Heterogeneous Effects of Informational Nudges on Pro-Social Behavior*

Jiayi Bao

University of Pennsylvania

Benjamin Ho

Vassar College

Revision Requested – BE Journal of Economics Analysis and Policy – Jan 2015

Abstract

Numerous experimental studies of informational nudges both in the lab and the field have demonstrated not just that informational nudges are effective policy tools for influencing behavior, but also that nudges have heterogeneous impacts that differ depending on the characteristics of the person involved and the situation. We adapt Andreoni's theory of warm glow impure altruism to account for how altruism motives responds differently depending on the disposition of the person and the situation. The model explains both positive spillovers (moral cleansing) and negative spillovers (moral licensing) for behavioral interventions, showing that targeting of informational campaigns depends on the complementarity between people's traits and the intervention's content. More importantly, the design of economic incentives (like Pigouvian taxes) to shift economic behavior should depend on both the distribution of social preferences in the population and the use of behavioral interventions.

Keywords: nudge, heterogeneity, moral licensing, moral cleansing, impure altruism, warm glow

---

* We thank seminar and conference participants at Harvard and Vassar for their helpful comments, as well as funding from the Mr. and Mrs. Noah Barnhart, Jr. Fund. In particular, we thank Evsen Turkay and Paul Ruud for very helpful feedback.

Contact Information: Ben Ho beho@vassar.edu.

1 Introduction

Over the past decade, informational nudges, a type of information provision with the purpose of altering social behavior, have become increasingly prominent in the literature (Allcott and Mullainathan, 2010; Thaler and Sunstein, 2008; Cialdini et al., 2006). Moreover, the real world implications have been acknowledged in multiple spaces. In the market place, research has shown that peer comparisons can induce pro-social behavior in energy consumption (Ayres, Raseman, and Shih, 2012). *Nudge*, a New York Times bestseller written by Cass Sunstein and Richard Thaler (2008) introduced and advocated the concept of "nudges," a type of government policy that helps people make better choices without limiting their freedom to choose. These policies have seen widespread popularity: the Behavioral Insights Team, launched in the UK in 2010, has successfully nudged tax debtors to pay more of their taxes owed by mailing them letters telling them about the behavior of others[1] while the White House has kicked off several federal projects in the second half of 2013 to nudge Americans to save more money for retirement, to curb energy uses, and to trim energy costs.[2]

Despite their increasing popularity, evidence for the success of programs using informational nudges to induce pro-social behavior has been mixed (Schultz et al., 2007). In particular, making the most of these policies requires understanding the heterogeneity in responses across sub-populations

---

[1] Peter John, "Nudges and information are means to assist conventional forms of policy implementation," *The British Politics and Policy Blog,* March 13, 2013, http://blogs.lse.ac.uk/politicsandpolicy/archives/33452.

[2] David Brown, "Nudge and Behavioural Insights," *Public Record Office Victoria*, September 28, 2012, http://prov.vic.gov.au/blog-only/nudge-and-behavioural-insigts.

Maxim Lott, "White House Progressives Create 'Nudge Squad' to Shape Behavior," *FoxNews.com*, July 30, 2013, http://newmediajournal.us/indx.php/item/10079.

and across intervention designs. One literature focuses on transient situational factors that affect the short-term response to a nudge such as the importance of the wording of the message, for example, whether the message is injunctive, "this is what you ought to do," or descriptive, "this is what others are doing" (Ferraro and Price, 2013; Merritt, Effron, and Monin, 2010; Cialdini et al., 2006). Another literature focuses on long-run dispositional difference across different groups like differences in social responsibility, group identity or prior behavior (Akerlof and Kennedy, 2013; Allcott, 2011; Beshears et al., 2011). The heterogeneity in impact can lead to unintended spillovers. For instance, a nudge toward good behavior in one domain may license individuals to behave badly in a different domain (Ho et al., 2014). Even successful nudges such as the letters to tax debtors sent in the UK are instructive of the limitations of the nudge approach. The UK policy focuses only on a subset of the worse offenders, and was only effective when the message carefully selected the comparison group (Hallsworth et al 2014).

Therefore, despite the popularity of nudges in popular discourse, important questions remain. How should we account for the heterogeneous impact of nudges across different populations or different interventions both in terms of designing behavioral policies but also for evaluating their welfare impact? More importantly, as the magnitude of change observed in most field studies are small—for example, 1-2% savings in the case of electricity (Allcott, 2011), or less than 1% in the case of water savings (Ferraro and Price, 2013)—do the benefits of the nudge outweigh the feelings of guilt they induce? Also, as nudges by themselves seem to have small effects, would nudge based policies work better in conjunction with traditional economic incentives like Pigouvian taxation?

In this paper, we organize what we know from the empirical literature into a coherent

framework to think about the heterogeneity of nudges. The framework allows us to address questions about how behavioral interventions interact with traditional economic policies and to assess impacts on social welfare. We focus on the interaction between two dimensions of heterogeneity, the effect of situational nudges and individual disposition, developing a model that considers short-term situation and long-run disposition as two separate channels for inducing the warm-glow or cold-prickle associated with impure altruism. While some of the past literature we have noted does acknowledge the heterogeneity, there has been no systematic study of how different kinds of nudges interact with disposition. Existing research focuses on each nudge intervention independently. Also, while the distinction between disposition and situation is a fundamental one in psychology, it is a relatively under utilized conceptual framework within the realm of economics.

We parameterize the situational impure altruism induced by nudges and the dispositional impure altruism inherent to the individual using cross derivatives, and predict behavior and welfare based on this relationship. By organizing the existing literature in this way, we hope to provide guidance for the empirical research needed to properly assess the welfare of nudge based policies.

The main policy context we are interested in is informational message campaigns targeted toward reducing consumption of goods that involve negative externalities: e.g. reducing consumption of gasoline in order to mitigate environmental damage. Therefore, the main disposition we will focus on is whether people are more or less inclined to have altruistic preferences toward curtailing their own consumption for the good of the public. Since the natural policy in this context is Pigouvian taxation, we will consider the interactions between our two

dimensions of impure altruism and optimal tax policy, although the model applies to any traditional economic incentive scheme designed to change consumption behavior. Even in domains where changes to tax policy are politically infeasible, design of behavioral interventions should depend on the existing tax environment.

The main findings of the paper include the following:

(i) The effect of pro-social nudges on welfare can be either positive or negative.

(ii) The observed heterogeneity across behavioral interventions can be explained by the cross derivative between situational and dispositional effects.

(iii) The optimal tax is a function of disposition and situational nudges.

a. Targeted taxes on certain sub-populations may be welfare enhancing.

b. Optimal economic policy accounts for interactions with situational nudges.

c. The inefficiency of a tax policy that does not take into account impure altruism and situational nudges depends on the complementarity between situational altruism and dispositional altruism.

The paper proceeds as follows. We first review the empirical literature on informational nudges and the explanations for their heterogeneous effects. We compare our model with past experimental evidence and similar theories before we introduce our model of impure altruism. We then present the main propositions about the effects of impure altruism on individual behavior, individual utility, and social welfare, and then examine the role of the interaction between situation and disposition. Next, we consider taxation as either an alternative or a supplement to nudges for social welfare enhancement and develop a model to explore how optimal taxes can depend on

situational factors. The implications for policy are discussed. Finally, some extensions of the model are considered for future research.

2 Literature Review

The gaining popular use of informational nudges is a result of the emerging behavioral economics and psychological literatures. Bergstrom et al. (1986) introduce the idea of impure altruism and acknowledge its role in driving individual choices. Andreoni (1990; 1995) extends the previous analysis and incorporates impure altruism into a more general model, reasoning that nudges work by inducing the feeling of a cold prickle when creating negative social externalities. Ayres, Raseman, and Shih (2012) find in two large field experiments partnered with private companies, Positive Energy and oPower, that by providing feedback to customers on home electricity and natural gas usage with a focus on peer comparisons, utilities can reduce energy consumption at a low cost. Berger and Rand (2008) also acknowledge the power of peer information provision in redressing heavy drinking in universities through laboratory and field experiments.

However, informational nudges are not always effective and their effects heavily rely on the type of people targeted and the framing of the message. While some nudges do effectively reduce production of a negative social externality, others have no impact or may even be counterproductive. Cialdini (1991) explains the heterogeneous effects by arguing that injunctive nudges, which guide behavior by shifting the perception of how most others would approve or disapprove of a person's conduct, are more likely to lead to beneficial social conduct than the descriptive types, which guide behavior through the perception of how most others would behave. Cialdini (2006) shows how some nudges backfire. In Arizona's Petrified Forest National Park, while the rate of petrified wood theft

decreased when people were given a injunctive message about what behaviors were not considered acceptable, another treatment using a descriptive message noting that others routinely violated the prohibition proved to be counterproductive. Merritt, Effron, and Monin (2010) explain how one act of pro-social behavior can lead to more pro-social behavior when it is seen as a commitment to the cause, but can lead to more anti-social behavior if it makes people feel like they have done "enough." Ferraro and Price (2013) revisit a field experiment about water consumption and find that messages based on social norms and social comparisons are more effective than instructive descriptions for reducing water usage.

The type of nudge is not the only source of heterogeneity in terms of effectiveness. Different groups of people are often affected differently by the same informational messaging for two general reasons. The first reason has to do with different people beginning with different status quos. Allcott (2011) notices the heterogeneous effects by evaluating oPower's Home Energy Report letters comparing residential utility customers' electricity usage to that of their neighbors. Residents using less than the average increased their energy usage, while though those using more decreased usage — mean energy usage was therefore largely unchanged. Akerlof and Kennedy (2013) further point out that campaigns deploying messages describing public levels of drug and alcohol use, recycling, and littering, et cetera, have often had little success in heightening adoption of pro-social behaviors. They attribute the ineffectiveness of informational nudges to their heterogeneous impact on people whose conducts are above social norms (socially desirable) and people whose conducts are below social norms (socially undesirable). The second source of heterogeneity comes from differences in individual disposition for pro-sociality due to factors such as group identity, perception of moral

obligations, and individual altruism. Beshears et al. (2011) find that peer information increased retirement savings of non-unionized recipients but decreased savings of unionized recipients, attributing the difference to differences in norms between the union and non-union workers. Jordan, Mullen and Murnighan (2011) find that people who were asked to recall their immoral behavior reported greater participation of moral activities than people who recalled moral behavior. Ho et al. (2014) find in paired lab and field studies of green electricity purchases that informational nudges have a stronger effect on those who were intrinsically inclined to be pro-social.

Our paper considers situation and disposition as two ways to induce the warm-glow or cold-prickle associated with impure altruism and parametrizes their effects into two temporal dimensions. Formally, disposition is a tendency by an individual to act in a specified way and is stable over time. We denote the effect of disposition by θ and call it the parameter for dispositional altruism. A person with high θ is more likely to be pro-social whereas a person with low θ is less likely. A nudge, on the other hand, has a situational effect that changes each period. We denote the effect of nudges by ω and call it situational altruism. In the case of informational nudges, someone who receives a nudge and thereafter perceives his behavior to be more socially desirable than average (above the norm) will experience low ω. Someone who perceives his personal behavior to be less socially desirable than the average (below the norm) will experience high ω.

