,d\}) &= G\left(\{a,b,c,d\}, \max_{y \in \varphi_r(\{a,b,c,d\})} w(y)\right) \\ &= G(\{a,b,c,d\}, w(a)) \\ &> G(\{a,b,c,d\}, w(d)) \\ &\geq G(\{b,c,d\}, w(d)) \\ &= V_{PS}(\{b,c,d\}) \end{aligned}$$

where the strict inequality follows from $w(a) < w(d)$. Moreover, because $c \in P_1(a)$, we have $g(c, w(a)) > u(a) = g(a, w(a))$ so $a \notin \mathcal{C}(\{a,b,c,d\})$. Thus, $a \succ^* b$. □

Proof of Theorem 2 (Continued).
(i) Suppose first $r(a) > r(b)$. Take some $c \in \text{int}(\Delta)$ such that $r(a) > r(c) > r(b)$.[51] By Lemma 30, $c \succ^* b$. Also, if $c \succ^* a$, then $r(c) > r(a)$ by Lemma 30, a contradiction. Therefore, $c \not\succ^* a$, hence $a \succ_r b$. Next, suppose $a \succ_r b$. If $a \succ^* b$, we have $r(a) > r(b)$ by Lemma 30. Now, consider the case in

[51] Take $c' = a\alpha b$ for some $\alpha \in (0,1)$. If $c' \in \text{int}(\Delta)$, let $c = c'$. Otherwise, take some $d \in \text{int}(\Delta)$ and let $c = \beta c' + (1-\beta)d$ where $\beta < 1$ is sufficiently close to 1.

which we have $c \not\succ^* a$ and $c \succ^* b$ for some $c \in \text{int}(\Delta)$. By Lemma 30, $r(c) > r(b)$. Also, if $r(c) > r(a)$, then $c \succ^* a$ by Lemma 30, a contradiction. Thus $r(a) \geq r(c) > r(b)$.

(ii) By Lemma 26, $a \succ_w b$ implies $w(a) > w(b)$. The converse follows from Lemma 22. □

A.3 Other Proofs

A.3.1 Proof of Proposition 1

It is easy to show that (ii) implies (i), so we only prove that (i) implies (ii). Let V_{PS} and V'_{PS} denote the function which represents $\succeq$ using (u, w, r, β) and (u', w', r', β'), respectively. Since V_{PS} is unique up to positive affine transformation, $u'(x) = V'_{PS}(\{x\}) = \alpha V_{PS}(\{x\}) + \gamma_u = \alpha u(x) + \gamma_u$ for some $\alpha > 0$ and $\gamma_u \in \mathbb{R}$. Now, by nondegeneracy, there exists $x, y \in \Delta$ such that $\{x, y\} \succ \{y\}$ and $\mathcal{C}(\{x, y\}) = \{y\}$. By the construction of w,[52] for any $z \in \Delta$,

$$\begin{aligned}
u'(z) + w'(z) - w'(x) &= \frac{1}{\delta} V'_{PS}(\{x, (1-\delta)y + \delta z\}) - \frac{1-\delta}{\delta} V'_{PS}(\{x, y\}) \\
&= \frac{1}{\delta} [\alpha V_{PS}(\{x, (1-\delta)y + \delta z\}) + \gamma_u] - \frac{1-\delta}{\delta} [\alpha V_{PS}(\{x, y\}) + \gamma_u] \\
&= \alpha \left[\frac{1}{\delta} V_{PS}(\{x, (1-\delta)y + \delta z\}) - \frac{1-\delta}{\delta} V_{PS}(\{x, y\})\right] + \gamma_u \\
&= \alpha \{u(z) + w(z) - w(x)\} + \gamma_u.
\end{aligned}$$

