NBER WORKING PAPER SERIES

AGGREGATING LOCAL PREFERENCES TO GUIDE MARGINAL POLICY ADJUSTMENTS

Daniel J. Benjamin
Ori Heffetz
Miles S. Kimball
Nichole Szembrot

Working Paper 18787
http://www.nber.org/papers/w18787

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2013

We thank Matthew Adler, Geir Asheim, Gregory Besharov, Aaron Bodoh-Creed, Alexander Cappelen, Gabriel Carroll, Steve Coate, Koen Decancq, Marc Fleurbaey, Erzo Luttmer, Guy Mayraz, Ted O'Donoghue, Alex Rees-Jones, Erik Sørensen, Alois Stutzer, Bertil Tungodden, our discussant, Ed Glaeser, and audiences at the University of Michigan, the AEA Annual Meeting, and the Norwegian School of Economics for valuable comments and suggestions. For financial support we are grateful to Cornell's Institute for Social Science, S. C. Johnson Graduate School of Management, and NIH/NIA grants R01-AG040787 to the University of Michigan and T32-AG00186 to the NBER. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Aggregating Local Preferences to Guide Marginal Policy Adjustments
Daniel J. Benjamin, Ori Heffetz, Miles S. Kimball, and Nichole Szembrot
NBER Working Paper No. 18787
February 2013
JEL No. D69,H0,I38

ABSTRACT

We propose a social choice rule for aggregating preferences elicited from surveys into a marginal adjustment of policy from the status quo. The mechanism is: (i) symmetric in its treatment of survey respondents; (ii) ordinal, using only the orientation of respondents' indifference surfaces; (iii) local, using only preferences in the neighborhood of current policy; and (iv) what we call "first-order strategy-proof," making the gains from misreporting preferences second order. The mechanism could be applied to guide policy based on how policy affects responses to subjective well-being surveys.

Daniel J. Benjamin
Economics Department
Cornell University
480 Uris Hall
Ithaca, NY 14853
and NBER
daniel.benjamin@gmail.com

Ori Heffetz
Cornell University
Samuel Curtis Johnson Graduate School
of Management
324 Sage Hall
Ithaca, NY 14853
oh33@cornell.edu

Miles S. Kimball
Department of Economics
University of Michigan
Ann Arbor, MI 48109-1220
and NBER
mkimball@umich.edu

Nichole Szembrot
Economics Department
Cornell University
455 Uris Hall
Ithaca, NY 14853
nes66@cornell.edu

Making policy choices is unavoidable. The standard economic approach to guiding policy consists of two steps: estimate policy effects on individuals' utility, and then use a social choice rule that aggregates across individuals to generate a policy recommendation. Traditionally, the policy-effect estimates come from market data. However, market data are insufficient in some contexts involving externalities, public goods, and non-market goods, and may not reliably reveal preferences in contexts where people are uninformed or make systematic mistakes. For these reasons, economists have increasingly been exploring survey-based approaches to estimating policy effects on utility. For our purposes, such surveys have two key features: they are unincentivized, and they generally work best for eliciting local preferences at the status quo (because respondents find it harder to introspect about how they would feel in counterfactual, unfamiliar situations). In this paper, we propose a social choice rule for aggregating local ordinal preferences elicited from surveys into a local policy improvement, while reducing the incentive for survey respondents to misreport their preferences.

We have in mind three examples of survey-based preference measurement to which our method could be applied. First, contingent valuation surveys often elicit respondents' marginal rates of substitution between different policies or between a policy and money (i.e., willingness to pay).

Second, much research on subjective well-being (SWB) treats the response to a SWB survey question, typically regarding happiness or life satisfaction, as a proxy for utility and estimates how policies affect it. Unlike contingent valuation or policy referendums, this approach sidesteps respondents' policy misconceptions or lack of information; and relative to traditional market-based indicators such as GDP, SWB data may capture a broader range of aspects of well-being (e.g., Stiglitz, Sen, and Fitoussi, 2009). However, evidence indicates that people make choices that systematically deviate from what they *believe* would maximize their responses to commonly-used SWB measures (Benjamin, Heffetz, Kimball, and Rees-Jones, 2012, 2013), suggesting that a single-question SWB measure is inadequate as a utility proxy.