Our theory shows how the heterogeneous effect of behavioral interventions depends on a third factor: the interaction between situation and disposition. We examine the individual psychic cost function for the relationship between situation and disposition. Situation and disposition are complements when people that are more altruistic are more receptive to information that confirms

their prior inclinations. Situation and disposition are substitutes when a situational nudge causes resentment in those most likely to give.

Our findings are consistent with the self-completion theory and moral regulation model in psychology. According to the self-completion theory of Jordan, Mullen, and Murnighan (2011), "recalling one's (im)moral behavior will lead to compensatory rather than consistent moral action as a way of completing the moral self." Sachdeva, Illiev, and Medin (2009) explain the compensatory moral action differently and argue that moral or immoral behavior can result from an internal balancing of moral self-worth and the cost inherent in altruistic behavior. When moral identity is threatened, moral behavior is a means to regain some lost self-worth (moral cleansing); however, affirming a moral identity leads people to feel licensed to act immorally (moral licensing). An experiment conducted by Darlington and Macker (1966) show that participants who were led to believe that they had harmed another person were more likely to subsequently engage in altruistic behavior such as donating blood to a local hospital, indicating guilt-induced moral cleansing. On the other hand, moral licensing can happen when good deeds of those who are above the norm establish moral credits that can be "withdrawn" to "purchase" the right to do bad deeds with impunity (Merritt, Effron, Monin, 2010). Monin and Miller (2001) notice that licensing operates by providing a temporary boost in self-concept, and an initial altruistic intent that boosts self-concept can liberate people to choose more indulgent option. Khan and Dhar (2006) and Ayal and Gino (2011) also note that people may feel licensed to refrain from good behavior when they have amassed a surplus of moral currency.

Formally, our model shares much with Andreoni (1990) as we both aim to extend the analysis of

Bergstrom et al. (1986) by decomposing their original idea of impure altruism. Andreoni's model characterizes people as either pure altruists whose preferences only depend on the total consumption of a public good or as egoists whose preferences only depend on the private consumption of the public good, with impure altruism combining the two. We add to Andreoni's idea of egoism and impure altruism by focusing on how impure altruism may arise from two distinct sources, i.e. situation and disposition, modeling the effect of situation and disposition through a psychic cost function. Moreover, just as Andreoni (1995) notices the behavioral asymmetry between the warm-glow of doing something good and the cold-prickle of doing something bad we uncover a different type of behavioral asymmetry, one not based on the consequences of the behavior but based on the disposition of people and the interaction between situation and disposition.

We also endogenize our measure of impure altruism by allowing the situational effects of a nudge to vary with the individual's own choices. A common behavioral policy intervention is to tell individuals how their behavior compares to the behavior of others. However, if implemented on a large scale, the "behavior of others" also depends on the intervention. Rotemberg (1994) examines how altruism may arise endogenously among a small set of strategically related individuals. Bowles (1998) discusses how markets and other economic institutions may give rise to endogenous preferences by changing the exogenous determinants in a cultural equilibrium. Casadesus-Masanell (2004) studies how principals may utilize motivational schemes that rely on social influences such as norms, ethical standards, and altruism to foster intrinsic motivation and trustworthy behavior. Our interest differs from theirs in that we are less interested in how these preferences arise. Instead we are interested in how the government needs to account for these preferences when considering

behavioral policy interventions. We formalize such endogeneity by deriving the optimal message that policy makers should send regarding the behavior of others and show that the optimal message should take into account the population's heterogeneity as well as the complementarity between dispositional and situational altruisms.

We discuss above examples of heterogeneity in response to such messages in trials involving messages about water usage (Ferraro and Price, 2013) or electricity usage (Alcott, 2011) or carbon footprint (Ho, 2014). Other studies in the lab and field shed light about the mechanisms for how information about the behavior of others influences behavior. For example, Azmat and Irberri (2010) examine the effect of a natural experiment where high school students learn their relative rank. They find the strongest effect of such information on the tails of the distribution and argue that such information works because people have competitive preferences that make especially high or low rankings more meaningful.

A different type of heterogeneity has been demonstrated in field data for cashiers (Mas and Moretti, 2009) and envelope stuffers (Falk and Ichino, 2006) who observe the performance of peers. Both find that working next to higher performing peers induces higher performance, but primarily for low performing workers. Mas and Moretti (2009) consider the alternative mechanisms of social pressure, contagious enthusiasm and knowledge spillovers, and argue that social pressure to conform to high performance is the driving mechanism.

Masclet et al. (2003) identify a different source of potential heterogeneity that operates through the mechanism of social pressure. They show that nonmonetary punishments, i.e. the verbal expression of disapproval will induce some people to contribute to a public good. Individuals

of different dispositions would likely respond differently to such nonmonetary sanctions. A similar nonmonetary effect can be seen in Ball et al. (2001) and Eckel et al. (2010) which show that people respond to arbitrarily assigned status in both market and public goods games. Knowing how you rank in some meaningless task domain affects behavior in other domains.

The main limitation to all of these studies is that even those that consider heterogeneity, tend to focus on a single intervention of fixed size. This paper highlights the need to consider how the heterogeneity in the response to a nudge varies as the size of the nudge varies. This paper studies such variation in a general model that accounts both for economic mechanisms such as social disapproval, peer comparisons, and status, but also for more psychological mechanisms like moral regulation and guilt.

Our paper further explores the policy implications of behavioral nudges in light of optimal tax treatments. Our tax model builds on Sandmo's (1975) model for optimal commodity tax, but with a damage that comes both from the externality as well as a psychic cost. Johansson (1997) and Diamond (2005) also consider warm-glow preferences to optimal tax calculations, but unlike Diamond, we focus on the cold-prickle from the creation of a negative externality rather than warm-glow, and additionally explore the interaction effects between taxes and impure altruism.

3 Main Model & Propositions

3.1 Model

Assume that there is a group of I individuals who are impurely altruistic, $I \geq 2$ and $I \in N^*$. The i^{th} individual chooses a quantity of g_i units of a good with negative externality to the whole group.

We first evaluate the individual's decision problem. Each individual chooses g^* to maximize the following individual utility function:

$$u_i = V(g_i) - N(G) - c(g_i, \theta, \omega) \quad (1)$$

where $V(g_i)$ is the value of g_i units of the good to the individual, $N(G)$ is the per person cost due to the negative externality of the total consumption of the good, and G is the total number of units consumed by everyone, i.e $G = \sum_{i=1}^{I} g_i$. $\frac{dV}{dg}$ is the marginal benefit for each additional unit of good consumed, and $\frac{d^2V}{dg^2}$ measures the change in the marginal benefit for each additional unit of good consumed. $\frac{dN}{dG}$ is the marginal cost for each individual for each additional unit of good consumed by the group, and $\frac{d^2N}{dG^2}$ is the change in the marginal cost for each individual for each additional unit of good consumed by the group.

The cost function, $c(g, \theta, \omega)$, is based on the units of good consumed (g), the individual's level of dispositional altruism (θ), and the situational altruism induced by nudges (ω). Assuming θ and ω are both exogenous, $\frac{\partial c}{\partial g}, \frac{\partial c}{\partial \theta}$, and $\frac{\partial c}{\partial \omega}$ are respectively the marginal cost of consumption conditional on psychological factors, the marginal psychic costs due to disposition, and the marginal psychic costs induced by nudges. The second derivative, $\frac{\partial^2 c}{\partial g^2}$, is the change in the marginal cost of good for each additional unit of good consumed. The cross derivatives, $\frac{\partial^2 c}{\partial \theta \partial g}$ and $\frac{\partial^2 c}{\partial \omega \partial g}$, are our proxies for the interaction between consumption and impure altruism. Specifically, $\frac{\partial^2 c}{\partial \theta \partial g}$ and $\frac{\partial^2 c}{\partial \omega \partial g}$ measure how the marginal cost of consumption changes respectively with people's dispositional and situational altruism. We use the concept of impure altruism and culpability interchangeably in the model as in Andreoni (1995), the warm glow of altruism you get from reducing consumption is equal to the cold prickle of guilt you feel when you increase consumption.

Finally, we examine the socially optimal level of consumption, G^{**}, which is found by maximizing the sum of utilities for the population U_G:

$$U_G = IV\left(\frac{G}{I}\right) - IN(G) - Ic\left(\frac{G}{I}, \theta, \omega\right). \quad (2)$$

3.2 Assumptions

The main model and its propositions are based on the following assumptions:

(A1) All individuals have the same utility function.

(A2) $\frac{dN}{dG} > 0$ and $\frac{d^2N}{dG^2} \geq 0$.

(A3) $V(g)$ is concave and increasing in g, that is, $\frac{dV}{dg} > 0$ and $\frac{d^2V}{dg^2} < 0$.

(A4) The impure altruism or culpability variables, θ and ω, are exogenous, that is, they are determined externally and do not change as g changes.

(A5) $\frac{\partial c}{\partial \theta} > 0, \frac{\partial c}{\partial \omega} > 0$, and $\frac{\partial c}{\partial g} > 0$. Purchase and transaction costs are subsumed by the cost function along with the psychic costs.

(A6) $\frac{\partial^2 c}{\partial g^2} \geq 0$.

(A7) Interaction between consumption and culpability: $\frac{\partial^2 c}{\partial \theta \partial g} > 0$ and $\frac{\partial^2 c}{\partial \omega \partial g} > 0$.

(A1)-(A7) are typical concavity assumptions. (A7) argues that the cold prickle of culpability increases in strength for larger deviations.

3.3 Propositions

Propositions 1 through 4 reproduce the standard tragedy of the commons results. We describe them briefly here but leave the details to the Appendix. Proposition 1 says that the individually rational choice of consumption of the anti-social good g^* and G^* is greater than the socially

optimal level of consumption, g^{**} and G^{**}. Proposition 2 says that consumption of the anti-social good is decreasing in dispositional or situational altruism. We interpret high situational altruism, ω, as the effect of moral cleansing. A person who is made to feel situationally guilty will act more altruistically to offset that guilt. We interpret low situational altruism as the effect of moral licensing. A subject who feels they have done something moral will consume more of the polluting good, g^{*}.

Proposition 3 and 4 evaluate the effect of impure altruism on utility and welfare. Proposition 3 says that if the marginal disutility of increased altruism is greater (less) than the marginal utility from mitigating the externality, then altruism is welfare decreasing (increasing). In other words, higher level of either situational or dispositional guilt potentially reduces individual utility (*1*). Increased guilt increases psychic cost but reduces consumption of the dirty good. The net effect on welfare (*2*) depends on the relative magnitudes between the two. Proposition 4 looks at social welfare loss, the gap between first best and equilibrium consumption as defined in Proposition 1, and decomposes the effect of altruism on welfare loss into two parts. The first part captures welfare gains due to change in marginal psychic costs. The second part captures welfare loss due to change in marginal social externality at the equilibrium outcome. The net effect of impure altruism on social welfare loss depends on the relative magnitude of these two parts.