Since $u'(z) = \alpha u(z) + \gamma_u$, $w'(z) = \alpha w(z) - \alpha w(x) + w'(x) \equiv \alpha w(z) + \gamma_w$. Next, by Theorem 2, both r and r' represent $\succeq_r$, so Lemma 1 implies $r' = \alpha_r r + \gamma_r$ for some $\alpha_r > 0$ and $\gamma_r \in \mathbb{R}$. Finally, because $(\succeq, \mathcal{C})$ is α-sensitive to shame for a unique $\alpha \in (0, 1)$, we have $\beta = \frac{1}{\alpha} - 1 = \beta'$ by construction. □

Proof of Proposition 2. Equivalence of (i) and (ii) follows from the construction of β in the proof of Theorem 1. To show the equivalence of (ii) and (iii), note that if DM i is α_i-sensitive to shame, then the PS representation implies

$$\alpha_i = \frac{V^i_{PS}(\{a, d\}) - V^i_{PS}(\{b, d\})}{V^i_{PS}(\{a, c\}) - V^i_{PS}(\{b, c\}) + V^i_{PS}(\{a, d\}) - V^i_{PS}(\{b, d\})}.$$

[52] Recall that the following result does not depend on the choice of x, y.

Then

$$
\begin{aligned}
& \alpha\{b,c\}+(1-\alpha)\left\{e_i^{a,d}\right\} \succeq_i \alpha\{a,c\}+(1-\alpha)\left\{e_i^{b,d}\right\} \\
\Leftrightarrow\ & \alpha V_{PS}^i(\{a,c\})+(1-\alpha)V_{PS}^i(\{b,d\}) \leq \alpha V_{PS}^i(\{b,c\})+(1-\alpha)V_{PS}^i(\{a,d\}) \\
\Leftrightarrow\ & \alpha \leq \alpha_i .
\end{aligned}
$$

Therefore, $\alpha_1 > (\geq)\alpha_2$ if and only if $\alpha\{b,c\}+(1-\alpha)\left\{e_2^{a,d}\right\} \succeq_2 \alpha\{a,c\}+(1-\alpha)\left\{e_2^{b,d}\right\}$ implies $\alpha\{b,c\}+(1-\alpha)\left\{e_1^{a,d}\right\} \succ_1 (\succeq_1)\alpha\{a,c\}+(1-\alpha)\left\{e_1^{b,d}\right\}$.
□

B Graphical Illustrations of Nondegeneracy and $\succ_r$

We provide graphical illustrations of the nondegeneracy concepts and the elicitation of $a \succ_r b$. Figure A1 illustrates the concepts of nondegeneracy and weak nondegeneracy, providing an example to distinguish the two. It also shows why, in Definition 1, $a \succ^* b$ needs to be elicited using a general menu $A \ni b$ and not just $A = \{b\}$, and presents a graphical illustration of Theorem 2. Figure A2 then demonstrates how Definition 1(ii-b) helps establish $a \succ_r b$ when we cannot establish the relation via condition (ii-a), i.e., $a \succ^* b$, which occurs when a is on the boundary of Δ.

Figure A1a illustrates nondegeneracy, which requires that there exist $x, y, y' \in \Delta$ such that $y \in \mathcal{P}(x)$ and $y' \in \mathcal{S}(x)$. For $\mathcal{P}(x)$ to be nonempty, we must have some $y \in \Delta$ such that $r(x) > r(y)$, $w(x) < w(y)$, and $g(x, w(x)) < g(y, w(x))$. The first two condition ensures that the reference point at $\{x, y\}$ is lower than that at $\{y\}$, and adding the third condition ensures that x is not chosen from $\{x, y\}$. Similarly, for $\mathcal{S}(x)$ to be nonempty, we must have some $y' \in \Delta$ such that $r(x) > r(y')$, $w(x) > w(y')$, and $g(x, w(x)) < g(y', w(x))$, ensuring that the reference point at $\{x, y'\}$ is higher than that at $\{y'\}$ and that x is not chosen from $\{x, y'\}$.