The primary application we have in mind is estimating the effects of policy on responses to a range of SWB questions that capture distinct aspects of well-being, including, for example, own and family happiness, health, security, and freedoms. The effect of policy on utility can be estimated as a weighted average of the effect of policy on these responses, with the weights derived from a separate survey that elicits individuals' marginal rates of substitution across aspects of well-being.[1] The theory and methods for combining responses to multiple SWB questions into a utility proxy are developed in a companion paper (Benjamin, Heffetz, Kimball, and Szembrot, 2012).

I. Key Properties

As our setup, we imagine a government agency that conducts a national survey—along the lines of one of the above examples—and uses it to estimate local preferences over policies. Our proposed mechanism for aggregating these preferences to yield a local policy change has four defining features.

First, the mechanism treats all individuals *symmetrically*. Second, it treats the preference data as *ordinal* and non-comparable across individuals. When aggregating policy effects on a SWB-based utility proxy, non-comparability is important not only because of well-known objections to assuming interpersonal comparability of utility, but also because different people might use the SWB response scales differently.

Third, the mechanism is *local* in using only the orientation of indifference surfaces in the neighborhood of a status-quo policy vector, and in generating a marginal adjustment of policy. Two attractions of a local approach to social choice are that local policy effects can be estimated empirically more credibly than global effects, and that designing policy *de novo* is usually less practical than adjusting policy (cf. Feldstein, 1976). Moreover, in the particular context of SWB data, we believe that using these data in a limited way (as a supplement to more familiar inputs to policy-making) is prudent because the enterprise is still exploratory and untested.

Finally, the mechanism is what we call "*first-order strategy-proof*" (*FOSP*). The Gibbard-Satterthwaite theorem rules out a mechanism being strategy-proof for all possible preferences; *FOSP* means strategy-proofness when indifference surfaces over policies are linear. That is, in the mechanism-design problem where each such individual reports her local preferences to the government agency, the mechanism makes truthful reporting a dominant strategy. Hence if the agency estimates policy preferences from SWB data, answering the SWB

[1] While such a combination of SWB measures may be the best available candidate for a survey-based utility proxy, we emphasize the need for work to deal with a variety of issues that bedevil the measurement of SWB (see, e.g., Adler, 2012), which we do not address. In addition, there are many econometric challenges to identifying the effects of policy on SWB responses. For example, a policy might affect an individual's use of the response scale.

survey honestly is a dominant strategy. More generally, since any smooth indifference surface is locally linear, FOSP makes the gains from misreporting second order.

While incentive-compatibility has been a concern in the contingent valuation literature, it is virtually unmentioned in the SWB literature. Indeed, there are reasons not to worry: any single individual would have negligible effect on policy but may incur a psychic cost from being dishonest, and many ways of advantageously manipulating responses would be hard to figure out because the survey responses only matter indirectly for policy (via the agency's estimates of policy effects).

Nonetheless, we *would be* worried if there were *easy* ways to advantageously manipulate policy, because interest groups could organize their members to do so.[2] FOSP rules out many straightforward incentives for survey-response manipulation.

II. The Aggregation Mechanism

Suppose that there are $P \geq 1$ policies. Each element of $\boldsymbol{p} \equiv (p_1, \ldots, p_P)$ represents the level of one policy and is a real number that the government can adjust independently of other policies. (It is straightforward to recast the policy vector to eliminate dependencies—e.g., those due to a budget constraint—by reducing the dimensionality of the policy space.) Let $\boldsymbol{p}_0$ denote the status-quo policy vector. There are $\Theta > 1$ types in the population, defined by having identical preferences over policies (a reduced-form representation of the effects of policies on the arguments of utility). For each type θ, let ϕ_θ denote its fraction of the population. Let $r_\theta(\boldsymbol{p})$ denote its ordinal utility as implied by its survey responses, which—allowing for misreporting—is not necessarily the same as $u_\theta(\boldsymbol{p})$, its true ordinal utility. We assume that each $u_\theta(\cdot)$ and $r_\theta(\cdot)$ is continuously differentiable, and hence we can locally approximate them as linear.