The heterogeneous effects of nudges are determined by the relationship between θ and ω, specifically, the cross derivative, $\frac{\partial^2 c}{\partial \omega \partial \theta}$ and $\frac{\partial^2 c}{\partial \theta \partial \omega}$. The cross derivatives indicate how the marginal psychic costs of additional guilt of one type change with the level of guilt of another type. From here, we assume $c(*)$ is continuous, so $\frac{\partial^2 c}{\partial \omega \partial \theta} = \frac{\partial^2 c}{\partial \theta \partial \omega}$. If $\frac{\partial^2 c}{\partial \omega \partial \theta}$ or $\frac{\partial^2 c}{\partial \theta \partial \omega} > 0$, we say θ and ω are

complements. If $\frac{\partial^2 c}{\partial \omega \partial \theta}$ or $\frac{\partial^2 c}{\partial \theta \partial \omega} < 0$, we say θ and ω are substitutes. We need to make one additional assumption before moving on to Proposition 5:

(A8) The psychic cost function is linear in g, that is, $\frac{\partial^2 c}{\partial g^2} = 0$. In this case, the cost function can be written as $c(g, \theta, \omega) = g \cdot c'(\theta, \omega)$.

Intuitively it means that marginal psychic cost remains constant as consumption increases (i.e. there is no "income" effect for psychic costs).

***Proposition 5: Interaction between* θ *and* ω**. *Assuming (A2), (A3), (A6), and (A8), we have the impacts of situation and disposition are complements (substitutes) if and only if the impacts of situation and disposition on psychic costs are complements (substitutes):*

$$sign\left(\frac{d^2 g^*}{d\omega d\theta}\right) = -sign\left(\frac{\partial^2 c}{\partial \omega \partial \theta}\right).$$

Symmetrically, we similarly have

$$sign\left(\frac{d^2 g^*}{d\theta d\omega}\right) = -sign\left(\frac{\partial^2 c}{\partial \theta \partial \omega}\right).$$

(See Appendices E. Proof of Proposition 5.)

This proposition effectively says:

(i) When situation and disposition are substitutes:

(a.) ω has a larger effect on g if θ is smaller.

That is, situational guilt has a larger effect on people inclined to be bad (versus people inclined to be good).

(b.) θ has a larger effect on g if ω is smaller.

That is, disposition has a larger effect on people who experience low situational

guilt (versus people who experience high situational guilt).

(ii) When situation and disposition are complements:

(a.) ω has a larger effect on g if θ is larger.

That is, situational guilt has a larger effect on people inclined to be good (versus people inclined to be bad).

(b.) θ has a larger effect on g if ω is larger.

That is, disposition has a larger effect on people who experience high situational guilt (versus people who experience low situational guilt).

Therefore,

(i) If $\frac{\partial^2 c}{\partial \omega \partial \theta} > 0$ (θ and ω are complements) then $\frac{d^2 g^*}{d\omega d\theta} < 0$.

(ii) If $\frac{\partial^2 c}{\partial \omega \partial \theta} < 0$ (θ and ω are substitutes) then $\frac{d^2 g^*}{d\omega d\theta} > 0$.

Proposition 5 begs the question: which types of situational nudges are complements and which types are substitutes. We can certainly speculate based on the mechanisms behind such nudges in the literature. For example, there is some evidence that nudges that appeal to social pressure would be more effective on those who have a disposition to be altruistic (e.g., Ho, 2014). Alternatively, nudges that threaten one's status may be counter-productive to those of an altruistic disposition. Peer effects for work performance seem to be a substitute for those with an intrinsic disposition toward good performance (Falk and Ichino, 2006; Mas and Moretti, 2009). However, to our knowledge there has been no systemic study of the complementarity between nudges and dispositional altruism.

The following section offers some guidance on how one might answer the complementarity

question by re-examining existing data for one type of situational intervention, the peer information nudge that tells people how their behavior compares to others.

3.4 Endogenous Situational Altruism

Up until now we have considered only cases where situation and disposition were exogenous. To understand the specific case of peer information nudges where people are targeted with a descriptive message telling them about the behavior of their peers, we consider the case where ω changes with g — that is, my situational guilt increases the more I choose to consume. We introduce $\hat{g}$ as an informational nudge where ω is a function of g and $\hat{g}$. Think of $\hat{g}$ as the informational message selected by the nudge designer that reads "other people like you chose to consume $\hat{g}$ units of the good." Clearly, such messages only work if this message, $\hat{g}$, differs from the person's prior belief about the consumption of others. We denote the prior beliefs regarding the consumption of others to be $\widehat{g_o}$. Even a truth telling policy maker has quite a bit of discretion of deciding $\hat{g}$, because the term "others like you" could refer to many different comparison groups (e.g. people in your neighborhood, or in your state, or in your age group) (Hallsworth et al 2014).

The key assumption for our model of endogenous altruism is that my desire to curtail consumption declines as I cut my own consumption, $\frac{\partial \omega}{\partial g} > 0$, but increases if I think others are cutting their consumption, $\frac{\partial \omega}{\partial \hat{g}} < 0$.

Proposition 6: The Case of Endogenous ω. *Assuming (A2), (A3), (A5), (A6) and (A7), if situational altruism depends on g such that $\frac{\partial \omega}{\partial g} > 0$, $\frac{\partial^2 \omega}{\partial g^2} > 0$, and $\frac{\partial \omega}{\partial \hat{g}} < 0$, then*

(i) $\frac{dg^*}{d\hat{g}} > 0$, *and (ii)* $\frac{dg^*}{d\theta} < 0$.

(See Appendices F. for Proof of Proposition 6.)

Proposition 6 reproduces our main results from Propositions 1-4 for the case of endogenous ω. In other words, when there is a positive complementarity between situational culpability and consumption, informational nudges influence social behavior by providing a reference point.

We use these results to derive the optimal message $\hat{g}$ so that individuals are nudged to choose the consumption level g that is optimal for society. In line with the discussion of exogenous impure altruism in section 3.3, we know that individuals choose $g^*(\widehat{g_0}, \theta)$, resulting in a total social utility of $I \cdot g^*(\widehat{g_0}, \theta)$, while a social planner would choose $G^{**}(\widehat{g_0}, \theta) < I \cdot g^*(\widehat{g_0}, \theta)$. Note that as specified by (*4*), G^{**} does not account for the psychic costs of receiving a social nudge. While this assumption is plausible given that current policy rarely considers psychic costs, it is not innocuous. We will see later that as in Propositions 3 and 4, impure altruism makes the welfare implications of nudges less straightforward. For now, we maintain the assumption that the social planner, knowing the individual demand function $g^*(\cdot, \theta)$ and disposition, θ, chooses an optimal message $\hat{g}_{opt}$ such that

$$I \cdot g^*\left(\hat{g}_{opt}, \theta\right) = G^{**}(\widehat{g_0}, \theta).$$

Proposition 7: Optimal Message under Endogenous ω. *In addition to the assumptions in Proposition 6, assuming (A6) and some technical conditions, a solution* $\hat{g} = \hat{g}_{opt}$ *to the following equation exists:*

$$I \cdot g^*(\hat{g}, \theta) = G^{**}(\widehat{g_0}, \theta).$$

Moreover, the optimal message depends on the dispositional altruism of the population:

$$\frac{d\hat{g}_{opt}}{d\theta} \neq 0.$$

(See Appendices G. for Proof of Proposition 7.)

The optimal message is chosen such that each individual maximizes

$$u_i = V(g_i) - N(G) - c\big(g_i, \theta, \omega(g_i, \hat{g}_{opt})\big), \tag{3}$$

and the social planner also maximizes total social utility (excluding psychic costs) given $\widehat{g_0}$ (we can think of $\widehat{g_0}$ as the true average consumption in the population),

$$U_G = IV\left(\frac{G}{I}\right) - IN(G) - Ic\left(\frac{G}{I}, \theta, \omega(\frac{G}{I}, \widehat{g_0})\right). \tag{4}$$

One caveat is that when the optimal message is chosen as such, the psychic cost function complicates the welfare implications. In a standard public good model without psychic costs, nudging individual consumption toward the best social choice has clear welfare improving consequences. However, when there is impure altruism, a message that induces a cold prickle would increase psychic costs and thus reduce individual utility. As in Proposition 3 and 4, the welfare loss depends on the degree of complementarity between situational and dispositional altruism.

In essence, the heterogeneous effects of informational nudges hinge on three factors: situational altruism induced by nudges (ω), long-run dispositional altruism (θ), and the relationship between the two. Differences in ω or θ alone for different groups of people can explain many cases of heterogeneity in consumption, but the net outcome is less obvious when they are inconsistent (e.g. ω high and θ low). Also, whether situation and disposition are viewed as complements or substitutes affects the effectiveness of nudges for certain types of people. A nudge

that induces situational guilt has a larger effect on people inclined to be bad (versus good) if people view situation and disposition as substitutes, but the nudge has a larger effect on people inclined to be good (versus bad) if people view situation and disposition as complements. Modeling ω as either exogenous or endogenous does not affect our conclusions. From the standpoint of policy interventions with the goal of enhancing the effectiveness of informational nudges, if a nudge can be designed so that people view it as a substitute for disposition, then it may be used to target people inclined to be bad. If a nudge can be designed so that people view it as a complement to disposition, then it may be used to target people inclined to be good.

In light of the results suggested by our model, additional or alternative measures should be considered when informational provision is less effective for certain groups of people. Some other forms of prevalent situational nudges include commitment devices, default options, implementation intentions, and exploitation of nonlinear demand curves (Allcott and Mullainathan, 2010). The identification of the relationship between situation and disposition, however, can be demanding. While such inferences may be possible by re-analyzing existing data, more research is necessary to avoid potentially adverse consequences of nudge-based policies.

4 The Role of Government: A Model of Taxation

As noted in the general model in Section 3, social welfare loss due to choices made by self-centered individuals may serve as the rationale behind governmental intervention. One possible way to alter choice of consumption is through the use of informational nudges, as discussed in the previous sections. Such interventions are less conventional but are favored by many researchers and

policy-makers over commands, requirements, and prohibitions (Thaler and Sunstein, 2008). Nevertheless, more conventional governmental interventions have been in place for a longer time and are still worth studying. This section will be devoted to the discussion of one specific type of traditional intervention – taxation. This section aims to provide insights into the optimal tax in the presence of dispositional and situational pro-sociality.

We first consider the first best policy, where government chooses both taxes and nudges optimally. However, policy makers often see nudges as a way to influence behavior when taxes are politically infeasible. In such cases, we hope our model serves as a reminder that nudges are not designed in a vacuum, and should be designed to complement existing tax policies.