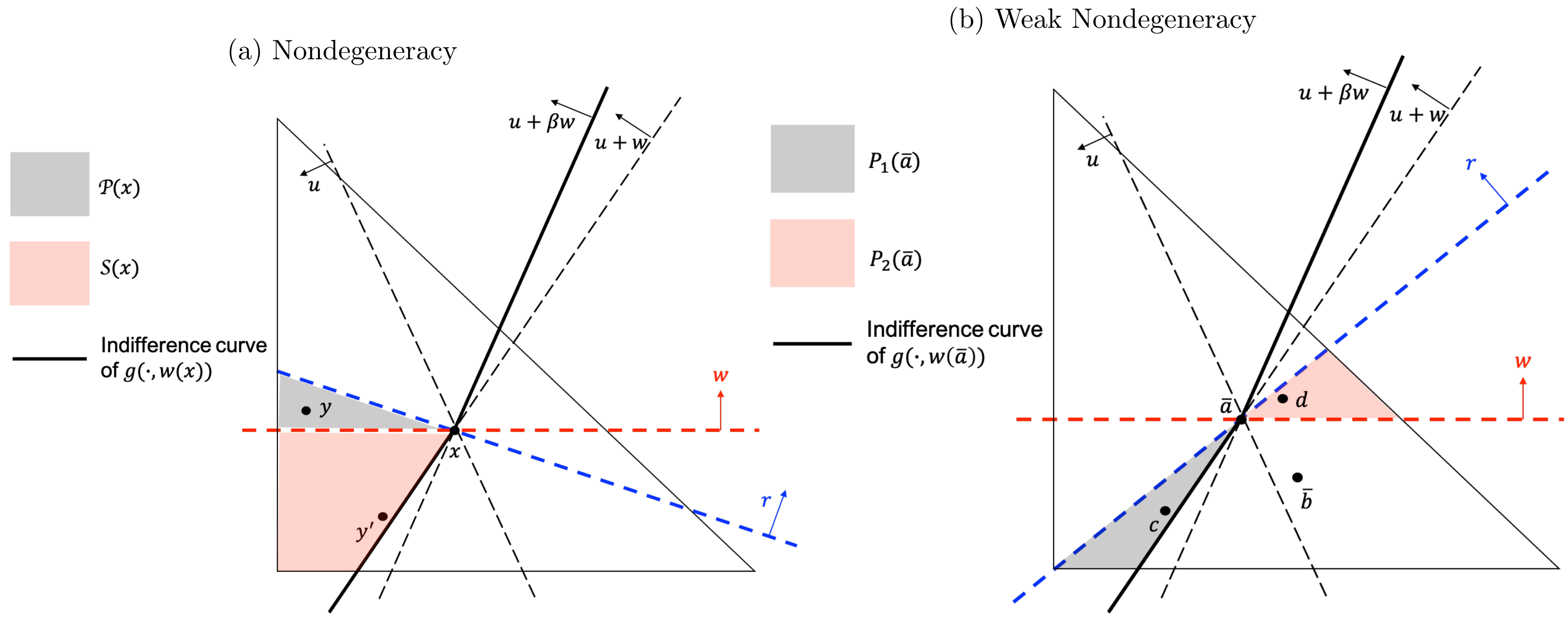
Figure A1b provides an example in which the nondegeneracy property is violated. To see this, note that for any $y \in \Delta$ such that $r(\bar{a}) > r(y)$ and $w(\bar{a}) < w(y)$, we have $g(\bar{a}, w(\bar{a})) > g(y, w(\bar{a}))$, so $\mathcal{P}(\bar{a})$ is empty. Since this observation holds for any $\bar{a} \in \Delta$, the nondegeneracy axiom is violated. In this case, the reference-lowering alternative $\bar{a}$ is also the chosen one, so observing $\{\bar{a}, y\} \succ \{y\}$ does not allow us to tell if the larger menu is preferred because $\bar{a}$ lowers the reference point or because $\bar{a}$ is a preferred choice.