A government agency conducts a national survey and, for each type θ, uses the survey responses to estimate the marginal rates of substitution, $\frac{\partial r_\theta / \partial p_j}{\partial r_\theta / \partial p_k}$, across each pair of policies j and

[2] For example, if SWB survey responses were treated as cardinal and added up to get a "social welfare function"—as is done implicitly in empirical work that estimates effects of policy on SWB averaged across all individuals—individuals could increase their impact on policy by spreading their responses to make use of the entire SWB response scale. To take an example that treats SWB data as ordinal, suppose that the equivalent variation of a policy were estimated as the change in income required to hold SWB constant—as is commonly done in the SWB literature—and these money-metric utilities were used for guiding policy. Then individuals could magnify their weight in policymaking by making their reported SWB less sensitive to changes in income—say, by answering the SWB question while focusing on non-income-related aspects of well-being.

k. Let $\boldsymbol{\nabla r_\theta} \equiv \left(\frac{\partial r_\theta}{\partial p_1}, \dots, \frac{\partial r_\theta}{\partial p_P}\right)$ and $\boldsymbol{\nabla u_\theta} \equiv \left(\frac{\partial u_\theta}{\partial p_1}, \dots, \frac{\partial u_\theta}{\partial p_P}\right)$ respectively denote the gradients at $\boldsymbol{p}_0$ of the (survey-based) utility-proxy and the (unobserved) utility.[3] The agency's aggregation mechanism is its algorithm for mapping the $\boldsymbol{\nabla r_\theta}$'s into a policy change $\Delta\boldsymbol{p}$.

Since the P policies are measured in different units (e.g., tax rate vs. dollars of spending), it is necessary to define what is meant by a "local change" in each policy. For each policy j, let the exogenous distance-metric parameter $m_j > 0$ denote the amount of change corresponding to "1 policy-unit." Three considerations may inform how the m_j's are set. First, since the mechanism assumes that the indifference surfaces are linear, the policy units should be small enough that this approximation is likely to be good based on prior beliefs. Second, since 1 policy unit will be the *maximum* Euclidean distance that the mechanism can prescribe, m_j should be small enough that policymakers are comfortable entrusting m_j amount of policy j to be determined by survey data. Third, the relative m_j's can correspond to intuitive judgments of similar-sized changes. The mechanism's key properties do not depend on the values of the m_j's but, as we discuss in section IV below, these values would matter in practice.

Define the matrix $\boldsymbol{M} \equiv \mathrm{diag}(m_1, m_2, \dots, m_P)$. Note that any policy change $\Delta\boldsymbol{p}$ in natural units is the change $\boldsymbol{M}^{-1}\Delta\boldsymbol{p}$ in policy units, and any utility gradient $\boldsymbol{\nabla u_\theta}$ corresponds to $\boldsymbol{M\nabla u_\theta}$ when the policy change is measured in policy units. If $\|\boldsymbol{M\nabla r_\theta}\| > 0$, define the "normalized gradient" $\widetilde{\boldsymbol{\nabla r_\theta}} \equiv \frac{\boldsymbol{M\nabla r_\theta}}{\|\boldsymbol{M\nabla r_\theta}\|}$, which is the utility-proxy gradient in policy units, normalized to have length 1. If $\|\boldsymbol{M\nabla r_\theta}\| = 0$, define $\widetilde{\boldsymbol{\nabla r_\theta}} \equiv \boldsymbol{0}$.

<u>Example:</u> Imagine two types, Young and Old, and two policies, the federal tax on distilled spirits and spending on national parks. Suppose the relevant government agency estimates that a \$1 increase in the tax per proof gallon on distilled spirits increases the utility proxy of the Old by 3 units and decreases the utility proxy of the Young by 3 units. The agency also estimates that a \$100-million increase in spending on national parks decreases the Old's utility proxy by 2 units and increases the Young's by 3 units. With these data, $\boldsymbol{\nabla r}_{\text{Old}} = \left(\frac{3 \text{ utils}}{\$1 \text{ per proof gallon}}, \frac{-2 \text{ utils}}{\$100 \text{ mil.}}\right)$ and $\boldsymbol{\nabla r}_{\text{Young}} = \left(\frac{-3 \text{ utils}}{\$1 \text{ per proof gallon}}, \frac{3 \text{ utils}}{\$100 \text{ mil.}}\right)$.