4.1 Impure Altruism and Taxation

To maximize total social welfare, the government can impose a Pigouvian corrective tax[3] of T on the purchase of each unit of the good to affect the choice of g. For simplicity, we focus this discussion on a combined measure of impure altruism parameter $\alpha(\theta, \omega)$. Let the density function of this combined function be $\phi(\alpha)$, $\alpha \in [\underline{\alpha}, \overline{\alpha}]$. Assumptions are adjusted to focus on α:

(A1') All the individuals have the same utility function.

(A2') $\frac{dN}{dG} > 0$ and $\frac{d^2N}{dG^2} \geq 0$.

(A3') $V(g)$ is concave and increasing in g, so $\frac{dV}{dg} > 0$ and $\frac{d^2V}{dg^2} < 0$.

(A4') The impure altruism/culpability variable, α, is exogenous, that is, it is determined externally

[3] A Pigouvian tax is a tax applied to a market activity that is generating negative externalities. The original argument is by Arthur C. Pigou in his work in 1920, *The Economics of Welfare*.

and do not change as g changes.

(A5') $\frac{\partial c}{\partial \alpha} > 0, \frac{\partial c}{\partial g} > 0,$ and purchase costs are subsumed by the cost function.

(A6') $\frac{\partial^2 c}{\partial g^2} \geq 0.$

(A7') Interaction between consumption and culpability: $\frac{\partial^2 c}{\partial \alpha \partial g} > 0.$

The relationship between the following propositions denoted in terms of α and the disposition and situation parameters θ and ω, follow in a straightforward way through application of the chain rule. We assume $\frac{\partial \alpha}{\partial \theta} > 0$ and $\frac{\partial \alpha}{\partial \omega} > 0$, and that the interaction between situation and disposition is again governed by the complementarity parameter $\frac{\partial^2 \alpha}{\partial \theta \partial \omega}$. We return to the three way interaction between tax rates, situation and disposition in section 4.4.

Assume the government knows the distributions of α. Moreover, the government also knows the response function of people, $g^* = h(\alpha, T)$, based on the tax T set by the government for each unit of consumption. People take T as exogenous, and $g^* = h(\alpha, T)$ is the solution to the utility maximization problem faced by each individual:

$$\max_g s = V(g) - N(G) - c(\alpha, g) - gT = u - gT.$$

Total social welfare is

$$\Pi = s_1 + \cdots + s_I + GT = u_1 + \cdots + u_I - (g_1 + \cdots + g_I)T + GT$$

$$= \sum_1^I u_i = \sum_1^I [V(g_i) - N(G) - c(\alpha, g_i)].$$

Since α is a random variable, the government maximizes the expected value of Π:

$$\max_T E[\Pi] = \sum_1^I E[V(g_i^*) - N(G^*) - c(\alpha, g_i^*)].$$

4.2 Stochastic Dominance of Impure Altruism and Effect on Taxation

This section explores the stochastic dominance of impure altruism and the corresponding effect on government's tax decision. First order stochastic dominance (FSD) transformations refer to stochastically larger impure altruism in the society and second order stochastic dominance (SSD) transformations refer to stochastically less volatile impure altruism in the society.

We now make pragmatic adjustments to the utility function of each individual by taking into consideration a budget constraint faced by individuals, b, and a real price of the good, x. Both b and x are positive, and we assume $x < b$. So each individual maximizes the following problem:

$$\max_{g} s = V(g) - N(G) - c(\alpha, g) - Tg - xg,$$

subject to

$$g(x + T) \leq b, g \geq 0.$$

These adjustments do not change Π, so the government faces the same problem:

$$\max_{T \geq 0} E[\Pi(T, \alpha)]$$

where, as we recall, $\Pi = \sum_1^I u_i$. Following Ormiston (1992), we have the next two propositions.

Proposition 8: First Order Stochastic Dominance. *By Lemma 8.1 (See Appendices H.), there is an existence of interior solution, T^*, for first order statistical dominance (FSD). T^* increases (decreases) for all FSD transformations of α if $u_{T\alpha}(T, \alpha) \geq 0$ ($u_{T\alpha}(T, \alpha) \leq 0$) everywhere.*

(See Appendices I. for Proof of Proposition 8.)

Therefore,

(i) The government will increase tax when impure altruism is stochastically larger in the

society when $\frac{d^2g^*}{dTd\alpha} \geq 0$, i.e. when tax and impure altruism are complements.

(ii) The government will decrease tax when impure altruism is stochastically larger in the society when $\frac{d^2g^*}{dTd\alpha} \leq 0$, i.e. when tax and impure altruism are substitutes.

Proposition 9: Second Order Stochastic Dominance. *T^* increases (decreases) for all SSD transformations of α if $u_{T\alpha}(T,\alpha) \geq 0$ and $u_{T\alpha\alpha}(T,\alpha) \leq 0$ ($u_{T\alpha}(T,\alpha) \leq 0$ and $u_{T\alpha\alpha}(T,\alpha) \geq 0$) everywhere.*

(See Appendices J. for Proof of Proposition 9.)

Therefore,

(i) The government will increase tax when impure altruism is stochastically less volatile provided that $\frac{d^2g^*}{d\alpha dT} \geq 0$ and $\frac{d}{d\alpha}\left(\frac{d^2u}{d\alpha dT}\right) \leq 0$, i.e. when tax and impure altruism are complements and $\frac{d}{d\alpha}\left(\frac{d^2u}{d\alpha dT}\right) = T \cdot \frac{d}{d\alpha}\left(\frac{d^2g^*}{d\alpha dT}\right) \leq 0$.

(ii) The government will decrease tax when impure altruism is stochastically less volatile provided that $\frac{d^2g^*}{d\alpha dT} \leq 0$ and $\frac{d}{d\alpha}\left(\frac{d^2u}{d\alpha dT}\right) \geq 0$, i.e. when tax and impure altruism are substitutes and $\frac{d}{d\alpha}\left(\frac{d^2u}{d\alpha dT}\right) = T \cdot \frac{d}{d\alpha}\left(\frac{d^2g^*}{d\alpha dT}\right) \geq 0$.

It is worth noting here that the complementarity we discuss in this proposition is a different complementarity than the one discussed in section 3. Whereas before, we were interested in how dispositional altruism and situational altruism interacted, Proposition 8 and 9 describe the altruism between impure altruism and taxes. We discuss the relationship between all three of these—taxes, dispositional altruism, and situational altruism—in section 4.4.

4.3 Taxation under Binary Types

To better understand the intuition of how government should set taxes in relation to altruistic preferences, consider government's taxation decision for a more specific functional form where there are only two types of α. Assume $\alpha \in \{\alpha^H, \alpha^L\}$, where α^H denotes high impure altruism and α^L denotes low impure altruism. More specifically,

$$\alpha = \begin{cases} \alpha^H, \text{with a probability of } p \\ \alpha^L, \text{with a probability of } (1-p) \end{cases}$$

where $p \in [0,1]$.

Proposition 10: Taxation under Binary Types of Impure Altruism. *Suppose the same assumptions needed for Proposition 8 and 9 also hold, we have*

$$sign\left(\frac{dT^*}{dp}\right) = sign\left(\frac{d^2g^*}{d\alpha dT}\right).$$

(See Appendices K for Proof of Proposition 10.)

Therefore, if $\frac{d^2g^*}{d\alpha dT} > 0$, that is, tax and impure altruism are complements, then the government will tax more if people are more likely to be impurely altruistic. If $\frac{d^2g^*}{d\alpha dT} < 0$, that is, tax and impure altruism are substitutes, then the government will tax less if people are more likely to be impurely altruistic. This proposition is consistent with the results in Section 4.2 since the binary assumption for impure altruism is just a specific case for the distribution of α.

So far, our discussion about the taxation implications has focused on the single parameter α that captures the randomness in impure altruism among the population. As mentioned earlier, α is defined as a function of both situation and disposition. In the next section, we decompose α to account for the individual effect of θ and ω.

4.4 Taxation under Impure Altruism Interactions

In this section, we assume that the government takes into consideration both dispositional altruism θ and situational altruism ω when determining the corrective tax, T. θ differs for each person in the economy with a density function of $\varphi(\theta)$, $\theta \in [\underline{\theta}, \overline{\theta}]$. ω also differs for each person in the economy with a density function of $\psi(\omega)$, $\omega \in [\underline{\omega}, \overline{\omega}]$. We assume that θ and ω are independently distributed. Assumptions are as (A1) - (A7) in Section 3.2. We make an additional assumption analogous to (A8) in the derivation of Proposition 5:

(A8') The psychic cost function $c(\theta, \omega, g) = \tilde{c}(\theta, \omega) \cdot \hat{c}(g)$ where $\frac{\partial \tilde{c}}{\partial \theta} > 0, \frac{\partial \tilde{c}}{\partial \omega} > 0, \frac{\partial \hat{c}}{\partial g} > 0$, and $\frac{\partial^2 \hat{c}}{\partial g^2} \geq 0$.

Assume the government knows the distributions of θ and ω. Moreover, the government also knows the response function , $g^* = h(\theta, \omega, T)$, based on the tax T set by the government for each unit of consumption. People take T as exogenous, and $g^* = h(\theta, \omega, T)$ is the solution to the utility maximization problem faced by each individual:

$$\max_g s = V(g) - N(G) - c(\theta, \omega, g) - gT = u - gT.$$

Total social welfare is

$$\Pi = s_1 + \cdots + s_I + GT = \sum_1^I u_i.$$

The government maximizes the expected social welfare:

$$\max_T E[\Pi] = \sum_1^I E[V(g_i^*) - N(G^*) - c(\theta, \omega, g_i^*)].$$

Proposition 11: Effect of Impure Altruism on Optimal Tax. *Assuming (A2), (A3), (A6), and (A8'), we have*

$$\frac{dT^*}{d\theta} \geq 0 \ \ and \ \ \frac{dT^*}{d\omega} \geq 0.$$

The equalities hold if and only if $\frac{\partial^2 \hat{c}}{\partial g^2} = 0$.

(See Appendices L. for Proof of Proposition 11.)

While our earlier results about the stochastic dominance of impure altruism suggest that the effect on taxation depends on the relationship between the tax rate and the degree of altruism, here the additional assumption (A8') allows us to determine the direction of the effect. When the psychic cost function is separable in terms of the impure altruism component and the consumption component, we derive that the optimal tax is non-decreasing in one type of impure altruism conditional on the other type.

The next proposition allows us to better understand the optimal tax level in light of two types of interaction. The relationship between each person's individual situation and disposition based on one's own psychic cost functions is referred to as individual substitutes (complements). The relationship between the population's situation and disposition and the choice of tax rate is referred to as policy substitutes (complements).

Proposition 12: Optimal Tax under Interacting Impure Altruism. *Assuming (A2), (A3), (A6), and (A8'), we have that the optimal tax depends on the relationship between situation and disposition:*

a. When the impacts of situation and disposition on psychic costs are individual substitutes, then the impacts of situation and disposition on optimal tax are policy complements:

$$\frac{d^2\tilde{c}}{d\omega d\theta} < 0 \Rightarrow \frac{d^2T^*}{d\omega d\theta} > 0.$$

b. When the impacts of situation and disposition on optimal tax are policy substitutes, we need the impacts of situation and disposition on psychic costs to be sufficiently individually complementary:

$$\frac{d^2T^*}{d\omega d\theta} < 0 \; if \; and \; only \; if \; \frac{d^2\tilde{c}}{d\omega d\theta} > 0 \; and \; large \; enough.$$

(See Appendices M. for Proof of Proposition 12.)