However, the preference illustrated in Figure A1b satisfies the weak non-degeneracy axiom. To see this, note that (i) the reference point at $\{\bar{a}, \bar{b}, c, d\}$, $w(\bar{a})$, is lower than the reference point at $\{\bar{b}, c, d\}$, $w(d)$, and (ii) $\mathcal{C}(\{\bar{b}, c, d\}) = \mathcal{C}(\{\bar{a}, \bar{b}, c, d\}) = \{c\}$. In this case, $\bar{a}$ makes the larger menu more desirable even though it is not chosen there, which implies it sets the reference point lower than the reference point at $\{\bar{b}, c, d\}$. Therefore, we have $\bar{a} \succ^* \bar{b}$. This example shows why we cannot confine Definition 1(i) to $A = \{\bar{b}\}$: Even if $r(\bar{a}) > r(\bar{b})$, we may have $\bar{a} \in \mathcal{C}(\{\bar{a}, \bar{b}\})$ ($\bar{a}$ is chosen) or $w(\bar{a}) \geq w(\bar{b})$ ($\bar{a}$ sets a weakly higher reference point), preventing us from concluding $\bar{a} \succ^* \bar{b}$ with $A = \{\bar{b}\}$. Thus, to conclude $\bar{a} \succ^* \bar{b}$, we may need a larger menu A which contains a "choice fixer" $c \in P_1(\bar{a})$ and a "higher reference setter" $d \in P_2(\bar{a})$. The figure also graphically illustrates Theorem 2, in particular that we observe $\bar{a} \succ^* \bar{b}$ whenever $r(\bar{a}) > r(\bar{b})$, as long as $P_1(\bar{a})$ and $P_2(\bar{a})$ are nonempty (which is a quite weak condition).

Figure A2 illustrates how we can elicit $\bar{a} \succ_r \bar{b}$ when $\bar{a}$ cannot satisfy $\bar{a} \succ^* \bar{b}$ even though data are generated by a PS preference with $r(\bar{a}) > r(\bar{b})$. Figure A2a depicts the indifference curves of the same PS preference as in Figure A1b. However, because $\bar{a}$ is on the boundary of Δ, $P_2(\bar{a})$ is empty, and we cannot establish $\bar{a} \succ^* \bar{b}$ with any A. Intuitively, when $\bar{a}$ is a unique reference alternative at $A \cup \{\bar{a}\}$ (i.e., $r(\bar{a}) > r(y)$ for all $y \in A$), the reference point is necessarily higher than that at A, preventing $\bar{a} \succ^* \bar{b}$.

However, we can still conclude $r(\bar{a}) > r(\bar{b})$ if there exists some $c \in \Delta$ such that $r(\bar{a}) \geq r(c) > r(\bar{b})$. As Figure A2b demonstrates, by taking $c \in \text{int}(\Delta)$ such that $r(\bar{a}) > r(c) > r(\bar{b})$, we can elicit $c \succ^* \bar{b}$. Moreover, we cannot have $c \succ^* \bar{a}$, as c cannot set a reference point at A whenever $\bar{a} \in A$. Thus, we can conclude $\bar{a} \succ_r \bar{b}$ via Definition 1(ii-b).