---

[3] We assume local non-satiation except at a bliss point. Then, without loss of generality, we choose a monotonic transformation for u_θ and r_θ such that neither $u_\theta(\boldsymbol{p}_0)$ nor $r_\theta(\boldsymbol{p}_0)$ is an inflection point. This ensures that $\boldsymbol{\nabla u_\theta} = \boldsymbol{0}$ if and only if $\boldsymbol{p}_0$ is a bliss point of $u_\theta(\cdot)$, and $\boldsymbol{\nabla r_\theta} = \boldsymbol{0}$ if and only if $\boldsymbol{p}_0$ is a bliss point of $r_\theta(\cdot)$.

In order to apply the mechanism, the agency must pin down the policy units. Suppose that one policy unit corresponds to a \$1.25 per proof gallon change in the tax on distilled spirits, or a \$125 million change in federal funding for the National Park Service: $m_1 = \$1.25$ per proof gallon and $m_2 = \$125$ million.

The normalized gradients are

$$\widetilde{\boldsymbol{\nabla r}}_{\text{Old}} = \left(\frac{\$1.25\left(\frac{3\text{ utils}}{\$1}\right)}{\sqrt{\left[\$1.25\left(\frac{3\text{ utils}}{\$1}\right)\right]^2 + \left[\$125\text{mil.}\left(\frac{-2\text{ utils}}{\$100\text{mil.}}\right)\right]^2}}, \frac{\$125\text{mil.}\left(\frac{-2\text{ utils}}{\$100\text{mil.}}\right)}{\sqrt{\left[\$1.25\left(\frac{3\text{ utils}}{\$1}\right)\right]^2 + \left[\$125\text{mil.}\left(\frac{-2\text{ utils}}{\$100\text{mil.}}\right)\right]^2}} \right) = (0.83, -0.55) \text{ and}$$

$$\widetilde{\boldsymbol{\nabla r}}_{\text{Young}} = \left(\frac{\$1.25\left(\frac{-3\text{ utils}}{\$1}\right)}{\sqrt{\left[\$1.25\left(\frac{-3\text{ utils}}{\$1}\right)\right]^2 + \left[\$125\text{mil.}\left(\frac{3\text{ utils}}{\$100\text{mil.}}\right)\right]^2}}, \frac{\$125\text{mil.}\left(\frac{3\text{ utils}}{\$100\text{mil.}}\right)}{\sqrt{\left[\$1.25\left(\frac{-3\text{ utils}}{\$1}\right)\right]^2 + \left[\$125\text{mil.}\left(\frac{3\text{ utils}}{\$100\text{mil.}}\right)\right]^2}} \right) = (-0.71, 0.71).$$

As illustrated in Figure 1, each is a vector of length 1 that points in the direction of maximal increase in the utility proxy for the group.

The aggregation mechanism is:

$$\boldsymbol{M}^{-1}\Delta\boldsymbol{p} = \sum_{\theta=1}^{\Theta} \phi_\theta \, \widetilde{\boldsymbol{\nabla r}}_{\boldsymbol{\theta}} \, .$$

In words, the policy change (in policy units) is the weighted sum of the normalized gradients, with weights equal to population shares.

<u>Example (continued):</u> Suppose that the population shares are $\phi_{\text{Old}} = 2/5$ and $\phi_{\text{Young}} = 3/5$. The mechanism prescribes a policy change (in policy units) of $\frac{2}{5}(0.83, -0.55) + \frac{3}{5}(-0.71, 0.71) = (-0.09, 0.22)$. Figure 2 illustrates this vector addition, with the resultant policy change from $\boldsymbol{p}_0$ to $\boldsymbol{p}_1$ shown by the solid vector. Converting back to natural units: reduce the tax by $0.09 \times \$1.25$ per proof gallon $= \$0.11$ per proof gallon, and increase spending on national parks by $0.22 \times \$125$ million $= \$25.3$ million.

To see that the mechanism is FOSP, notice that it decentralizes the policy change: each type contributes a vector (of fixed length) that it chooses via its survey responses. Each type θ evaluates its most-preferred direction starting at $\boldsymbol{p}_0 + \sum_{\hat{\theta} \neq \theta} \phi_{\hat{\theta}} \, \widetilde{\boldsymbol{\nabla r}}_{\hat{\boldsymbol{\theta}}}$ (i.e., taking into account the summed contributions of the other types), but with linear preferences, the most-preferred direction there is the same as at $\boldsymbol{p}_0$. Therefore, each type does best by ensuring that the agency's estimate of its most-preferred direction at $\boldsymbol{p}_0$ is correct—i.e., that $\boldsymbol{\nabla r_\theta}$ is proportional to $\boldsymbol{\nabla u_\theta}$.