This proposition effectively says:

(i) When situation and disposition are individual substitutes:

a) ω has a larger effect on T^* if θ is larger.

Situation has a larger effect on the optimal tax when people are inclined to be good. The optimal dispersion of tax rates for subpopulations experiencing different levels of situational guilt is larger when the population disposition is more altruistic. If tax rates cannot vary by situational guilt, the inefficiency becomes larger when the population has a more altruistic disposition.

b) θ has a larger effect on T^* if ω is larger.

Disposition has a larger effect on the optimal tax when people experience high situational guilt. The optimal dispersion of tax rates for subpopulations with different dispositions for altruism is larger when the experienced situational guilt is higher. If tax rates cannot vary by dispositions, then the inefficiency becomes larger when the population is more situationally guilty.

(ii) When situational and dispositional altruism are individual complements, the interaction effect on T^* is ambiguous. However, for situation and disposition sufficiently complementary, we have a symmetric results to (i).

a) ω has a larger effect on T^* if θ is smaller.

Situation has a larger effect on the optimal tax when people are inclined to be bad. Optimal tax dispersion for subpopulations experiencing different levels of situational guilt is larger when the population disposition is lower. If tax rates cannot vary by situational guilt, the inefficiency becomes larger when the population has a less altruistic disposition.

b) θ has a larger effect on T^* if ω is smaller.

Disposition has a larger effect on the optimal tax when people experience low situational guilt. The optimal tax dispersion for subpopulations with different dispositions is larger when the experienced situational guilt is lower. If tax rates cannot vary by situational guilt, the inefficiency becomes larger when the population is less situationally guilty.

5 Discussion: Implications for Policy

As mentioned in the introduction, local authorities and central governments in the UK, US, and France have been avidly engaged in the use of informational nudges to increase tax pay-offs, to encourage more retirement savings, to improve energy efficiency, just for a start. However, Akerlof and Kennedy (2013) note limited success over the past 15 years of campaigns using messages of social comparison in areas including drug and alcohol use, recycling, and littering. Our model shows

that informational nudges may have very limited impact for three reasons. First, a nudge is less effective when it fails to induce high situational altruism. Cialdini et al. (2006) point out that informational nudges, when purely descriptive, may not be effective in mitigating socially disapproved conduct and may even worsen the problem. To address this counterproductive licensing effect, policy makers should pay attention to how social messages are conveyed.

Second, our model shows that nudges may have very limited impact if low dispositional altruism is dominant in the targeted population. Hence, it is crucial for the government to understand the disposition of certain groups of people when designing public programs to induce pro-social behavior through nudges. Moreover, better understanding of people's disposition will allow the government to target certain sub-populations and use its resources more efficiently.

Third, even when a nudge successfully induces high situational altruism and when people's disposition can be identified, the effectiveness of nudges still depend on the relationship between nudges and disposition. While existing work on mechanisms behind situational nudges can provide guidance on this relationship, experiments should be re-analyzed and re-designed to test theories about how certain sub-populations' situation and disposition interact, and how different types of nudges target different sub-populations. If a nudge can be designed so that people view them as a substitute for disposition, then it may be used to target people with a bad disposition. If a nudge can be designed so that people view it as a complement to disposition, then that nudge should be used to target people inclined to be good.

Some types of dispositional impure altruism include conformity, moral obligation, group identity, et cetera. Some examples of informational nudges that can induce situational altruism

include the use of heuristics (anchoring, availability, and representativeness), framing, emotion arousing, feedback, depiction of social norms, and peer comparison (Sunstein and Thaler, 2008). Others are based on social disapproval, status, or social pressure (Kandel and Lazear, 1992). While our model suggests that pro-social behavior following a nudge is affected by the relationship between disposition and situation, we have not addressed how to identify whether disposition and situation are complements or substitutes in a specific real life case. However, our model provides a foundation for testing hypotheses based on field experiments. For instance, if situation and disposition are complements, then according to our model, a nudge that induces situational guilt will be more effective for those with high disposition. Costa and Kahn (2012) find in a field experiment that environmentalists are more responsive to green nudges than the average person, implying that situation and disposition are very likely complements in this case.

This paper also considers taxes as an alternative or supplemental tool of public intervention in inducing socially optimal behavior and that the optimal taxes depend on the relationship between the tax and impure altruism. Consequently, two important implications arise. First, if we consider the effect of taxes on dispositional altruism, then identifying sub-populations with different distributions of θ and targeting them with different optimal taxes can be eventually welfare enhancing. Alternatively, if we consider the effect of taxes on situational altruism induced by nudges, then the optimal taxes should be responsive to the behavioral policies being implemented. Also, while in principle, it would be possible to implement a tax system that induces truthful revelation of private information about one's disposition for altruism along the lines of Mirrlees (1971), in practice such a tax policy would likely require large information rents and would be unlikely in practice. Our tax

results, therefore, also provide guidance regarding the size of the welfare loss, when tax policy cannot account for the altruism of the population. In our framework, it is also worth noting that given a choice between a tax and a nudge, the tax is generally preferable. Assuming the government revenue is efficiently allocated, the tax costlessly reduces the externality, while the nudge imposes additional psychic costs.

6 Conclusions

Ideally, informational nudges can be used as tools to induce pro-social behavior at the individual level and to reduce negative externality at the public level. The ultimate question is whether or when the nudges are effective. When nudges fail to induce high situational altruism, they are less effective or can even have counterproductive moral licensing effect. But even when nudges do induce high situational altruism, the pro-social impact of a nudge can still be foiled if the intervention targets low disposition individuals. Furthermore, the interaction between situational nudges and disposition serves as an additional explanation for the heterogeneous effects. An effective nudge has a larger impact on people inclined to be bad if people view situation and disposition as substitutes, but the nudge has a larger impact on people inclined to be good if people view situation and disposition as complements. Our conclusions can also be extended to environments with multiple goods or multiple domains.

From a government standpoint, taxes serve as the standard tool used to enhance social welfare given externalities. We show that the optimal tax level is dependent on the distribution of preferences for altruism (both dispositional and situational) as well as the interdependence between

people's altruistic preferences. We show that limiting the tax base by targeting taxes on certain sub-populations may be welfare enhancing.

Finally, the relationship between situation and disposition is unobservable and relatively unexplored. So is the relationship between taxes and impure altruism. This paper aims to fill the theoretic gap in the literature for the policy implications of nudges and to provide a framework for future empirical research uncovering these relationships.

REFERENCES

Akerlof, George A. and Rachel E. Kranton. "Economics and Identity." *The Quarterly Journal of Economics* 115, no. 3 (2000): 715-753.

Akerlof and Kennedy. "Nudging toward a Healthy Natural Environment: How Behavior Change Research can Inform Conservation." June 10, 2013. http://climatechangecommunication.org/sites/default/files/reports/NudgesforConservation_GMU_061013.pdf, accessed November 13, 2014.

Allcott, Hunt. "Social Norms and Energy Conservation." *Journal of Public Economics* 95, no. 9 (2011): 1082-1095.

Allcott, Hunt and Sendhil Mullainathan. "Behavioral Science and Energy Policy." *Science* 327, no. 5970 (2010): 1204-1205.

Andreoni, James. "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving." *The Economic Journal* 100, no. 401 (1990): 464-477.

Andreoni, James. "Warm-Glow Versus Cold-Prickle: The Effects of Positive and Negative Framing on Cooperation in Experiments." *The Quarterly Journal of Economics* 110, no. 1 (1995): 1-21.

Ayal, Shahar and Francesca Gino. "Honest Rationales for Dishonest Behavior." The Social Psychology of Morality: Exploring the Causes of Good and Evil. Washington, DC: American Psychological Association (2011).

Ayres, Ian, Sophie Raseman, and Alice Shih. "Evidence from two large field experiments that peer comparison feedback can reduce residential energy usage." *Journal of Law, Economics, and Organization* (2012): ews020.

Azmat, Ghazala, and Nagore Iriberri. "The importance of relative performance feedback information: Evidence from a natural experiment using high school students." *Journal of Public Economics* 94, no. 7 (2010): 435-452.

Ball, Sheryl, Catherine Eckel, Philip J. Grossman, and William Zame.. "Status in markets." Quarterly Journal of Economics - 116.1 (2001): 161-18

Berger, Jonah and Lindsay Rand. "Shifting Signals to Help Health: Using Identity Signaling to Reduce Risky Health Behaviors." *Journal of Consumer Research* 35, no. 3 (2008): 509-518.

Bergstrom, Theodore, Lawrence Blume, and Hal Varian. "On the private provision of public goods." *Journal of public economics* 29, no. 1 (1986): 25-49.

Beshears, John, James J. Choi, David Laibson, Brigitte C. Madrian, and Katherine L. Milkman. *The effect of providing peer information on retirement savings decisions*. No. w17345. National Bureau of Economic Research, 2011.

Bowles, Samuel. "Endogenous preferences: The cultural consequences of markets and other economic institutions." *Journal of economic literature* 36, no. 1 (1998): 75-111.

Casadesus-Masanell, Ramon. "Trust in agency." *Journal of Economics & Management Strategy* 13, no. 3 (2004): 375-404.

Cialdini, Robert B., Linda J. Demaine, Brad J. Sagarin, Daniel W. Barrett, Kelton Rhoads, and Patricia L. Winter. "Managing Social Norms for Persuasive Impact." *Social Influence* 1, no. 1 (2006): 3-15.

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." *Advances in Experimental Social Psychology* 24, no. 20 (1991): 1-243.

Costa, Dora L., and Matthew E. Kahn. Why has California's residential electricity consumption been so flat since the 1980s?: a microeconometric approach. No. w15978. National Bureau of Economic Research, 2010.

Darlington, Richard B. and Clifford E. Macker. "Displacement of Guilt-Produced Altruistic Behavior." *Journal of Personality and Social Psychology* 4, no. 4 (1966): 442.

Diamond, Peter. "Optimal tax treatment of private contributions for public goods with and without warm glow preferences." *Journal of Public Economics* 90, no. 4 (2006): 897-919.

Eckel, Catherine C., Enrique Fatas, and Rick Wilson. "Cooperation and status in organizations." Journal of Public Economic Theory 12.4 (2010): 737-762.

Etzioni, Amitai. "The case for a multiple-utility conception." *Economics and Philosophy* 2, no. 02 (1986): 159-184.

Falk, Armin and Andrea Ichino. "Clean Evidence of Peer Effects." *Journal of Labor Economics* 24, no. 1 (2006): 39-57.

Ferraro, Paul J. and Michael K. Price. "Using Nonpecuniary Strategies to Influence Behavior: Evidence from a Large-Scale Field Experiment." *Review of Economics and Statistics* 95, no. 1 (2013): 64-73.