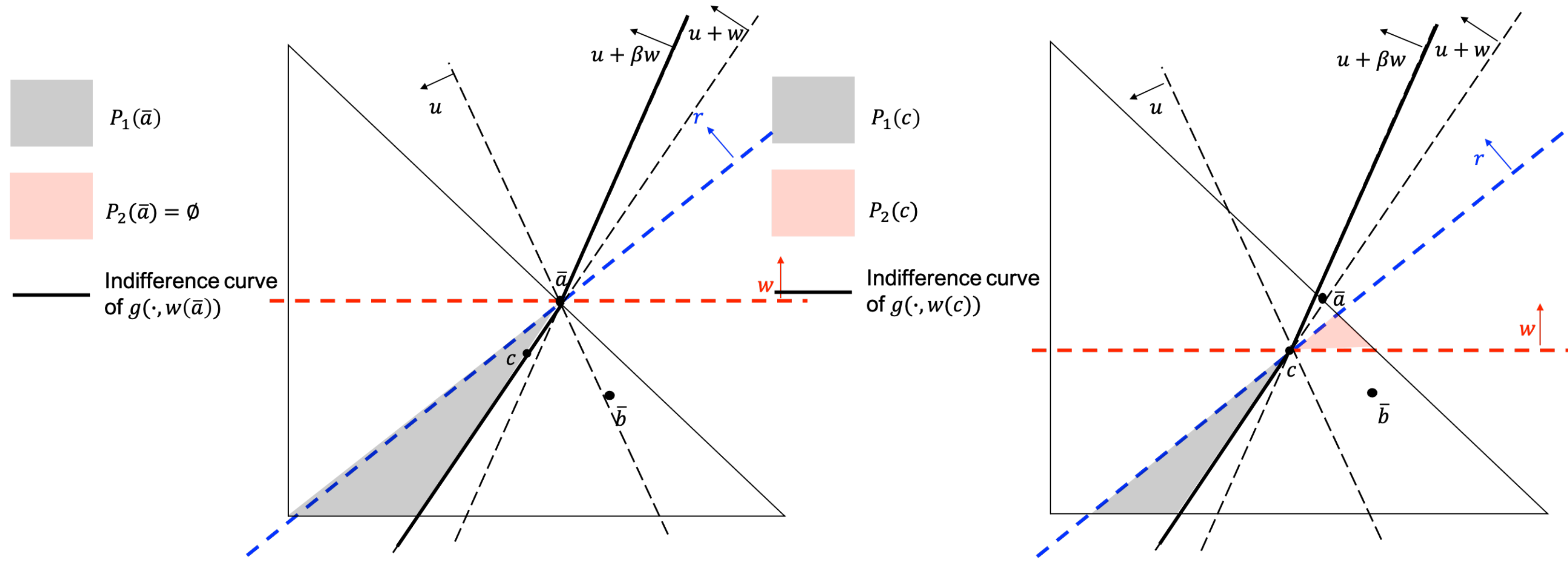
Figure A1: Nondegeneracy and Weak Nondegeneracy



Notes: Panel (a) presents an example of a PS preference which satisfies nondegeneracy, and Panel (b) presents an example of a PS preference which satisfies weak nondegeneracy but not nondegeneracy. Each dashed or solid straight line represents an indifference curve of u, w, $u+w$, $u+\beta w$ or r, with an allow indicating the increasing direction of the utility function. The bold solid line kinked at x in Panel (a) (at $\bar{a}$ in Panel (b)) denotes the indifference curve of the function $g(\cdot, w(x))$ $(g(\cdot, w(\bar{a})))$ defined in Lemma 20. In Panel (a), the black and red shaded areas depict $\mathcal{P}(x)$ and $\mathcal{S}(x)$, respectively, defined in Eq. (4) and (5) in Section 3.1. In Panel (b), the black and red shaded area depicts $P_1(\bar{a})$ and $P_2(\bar{a})$, respectively, defined in Lemma 29. See the text in Appendix B for details.

Figure A2: Reference Elicitation on the Boundary

(a) Non-existence of $d \in P_2(\bar{a})$ (b) Mediating alternative $c \in \text{int}(\Delta)$