To provide some intuition regarding the form of the mechanism, consider the problem faced by a social planner who directly observed the cardinal utilities of the individuals and maximized a social welfare function, $W(u_1, \dots, u_\Theta)$, where $W(\cdot)$ is strictly increasing, strictly concave, and continuously differentiable. To ensure that the policy change is local, we constrain it to be at most 1 policy unit:

$$\max_{\Delta \boldsymbol{p}} W\big(u_1(\boldsymbol{p}_0 + \Delta \boldsymbol{p}), \dots, u_\Theta(\boldsymbol{p}_0 + \Delta \boldsymbol{p})\big) \text{ subject to } (\boldsymbol{M}^{-1}\Delta \boldsymbol{p})'(\boldsymbol{M}^{-1}\Delta \boldsymbol{p}) \leq 1.$$

The vector first-order condition is $\sum_{\theta=1}^{\Theta} \frac{\partial W}{\partial u_\theta} \boldsymbol{\nabla u_\theta} = \lambda \boldsymbol{M}^{-2} \Delta \boldsymbol{p}$, where $\lambda > 0$ is the Lagrange multiplier. Hence the optimal $\boldsymbol{M}^{-1}\Delta \boldsymbol{p}$ has length 1 policy unit and direction $\sum_{\theta=1}^{\Theta} \frac{\partial W}{\partial u_\theta} \boldsymbol{M \nabla u_\theta}$. Notice that our mechanism *is* this solution, except: (*i*) $\frac{\partial W}{\partial u_\theta}$ is replaced by ϕ_θ (because respondents are treated symmetrically); (*ii*) $\boldsymbol{M \nabla u_\theta}$ is replaced by $\frac{\boldsymbol{M \nabla r_\theta}}{\|\boldsymbol{M \nabla r_\theta}\|}$ (because only a survey-based gradient is available, and it must be normalized since its magnitude has no meaning); and (*iii*) our mechanism changes policy by less than 1 policy unit unless all types have the same local preferences.

This set of observations yields three insights. First, a (non-symmetric) variant of the mechanism could give different weights to different types (e.g., more weight to worse-off individuals). Second, given that every type has an incentive to ensure that its $\boldsymbol{\nabla r_\theta}$ is proportional to its $\boldsymbol{\nabla u_\theta}$, the direction of change yielded by the mechanism is optimal for the class of social welfare functions satisfying $\frac{\partial W}{\partial u_\theta} = \frac{\phi_\theta}{\|\boldsymbol{M \nabla r_\theta}\|}$ for each θ. Therefore, if the agency had only ordinal preference data but were *not* concerned about incentive-compatibility, it might nonetheless use the mechanism to determine the direction of policy change—but then implement a change of 1 policy unit. Finally, the inefficiency that arises from changing policy by less than 1 policy unit can be interpreted as the social cost incurred by having FOSP. As illustrated in Figure 3, if the policy change were instead *guaranteed* to have length 1 policy unit, then in general there would be an incentive to misreport one's preferences.[4]

[4] An alternative way of motivating the mechanism—as if it were balancing social welfare against a quadratic cost of change—generates a smaller policy change when there is disagreement, even if one leaves aside any concern about truthful reporting. Intuitively, at an optimum, the marginal cost of change in dimension *j* (proportional to m_j^{-1}) is equal to the marginal social benefit, which is smaller with more disagreement.

III. Additional Properties

The mechanism satisfies a local version of the Paretian principle: if some direction of policy change would make all groups better off, then the mechanism will implement change in that direction. Formally, for any policy-change vector $\boldsymbol{\delta}$ such that $\boldsymbol{\delta}'(\boldsymbol{M}\nabla\boldsymbol{u}_{\boldsymbol{\theta}}) > 0$ for all θ, it follows that $\boldsymbol{\delta}'\left(\sum_{\theta=1}^{\Theta} \phi_\theta \frac{\boldsymbol{M}\nabla\boldsymbol{u}_{\boldsymbol{\theta}}}{\|\boldsymbol{M}\nabla\boldsymbol{u}_{\boldsymbol{\theta}}\|}\right) > 0$ and hence $\boldsymbol{\delta}'(\boldsymbol{M}^{-1}\Delta\boldsymbol{p}) > 0$. This property implies that, even if respondents have divergent interests on some policy dimensions, the mechanism will find and implement changes on policy dimensions where respondents' interests coincide. The mechanism can even disfavor policies, such as certain restrictions on trade, that benefit a few individuals a great deal but hurt almost everyone else a little bit.