Gibbons, Robert. *Game Theory for Applied Economists.* Princeton University Press, 1992.

Hallsworth, Michael, et al. *The behavioralist as tax collector: Using natural field experiments to enhance tax compliance*. No. w20007. National Bureau of Economic Research, 2014.

Ho, Benjamin, Greg Poe, John Taber, and Antonio Bento. "Culpability and Willingness to Pay to Reduce Negative Externalities: A Contingent Valuation and Experimental Economics Study." forthcoming Environment and Resource Economics (2014).

Johansson, Olof. "Optimal Pigovian Taxes under Altruism." *Land Economics* 73, no. 3 (1997).

Jordan, Jennifer, Elizabeth Mullen, and J. Keith Murnighan. "Striving for the Moral Self: The Effects of Recalling Past Moral Actions on Future Moral Behavior." *Personality and Social Psychology Bulletin* 37, no. 5 (2011): 701-713.

Kandel, Eugene, and Edward P. Lazear. "Peer pressure and partnerships." Journal of political Economy (1992): 801-817.

Khan, Uzma and Ravi Dhar. "Licensing Effect in Consumer Choice." *Journal of Marketing Research* (2006): 259-266.

Loewenstein, George F., Leigh Thompson, and Max H. Bazerman. "Social Utility and Decision Making in Interpersonal Contexts." *Journal of Personality and Social Psychology* 57, no. 3 (1989): 426.

Mas, Alexandre and Enrico Moretti. "Peers at work." *American Economic Review* 99, no. 1 (2009): 112-145.

Masclet, David, Charles Noussair, Steven Tucker and Marie-Claire Villeval. "Monetary and Nonmonetary Punishment in the Voluntary Contributions Mechanism." *American Economic Review* 93, no. 1 (2003): 366-380.

Merritt, Anna C., Daniel A. Effron, and Benoît Monin. "Moral Self-Licensing: When being Good Frees Us to be Bad." *Social and Personality Psychology Compass* 4, no. 5 (2010): 344-357.

Mirrlees, James A. "An exploration in the theory of optimum income taxation."*The review of economic studies* (1971): 175-208.

Monin, Benoît and Dale T. Miller. "Moral Credentials and the Expression of Prejudice." *Journal of Personality and Social Psychology* 81, no. 1 (2001): 33.

Ormiston, Michael B. "First and second degree transformations and comparative statics under uncertainty." *International Economic Review* (1992): 33-44.

Rotemberg, Julio J. "Human Relations in the Workplace. " *Journal of Political Economy* 102, no. 4 (1994): 684-717.

Sachdeva, Sonya, Rumen Iliev, and Douglas L. Medin. "Sinning Saints and Saintly Sinners the Paradox of Moral Self-Regulation." *Psychological Science* 20, no. 4 (2009): 523-528.

Sandmo, Agnar. "Optimal taxation in the presence of externalities." *The Swedish Journal of Economics* (1975): 86-98.

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* 18, no. 5 (2007): 429-434.

Sunstein, Cass R. "Social Norms and Social Roles." *Columbia Law Review* 96, no. 4 (1996): 903-968.

Tang, Zhongjun, Xiaohong Chen, and Jianghong Luo. "Determining Socio-Psychological Drivers for Rural Household Recycling Behavior in Developing Countries A Case Study from Wugan, Hunan, China." *Environment and Behavior* 43, no. 6 (2011): 848-877.

Thaler, Richard H. and Cass R. Sunstein. *Nudge: Improving Decisions about Health, Wealth, and Happiness* Yale University Press, 2008.

Warr, Peter G. "Pareto optimal redistribution and private charity." *Journal of Public Economics* 19, no. 1 (1982): 131-138.

APPENDICES

A. Proposition 1 and Proof

Proposition 1: Tragedy of the Commons. *Assuming (A2), (A3), and (A6), the individually rational choice of consumption of the anti-social good* g^* *and* G^* *is greater than the socially optimal level of consumption,* g^{**} *and* G^{**}*:*

$$G^* > G^{**} \text{ and } g^* > g^{**}.$$

Proof:

Each individual faces the maximization problem below:

$max_{g_i} u_i = V(g_i) - N(G) - c(g_i, \theta, \omega)$ (EQ1)

where u_i is the utility of the i^{th} individual and g_i is the consumption level chosen by him. The first order condition is:

$$\frac{dV(g_i)}{dg_i} - \frac{dN(G)}{dg_i} - \frac{dc(g_i, \theta, \omega)}{dg_i} = 0.$$

If $(g_1^*, g_2^*, \dots, g_I^*)$ is a Nash equilibrium, then g_i^* maximizes (EQ1) given that the others choose $(g_1^*, \dots, g_{i-1}^*, g_{i+1}^* \dots, g_I^*)$. Because of the symmetry of the equilibrium, we have $g_1^* = \cdots = g_i^* \dots = g_I^* = g^*$. The total units of goods consumed is $G^* = Ig_i^* = Ig^*$ and total social utility is $U_G^* = Iu_i^*$.

We have

$\frac{dV(g^*)}{dg} - \frac{dN(G^*)}{dg_i} - \frac{dc(g^*, \theta, \omega)}{dg} = 0.$ (EQ2)

However, the socially optimal level of consumption, $G^{**} = Ig^{**}$, is found by maximizing the total utility of the society, U_G, through solving the problem below:

$$\max_G U_G = IV\left(\frac{G}{I}\right) - IN(G) - Ic\left(\frac{G}{I}, \theta, \omega\right)$$

and the first order condition is:

$$I\frac{dV\left(\frac{G}{I}\right)}{dG} - I\frac{dN(G)}{dG} - I\frac{dc\left(\frac{G}{I}, \theta, \omega\right)}{dG} = 0.$$

So we have

$\frac{dV(g^{**})}{dg} - \frac{dN(G^{**})}{dg} - \frac{dc(g^{**},\theta,\omega)}{dg} = 0.$ (EQ3)

Now we compare (EQ2) and (EQ3). We have

$$\frac{dV(g^*)}{dg} - \frac{dN(G^*)}{dg} - \frac{dc(g^*,\theta,\omega)}{dg} = \frac{dV(g^{**})}{dg} - \frac{dN(G^{**})}{dg} - \frac{dc(g^{**},\theta,\omega)}{dg}.$$

Suppose $G^* \leq G^{**}$, $g^* \leq g^{**}$, then by (A3) and (A6), we have $\frac{dV(g^*)}{dg} \geq \frac{dV(g^{**})}{dg}$, and $\frac{dc(g^*,\theta,\omega)}{dg} \leq \frac{dc(g^{**},\theta,\omega)}{dg}$. Then we must $\frac{dN(G^*)}{dg} > \frac{dN(G^{**})}{dg}$. By (A2), $\frac{dN(G^*)}{dG} \leq \frac{dN(G^{**})}{dG}$, so we have $\frac{dN(G^*)}{dg} \leq \frac{dN(G^{**})}{I \cdot dg}$. This implies that $I \cdot \frac{dN(G^*)}{dg} \leq \frac{dN(G^{**})}{dg} < \frac{dN(G^*)}{dg}$. Note that since $\frac{dN(G^*)}{dg} > 0$, when $I \geq 1$, it cannot be that $I \cdot \frac{dN(G^*)}{dg} < \frac{dN(G^*)}{dg}$, presenting a contradiction. Therefore, $G^* > G^{**}, g^* > g^{**}$.

If every individual aims to maximize his or her own utility by choosing g privately instead of coordinating and choosing G collectively, the group will choose a higher level of total amount of the good and cause more negative externality. Then,

$$U_G(G^{**}) > U_G(G^*) = Iu_i = IV(g_i) - IN(G) - Ic(g_i,\theta,\omega)$$

$$\Rightarrow \frac{U_G(G^{**})}{I} > \frac{U_G(G^*)}{I} \Rightarrow u_i^{**} > u_i^*.$$

B. Proposition 2 and Proof

Proposition 2: Effect of Impure Altruism/Culpability on Consumption. *Assuming (A2), (A3), (A6) and (A7), consumption of the anti-social good is decreasing in disposition and situation:*

$$\frac{dg^*}{d\theta} < 0 \text{ and } \frac{dg^*}{d\omega} < 0.$$

In other words, an individual with high disposition or situational culpability will choose a smaller g

than people with low disposition or situational culpability.[4]

Proof:

According to the Implicit Function Theorem and by (A2), (A3), (A6) and (A7),

$$\frac{dg^*}{d\theta} = \frac{\frac{\partial^2 c}{\partial\theta\partial g}}{\frac{d^2V}{dg^2} - \frac{d^2N}{dg^2} - \frac{d^2c}{dg^2}} < 0,$$

$$\frac{dg^*}{d\omega} = \frac{\frac{\partial^2 c}{\partial\omega\partial g}}{\frac{d^2V}{dg^2} - \frac{d^2N}{dg^2} - \frac{d^2c}{dg^2}} < 0.$$

C. Proposition 3 and Proof

Proposition 3: Effect of Impure Altruism/Culpability on Utility*. Assuming (A2), (A3), (A5), and (A7), if the marginal disutility of increased altruism is greater (less) than the marginal utility from mitigating the externality, then altruism is welfare decreasing (increasing).*

(i) If $\left.\frac{\partial\hat{u}}{\partial\theta}\right|_{g^*} > \left.\frac{dN}{d\theta}\right|_{g^*}$, then $\left.\frac{\partial iu}{\partial\theta}\right|_{g^*} > 0$. If $\left.\frac{\partial\hat{u}}{\partial\omega}\right|_{g^*} > \left.\frac{dN}{d\omega}\right|_{g^*}$, then $\left.\frac{\partial iu}{\partial\omega}\right|_{g^*} > 0$.

(ii) If $\left.\frac{\partial\hat{u}}{\partial\theta}\right|_{g^*} < \left.\frac{dN}{d\theta}\right|_{g^*}$, then $\left.\frac{\partial iu}{\partial\theta}\right|_{g^*} < 0$. If $\left.\frac{\partial\hat{u}}{\partial\omega}\right|_{g^*} < \left.\frac{dN}{d\omega}\right|_{g^*}$, then $\left.\frac{\partial iu}{\partial\omega}\right|_{g^*} < 0$.

[4] We note that Proposition 2 needs the additional assumption (A7), that is, $\frac{\partial^2 c}{\partial\theta\partial g} > 0$ and $\frac{\partial^2 c}{\partial\omega\partial g} > 0$.[4] The assumption $\frac{\partial^2 c}{\partial\omega\partial g} > 0$ means that the impact of culpability is larger when the consumption of good is higher, that is, there is a positive complementarity between ω and g. A supermodularity condition on the cost function would yield the same results for non-differentiable cost functions. This complementarity can either work for or against the benefit of society. In the case that nudges have positive situational altruism ω, the complementarity amplifies the cleansing effect for people who are consuming large quantities of the good. On the other hand, nudges that decrease situational altruism ω work in a counterproductive way when moral licensing comes into play; the complementarity between ω and g amplifies this undesired effect as well.