Notes: Panel (a) presents an example of alternatives $\bar{a}, \bar{b} \in \Delta$ such that $r(\bar{a}) > r(\bar{b})$ but $\bar{a} \not\succ^* \bar{b}$, and Panel (b) illustrates how we can establish $\bar{a} \succ_r \bar{b}$ via Definition 1(ii-b) using a mediating alternative $c \in \text{int}(\Delta)$. Each dashed or solid straight line represents an indifference curve of u, w, $u + w$, $u + \beta w$ or r, with an allow indicating the increasing direction of the utility function. The bold solid line kinked at $\bar{a}$ in Panel (a) (at c in Panel (b)) denotes the indifference curve of the function $g(\cdot, w(\bar{a}))$ $(g(\cdot, w(c)))$ defined in Lemma 20. In Panel (a), the black and red shaded areas depict $P_1(\bar{a})$ and $P_2(\bar{a})$, respectively, defined in Lemma 29. In Panel (b), the black and red shaded areas depict $P_1(c)$ and $P_2(c)$, respectively. See the text in Appendix B for details.

C Other Predictions from the Simple Model

In Section 2.1, we developed a simple PS model of prosocial behavior and showed its insightfullness. In this section, we further discuss implications of the model, such as conformative versus pride-seeking behavior, and boomerang effects.

Recall that $x = 1$ denotes engaging in prosocial behavior, and $x = 0$ denotes non-engagement. Private and social payoffs are in conflict: $u(0) = \bar{u} > 0 = u(1)$ and $w(0) = 0 < \bar{w} = w(1)$, with $\beta\bar{w} < \bar{u} < \bar{w}$. At menu $\{0, 1\}$, the DM chooses an action by comparing the ex post utility of action 0, $U(0; \{0, 1\}) = \bar{u} - w(\varphi_r(\{0, 1\}))$, with that of action 1, $U(1; \{0, 1\}) = \beta\,[\bar{w} - w(\varphi_r(\{0, 1\}))]$. We compare decisions in the benchmark case $r(0) > r(1)$ (prosocial behavior is perceived to be uncommon) and in a post-intervention case $r'(0) < r'(1)$ (prosocial behavior is perceived to be common).

Conformity and pride seeking. The DM conforms to the reference alternative both in the benchmark case ($x = \varphi_r(\{0, 1\}) = 0$) and post-intervention case ($x = \varphi_{r'}(\{0, 1\}) = 1$). By contrast, if we modify the benchmark assumption so that $\beta\bar{w} > \bar{u}$, then the DM engages in prosocial behavior under both scenarios. In the modified benchmark, the DM deviates from the reference to seek pride. Thus, our model can produce conformative or pride-seeking behavior depending on β. Typical empirical findings suggest β is small (see footnote 18); still, in some contexts, individuals may seek to perform better than a natural reference point (Birke 2020, and see also footnote 16).

Boomerang effect. In a field experiment on electricity consumption, Schultz et al. (2007) find that providing descriptive information on neighbors' electricity usage led to desired electricity saving by high-consuming households but increased consumption by low-consuming households. To explain the latter result (which Schultz et al. (2007) call a "boomerang effect") without complicating the model, let $x = 0$ and $x = 1$ denote high consumption and low consumption of electricity, respectively, and suppose that the low-consuming households originally perceive norms (w, r') but the intervention updates the perceptions to (w, r). By the analysis in Section 2.1, the low-consuming households originally choose $x = 1$ but the intervention causes them to switch to $x = 0$. Thus, our model can explain the boomerang effect by an upward shift of the perceived descriptive norm for those households.

The model can also explain why penalizing certain behavior may induce the behavior, or rewarding certain behavior can discourage it. An introduction of a fine (reward) is often accompanied by information that the behavior is widespread (rare), which may push the descriptive norm in the unintended direction. Thus, changes in perceived norms might be a reason for these counter-intuitive empirical findings.

The purpose of the above example is to illustrate the importance of considering the perceived norms of individuals when introducing a policy, rather than develop a more thorough model. For example, the reduction in the electricity consumption by high-consuming households can be explained by the opposite shift in the perceived descriptive norm. Instead of developing a model which accommodates both types of households (possibly requiring more than two options), we note that even the direction of a policy effect, as well as its magnitude, crucially depends on what norms the households perceive prior to the intervention.