Another important property is that, when $P > 1$, each type has more influence on the policy change along the policy dimensions where a marginal adjustment matters *relatively* more to that type. Consequently, the direction prescribed by the mechanism can depend on which and how many dimensions are included. Figure 4 illustrates an example with three types that have equal population weights. With only information about preferences over policy dimension 1 (Figure 4b), the mechanism, which specializes to $\Delta p_1 = \sum_{\theta=1}^{\Theta} \phi_\theta \text{ sign}\left(\frac{dr_\theta}{dp_1}\right)$, prescribes an increase in policy 1. Incorporating information about preferences over dimension 2 (Figure 4a), however, the mechanism prescribes a *decrease* in policy 1. Since including more dimensions uses more information on tradeoffs, the mechanism ideally would be used with as many policy dimensions included as possible.

IV. Discussion

The mechanism's biggest practical limitation may be the degrees of freedom afforded by the several considerations that inform the choice of the m_j's. All else equal, with a larger m_j the mechanism will generate a policy change that is relatively larger in magnitude in dimension *j*; it could also reverse the direction of change in another dimension (in Figure 4, changing m_2 from zero to positive reverses dimension 1's direction of change). To minimize the risk of manipulation by government (which is akin to agenda-setting), it is crucial that the agency that determines the m_j's be independent and non-political.

Since in practice the mechanism would be likely to be applied iteratively, the most urgent next step is to understand its dynamic properties. Our preliminary analysis suggests that if individuals differ in their preferences but truthfully report them, the iterated mechanism could in some circumstances have a limit cycle. We view such non-convergence as an undesirable property similar in spirit to the intransitivity of voting—a limitation that the public seems willing to tolerate in democracies. Moreover, *any* social-preference ordering that uses only local information about individuals' preferences will be intransitive (Fleurbaey, Suzumura, and Tadenuma, 2005); indeed, it is well-known that widely-used criteria such as compensating variation share this limitation.

We view this paper primarily as a first step that merits further development. Nonetheless, there are several immediately actionable implications for empirical researchers that would facilitate exploring realistic applications of the mechanism. First, researchers should report the quantile of SWB respondents at which the estimated effect of a policy is zero; the mechanism as applied only to that policy dimension would dictate adjustment proportional to the implied vote margin. Second, when studying the effect of more than one policy, researchers should report the marginal rate of substitution between policies. Third, researchers should report policy effects on different subgroups of the population in order to move toward identifying types and analyzing the implied tradeoffs between them.

References

Adler, Matthew D. 2012. “Happiness Surveys and Public Policy: What’s the Use?” *Duke Law Journal,* forthcoming.

Benjamin, Daniel J., Ori Heffetz, Miles S. Kimball, and Alex Rees-Jones. 2012. “What Do You Think Would Make You Happier? What Do You Think You Would Choose?” *American Economic Review*, 102(5), 2083–2110.

Benjamin, Daniel J., Ori Heffetz, Miles S. Kimball, and Alex Rees-Jones. 2013. “Can Revealed-Preference Tradeoffs Be Inferred From Happiness Data? Evidence from Residency Choices.” Cornell University manuscript, January.

Benjamin, Daniel J., Ori Heffetz, Miles S. Kimball, and Nichole Szembrot. 2012. “Beyond Happiness and Satisfaction: Toward Well-Being Indices Based on Stated Preference.” NBER Working Paper No. 18374.

Feldstein, Martin. 1976. “On the theory of tax reform.” *Journal of Public Economics*, 6, 77-104.

Fleurbaey, Marc, Kotaro Suzumura, and Koichi Tadenuma. 2005. “Arrovian aggregation in economic environments: how much should we know about indifference surfaces?” *Journal of Economic Theory*, 124, 22–44.

Stiglitz, Joseph E., Amartya Sen, and Jean-Paul Fitoussi. 2009. *Report by the Commission on the Measurement of Economic Performance and Social Progress*.