Proof:

By Envelope Theorem, we have

$$\begin{aligned}\frac{\partial \hat{u}}{\partial \theta} = \frac{d\hat{u}^*}{d\theta} &= \frac{d}{d\theta}[V(g^*) - c(g^*, \theta, \omega)] \\ &= \frac{dV}{dg}\frac{dg^*}{d\theta} - \frac{\partial c}{\partial g}\frac{dg^*}{d\theta} - \frac{\partial c}{\partial \theta} \\ &= \left[\left(\frac{dV}{dg} - \frac{\partial c}{\partial g}\right)\frac{dg}{d\theta} - \frac{\partial c}{\partial \theta}\right]\Big|_{g^*} \\ &= \left[\left(\frac{dV}{dg} - \frac{dc}{dg}\right)\frac{dg}{d\theta} - \frac{\partial c}{\partial \theta}\right]\Big|_{g^*} \\ &= \left(-\frac{\partial c}{\partial \theta}\right)\Big|_{g^*} < 0\end{aligned}$$

as $\frac{\partial c}{\partial g} = \frac{dc}{dg}$, $\left(\frac{dV}{dg} - \frac{dc}{dg}\right)\Big|_{g^*} = 0$ (by first order condition), and $\frac{\partial c}{\partial \theta} > 0$ by our assumption (A5).

Similarly,

$$\frac{\partial \hat{u}}{\partial \omega} < 0.$$

The total social externality is $iN(G^*)$. We have

$$\frac{dN(G^*)}{d\theta} = \frac{dN}{dG}\frac{dG}{dg}\frac{dg}{d\theta}\Big|_{g^*} = \frac{dN}{dG}\frac{dg}{d\theta}\Big|_{g^*} < 0$$

as $\frac{dN}{dG} > 0$ by assumption (A2) and $\frac{dg^*}{d\theta} < 0$ by Proposition 2.

Similarly,

$$\frac{dN(G^*)}{d\omega} < 0.$$

To study the effect of impure altruism on total social utility at g^*, we examine

$$\frac{\partial iu}{\partial \theta}\Big|_{g^*} = i\frac{\partial u}{\partial \theta}\Big|_{g^*} = i\left(\frac{\partial \hat{u}}{\partial \theta} - \frac{\partial N}{\partial \theta}\right)\Big|_{g^*} = i\frac{\partial \hat{u}}{\partial \theta}\Big|_{g^*} - i\frac{dN}{d\theta}\Big|_{g^*} = i\frac{\partial \hat{u}}{\partial \theta}\Big|_{g^*} - i\frac{dN(G^*)}{d\theta}$$

and

$$\frac{\partial iu}{\partial \omega}\Big|_{g^*} = i\frac{\partial \hat{u}}{\partial \omega}\Big|_{g^*} - i\frac{dN}{d\omega}\Big|_{g^*}.$$

D. Proposition 4 and Proof

Proposition 4: Effect of Impure Altruism/Culpability on Social Welfare Loss. *Assuming (A2), (A3), (A5), and (A7), the change in Welfare Loss due to dispositional and situational altruism is*

$$\frac{\partial WL}{\partial \theta} = \underbrace{i\left[\frac{\partial c(g^*,\theta,\omega)}{\partial \theta} - \frac{\partial c(g^{**},\theta,\omega)}{\partial \theta}\right]}_{\text{Part I},>0} + \underbrace{i\frac{dN(G^*)}{d\theta}}_{\text{Part II},<0},$$

and

$$\frac{\partial WL}{\partial \omega} = \underbrace{i\left[\frac{\partial c(g^*,\theta,\omega)}{\partial \omega} - \frac{\partial c(g^{**},\theta,\omega)}{\partial \omega}\right]}_{\text{Part I},>0} + \underbrace{i\frac{dN(G^*)}{d\omega}}_{\text{Part II},<0}.$$

Social welfare loss represents the gap between first best and equilibrium consumption as defined in Proposition 1. Part I represents how the difference in the marginal psychic costs changes with θ or ω and this part is positive. Part II represents how the marginal social externality at g^* (or equivalently, G^*) changes with θ or ω and this part is negative. The net effect of impure altruism on social welfare loss depends on the relative magnitude of Part I and Part II.

Proof:

$$\begin{aligned}\frac{\partial WL}{\partial \omega} = \frac{\partial U^{**}}{\partial \omega} - \frac{\partial U^{*}}{\partial \omega} &= \frac{\partial U_G(G^{**})}{\partial \omega} - \frac{\partial}{\partial \omega}[i\hat{u}(g^*) - iN(G^*)] \\ &= \frac{\partial}{\partial \omega}\left[iV\left(\frac{G^{**}}{i}\right) - iN(G^{**}) - ic\left(\frac{G^{**}}{i},\theta,\omega\right)\right] - i\frac{\partial}{\partial \omega}[V(g^*) - c(g^*,\theta,\omega)] \\ &+ i\frac{\partial N(G^*)}{\partial \omega} = i\left[-\frac{\partial c(g^{**},\theta,\omega)}{\partial \omega}\right] - i\left[-\frac{\partial c(g^*,\theta,\omega)}{\partial \omega}\right] + i\frac{dN(G^*)}{d\omega} \\ &= \underbrace{i\left[\frac{\partial c(g^*,\theta,\omega)}{\partial \omega} - \frac{\partial c(g^{**},\theta,\omega)}{\partial \omega}\right]}_{\text{Part I},>0} + \underbrace{i\frac{dN(G^*)}{d\omega}}_{\text{Part II},<0}\end{aligned}$$

by assumption (A2), (A7) , $g^* > g^{**}$ (Proposition 1), and $\frac{dN(G^*)}{d\omega} < 0$ (C. Proof of Proposition 3). The sign of $\frac{dWL}{d\omega}$ can be either positive or negative, depending on the relative magnitude of Part I and Part II.

Similarly,

$$\frac{\partial WL}{\partial \theta} = \underbrace{i\left[\frac{\partial c(g^*,\theta,\omega)}{\partial \theta} - \frac{\partial c(g^{**},\theta,\omega)}{\partial \theta}\right]}_{\text{Part I},>0} + \underbrace{i\frac{dN(G^*)}{d\theta}}_{\text{Part II},<0}.$$

E. Proof of Proposition 5

Taking the cross-derivative, we have

$$\frac{d}{d\omega}\left(\frac{dg^*}{d\theta}\right) = \frac{d}{d\omega}\left(\frac{\frac{\partial^2 c}{\partial\theta\partial g}}{\frac{d^2V}{dg^2} - \frac{d^2N}{dg^2} - \frac{d^2c}{dg^2}}\right)\Bigg|_{g^*}$$
$$= \frac{\left(\frac{d^2V}{dg^2} - \frac{d^2N}{dg^2} - \frac{d^2c}{dg^2}\right)\cdot\frac{d}{d\omega}\left(\frac{\partial^2 c}{\partial\theta\partial g}\right) + \frac{\partial^2 c}{\partial\theta\partial g}\cdot\frac{d}{d\omega}\left(\frac{\partial^2 c}{\partial g^2}\right)}{\left(\frac{d^2V}{dg^2} - \frac{d^2N}{dg^2} - \frac{d^2c}{dg^2}\right)^2}\Bigg|_{g^*}.$$

Based on (A2), (A3), (A6), and (A8),

$$\frac{d}{d\omega}\left(\frac{dg^*}{d\theta}\right) = \frac{d}{d\omega}\left(\frac{\frac{\partial^2 c}{\partial\theta\partial g}}{\frac{d^2V}{dg^2} - \frac{d^2N}{dg^2}}\right)\Bigg|_{g^*}$$
$$= \frac{\left(\frac{d^2V}{dg^2} - \frac{d^2N}{dg^2}\right)\cdot\frac{d}{d\omega}\left(\frac{\partial^2 c}{\partial\theta\partial g}\right) - \frac{\partial^2 c}{\partial\theta\partial g}\cdot\frac{d}{d\omega}\left(\frac{d^2V}{dg^2} - \frac{d^2N}{dg^2}\right)}{\left(\frac{d^2V}{dg^2} - \frac{d^2N}{dg^2}\right)^2}\Bigg|_{g^*}$$
$$= \frac{\left(\frac{d^2V}{dg^2} - \frac{d^2N}{dg^2}\right)\cdot\frac{d}{d\omega}\left(\frac{\partial^2 c}{\partial\theta\partial g}\right)}{\left(\frac{d^2V}{dg^2} - \frac{d^2N}{dg^2}\right)^2}\Bigg|_{g^*}.$$

Since $\frac{d^2V}{dg^2} - \frac{d^2N}{dg^2} < 0$,

$$\text{sign of } \frac{d}{d\omega}\left(\frac{dg^*}{d\theta}\right) \text{ is opposite to that of } \frac{d}{d\omega}\left(\frac{\partial^2 c}{\partial\theta\partial g}\right)\Big|_{g^*}.$$

By (A8), we have $c(g,\theta,\omega) = g\cdot c'(\theta,\omega)$. Thus, $\frac{\partial^2 c}{\partial\theta\partial g} = \frac{\partial^2[gc'(\theta,\omega)]}{\partial\theta\partial g} = \frac{\partial}{\partial\theta}[c'(\theta,\omega)] = \frac{\partial c'}{\partial\theta}$. Then

$$\frac{d}{d\omega}\left(\frac{\partial^2 c}{\partial\theta\partial g}\right)\Big|_{g^*} = \frac{\partial}{\partial\omega}\left(\frac{\partial^2 c}{\partial\theta\partial g}\right) + \frac{\partial}{\partial g}\left(\frac{\partial^2 c}{\partial\theta\partial g}\right)\frac{\partial g}{\partial\omega}\Big|_{g^*}$$

$$= \frac{\partial^2 c'}{\partial\omega\partial\theta} + \frac{\partial^2 c'}{\partial g\partial\theta}\frac{\partial g}{\partial\omega}\bigg|_{g^*} = \frac{\partial^2 c'}{\partial\omega\partial\theta}\bigg|_{g^*}$$

as $\frac{\partial^2 c'}{\partial g\partial\theta} = 0$. We also have $\frac{\partial^2 c}{\partial\omega\partial\theta} = \frac{\partial^2\left(gc'(\theta,\omega)\right)}{\partial\omega\partial\theta} = g\frac{\partial^2 c\prime}{\partial\omega\partial\theta'}$, so we have

$$\text{sign of } \frac{\partial^2 c}{\partial\omega\partial\theta} \text{ is the same as that of } \frac{\partial^2 c'}{\partial\omega\partial\theta}$$

and thus,

$$\text{sign of } \frac{d}{d\omega}\left(\frac{\partial^2 c}{\partial\theta\partial g}\right)\bigg|_{g^*} \text{ is the same as that of } \frac{\partial^2 c}{\partial\omega\partial\theta}.$$

Therefore, we have

$$\text{sign of } \frac{d}{d\omega}\left(\frac{dg^*}{d\theta}\right) \text{ is opposite to that of } \frac{\partial^2 c}{\partial\omega\partial\theta}.$$

Similarly,

$$\text{sign of } \frac{d}{d\theta}\left(\frac{dg^*}{d\omega}\right) \text{ is opposite to that of } \frac{\partial^2 c}{\partial\theta\partial\omega}.$$

F. Proof of Proposition 6

In the case of endogenous ω, the utility maximization problem is

$$\max_{g_i} u_i = V(g_i) - N(G) - c[g_i, \omega(g_i, \hat{g_i}), \theta].$$

$$F.O.C. \Rightarrow \frac{du_i}{dg_i} = \frac{dV}{dg} - \frac{dN}{dg} - \frac{dc}{dg} = \frac{dV}{dg} - \frac{dN}{dg} - \frac{\partial c}{\partial\omega}\frac{\partial\omega}{\partial g} - \frac{\partial c}{\partial g} = 0$$

In addition to (A2), (A3), (A5), and (A7), assume that $\frac{\partial w}{\partial g} > 0$, $\frac{\partial^2 w}{\partial g^2} > 0$, and $\frac{\partial w}{\partial\hat{g}} < 0$, and that the psychic cost function is continuous so $\frac{\partial^2 c}{\partial g\partial\omega} = \frac{\partial^2 c}{\partial\omega\partial g} > 0$, $\frac{\partial^2 c}{\partial g\partial\theta} = \frac{\partial^2 c}{\partial\theta\partial g} > 0$.