Figure 1: Local Indifference Curves and Corresponding Gradients

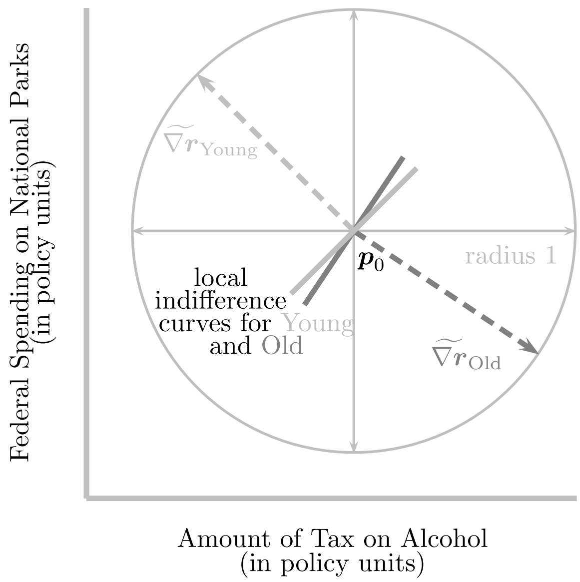

Figure 2: The Mechanism: A Weighted Average of Normalized Gradients

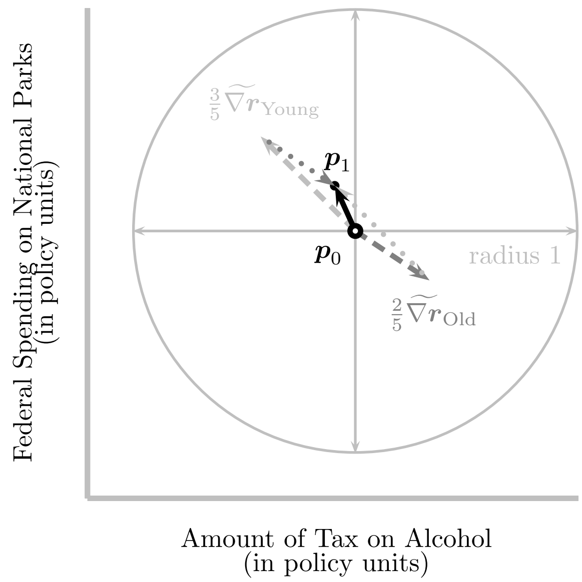

Figure 3: Incentive to Misreport Preferences if Policy Change is Always 1 Policy Unit

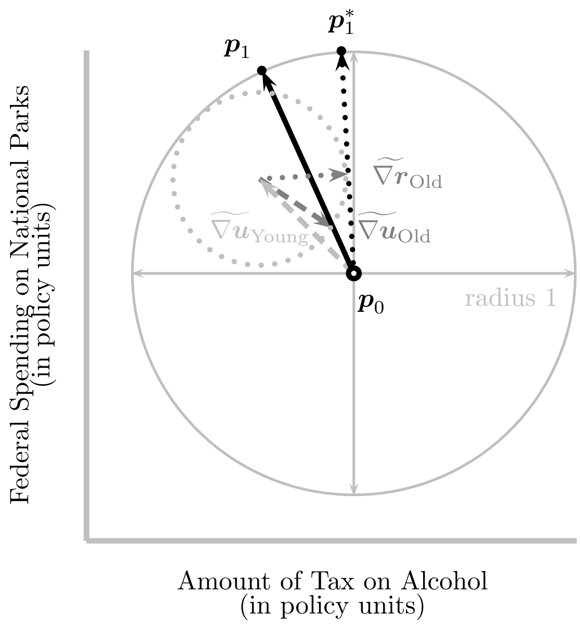

Figure 4a: Policy Change Generated by the Mechanism in 2 Dimensions

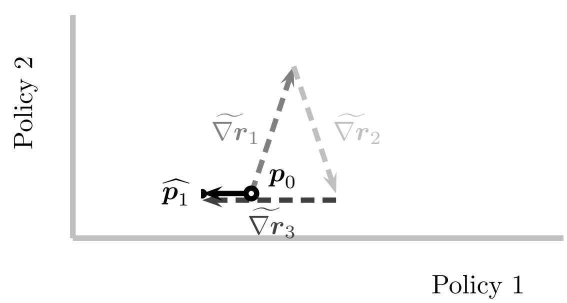

Figure 4b: Policy Change Generated by the Mechanism in 1 Dimension