According to the Implicit Function Theorem, we have

$$\frac{dg^*}{d\hat{g}} = \frac{\frac{\partial}{\partial\hat{g}}\left(\frac{\partial c}{\partial\omega}\frac{\partial\omega}{\partial g}\right) + \frac{\partial}{\partial\hat{g}}\left(\frac{\partial c}{\partial g}\right)}{\frac{d^2V}{dg^2} - \frac{\partial}{\partial g}\left(\frac{dN}{dG}\right) - \frac{\partial}{\partial g}\left(\frac{\partial c}{\partial\omega}\frac{\partial\omega}{\partial g}\right) - \frac{\partial^2 c}{\partial g^2}} > 0.$$

Since $\frac{d^2V}{dg^2} < 0$, $\frac{\partial}{\partial g}\left(\frac{dN}{dG}\right) \geq 0$, $\frac{\partial^2 c}{\partial g^2} \geq 0$, the sign of the denominator is determined by the sign of $\frac{\partial}{\partial g}\left(\frac{\partial c}{\partial \omega}\frac{\partial \omega}{\partial g}\right) = \frac{\partial \omega}{\partial g}\frac{\partial^2 c}{\partial g \partial \omega} + \frac{\partial c}{\partial \omega}\frac{\partial^2 \omega}{\partial g^2}$. Since $\frac{\partial c}{\partial \omega} > 0, \frac{\partial^2 \omega}{\partial g^2} > 0$, $\frac{\partial w}{\partial g} > 0$, and $\frac{\partial^2 c}{\partial g \partial \omega} > 0$, we have $\frac{\partial}{\partial g}\left(\frac{\partial c}{\partial \omega}\frac{\partial \omega}{\partial g}\right) > 0$. Therefore, the sign of the denominator is negative. The sign of the numerator is also negative since $\frac{\partial}{\partial \hat{g}}\left(\frac{\partial c}{\partial \omega}\frac{\partial \omega}{\partial g}\right) + \frac{\partial}{\partial \hat{g}}\left(\frac{\partial c}{\partial g}\right) = 2\frac{\partial}{\partial \hat{g}}\left(\frac{\partial c}{\partial g}\right) = 2\frac{\partial}{\partial \omega}\left(\frac{\partial c}{\partial g}\right) \cdot \frac{\partial \omega}{\partial \hat{g}} < 0$. Thus,

$$\frac{dg^*}{d\hat{g}} > 0.$$

According to the Implicit Function Theorem, we have

$$\frac{dg^*}{d\theta} = \frac{\frac{\partial}{\partial \theta}\left(\frac{\partial c}{\partial \omega}\frac{\partial \omega}{\partial g}\right) + \frac{\partial}{\partial \theta}\left(\frac{\partial c}{\partial g}\right)}{\frac{d^2V}{dg^2} - \frac{\partial}{\partial g}\left(\frac{dN}{dG}\right) - \frac{\partial}{\partial g}\left(\frac{\partial c}{\partial \omega}\frac{\partial \omega}{\partial g}\right) - \frac{\partial^2 c}{\partial g^2}}.$$

As before, the sign of the denominator is negative. Now we discuss the sign of the numerator. Since $\frac{\partial}{\partial \theta}\left(\frac{\partial c}{\partial \omega}\frac{\partial \omega}{\partial g}\right) + \frac{\partial}{\partial \theta}\left(\frac{\partial c}{\partial g}\right) = 2\frac{\partial}{\partial \theta}\left(\frac{\partial c}{\partial g}\right) > 0$, we have

$$\frac{dg^*}{d\theta} < 0.$$

G. Proof of Proposition 7

We first justify the existence claim. Let $\widehat{g_0}$ and θ be given. Note that by Proposition 6, $\frac{dg^*}{d\hat{g}} > 0$. We assume that we can find $\hat{g}$ small enough such that $I \cdot g^*(\hat{g}', \theta) < G^{**}(\widehat{g_0}, \theta)$. Since $I \cdot g^*(\widehat{g_0}, \theta) > G^{**}(\widehat{g_0}, \theta)$, under some technical conditions we have that there exists $\hat{g}_{opt}$ between $\hat{g}'$ and $\widehat{g_0}$ such that $\hat{g} = \hat{g}_{opt}$ solves the following equation:

$$I \cdot g^*(\hat{g}, \theta) = G^{**}(\widehat{g_0}, \theta),$$

or equivalently,

$$g^*(\hat{g}, \theta) - \frac{1}{I}G^{**}(\widehat{g_0}, \theta) = 0.$$

To show for the rest of the proposition, note that by Implicit Function Theorem, we have

$$\frac{d\hat{g}_{opt}}{d\theta} = -\frac{\frac{dg^*(\hat{g},\theta)}{d\theta} - \frac{1}{I}\frac{dG^{**}(\widehat{g_0},\theta)}{d\theta}}{\frac{dg^*(\hat{g},\theta)}{d\hat{g}}} = \frac{\frac{dg^{**}}{d\theta} - \frac{dg^*}{d\theta}}{\frac{dg^*}{d\hat{g}}}.$$

By Proposition 6, $\frac{dg^*}{d\hat{g}} > 0$. So we have

$$\text{sign of } \frac{d\hat{g}_{opt}}{d\theta} \text{ is the same as that of } \frac{dg^{**}}{d\theta} - \frac{dg^*}{d\theta}.$$

Since $\frac{dg^*}{d\theta} = \left.\frac{\frac{\partial^2 c}{\partial\theta\partial g}}{\frac{d^2V}{dg^2} - \frac{d^2N}{dg^2} - \frac{d^2c}{dg^2}}\right|_{g^*}$ and $\frac{dg^{**}}{d\theta} = \left.\frac{\frac{\partial^2 c}{\partial\theta\partial g}}{\frac{d^2V}{dg^2} - \frac{d^2N}{dg^2} - \frac{d^2c}{dg^2}}\right|_{g^{**}}$, then we have

$$\frac{d\hat{g}_{opt}}{d\theta} \neq 0 \Leftrightarrow \frac{\partial}{\partial g}\left(\frac{\partial^2 c}{\partial\theta\partial g}\right) \neq 0.$$

Therefore, as long as we assume $\frac{\partial}{\partial g}\left(\frac{\partial^2 c}{\partial\theta\partial g}\right) \neq 0$, then the optimal message would depend on the dispositional altruism θ.

H. Lemma 8.1: Existence of Interior Solution T^* for First Order Stochastic Dominance

In the case of the more realistic modifications of the maximization problem (a budget constraint and a real price), we have

$$\frac{d^2u}{dT^2} = \frac{d^2g^*}{dT^2}\cdot T + \frac{dg^*}{dT} = \frac{dg^*}{dT} < 0.$$

So $\Pi(T,\alpha)$ is strictly concave in T since $\Pi = \frac{1}{I}\sum_1^I u_i$. Therefore, the existence of an interior solution T^* that solves $E[u_T(T,\alpha)] = 0$ and $E[u_{TT}(T,\alpha)] < 0$ can be assumed and first order stochastic dominance works.

I. Proof of Proposition 8

Note that T^* increases (decreases) for all FSD transformations of the random variable α if

$u_{T\alpha}(T,\alpha) \geq 0$ $(u_{T\alpha}(T,\alpha) \leq 0)$ everywhere. We want to show that $u_{T\alpha}(T,\alpha) = \frac{d^2u}{d\alpha dT} \geq 0$ is equivalent to $\frac{d^2g^*}{d\alpha dT} \geq 0$.

In the case of the more realistic modifications of the maximization problem, we have

$$\frac{d^2u}{d\alpha dT} = \frac{d}{d\alpha}\left(\frac{dV}{dT} - \frac{dN}{dT} - \frac{dc}{dT} - x\frac{dg}{dT}\right) = \frac{d}{d\alpha}\left(\frac{dV}{dg^*}\frac{dg^*}{dT} - \frac{dN}{dg^*}\frac{dg^*}{dT} - \frac{dc}{dg^*}\frac{dg^*}{dT} - x\frac{dg}{dT}\right)$$
$$= \frac{d^2g^*}{d\alpha dT}\left(\frac{dV}{dg^*} - \frac{dN}{dg^*} - \frac{dc}{dg^*}\right) + \frac{dg^*}{dT}\frac{d}{d\alpha}\left(\frac{dV}{dg^*} - \frac{dN}{dg^*} - \frac{dc}{dg^*}\right) - x\frac{d}{d\alpha}\left(\frac{dg}{dT}\right).$$

Since g^* is the solution to individual utility maximization problem, that is, g^* is the solution to $\frac{dV}{dg} - \frac{dN}{dg} - \frac{dc}{dg} - T - x = 0$, we have

$$\frac{dV}{dg^*} - \frac{dN}{dg^*} - \frac{dc}{dg^*} = T + x.$$

Then

$$\frac{d^2u}{d\alpha dT} = \frac{d^2g^*}{d\alpha dT}(T + x) + \frac{dg^*}{dT}\frac{dT}{d\alpha} - \frac{d^2g^*}{d\alpha dT}x = \frac{d^2g^*}{d\alpha dT}T.$$

With $T \geq 0$, the sign of $\frac{d^2u}{d\alpha dT}$ is the same as that of $\frac{d^2g^*}{d\alpha dT}$.

J. Proof of Proposition 9

Refer to I. Proof of Proposition 8 to see that the sign of $\frac{d^2u}{d\alpha dT}$ is the same as that of $\frac{d^2g^*}{d\alpha dT}$.

K. Proof of Proposition 10

Under the binary assumption for α, the government faces the following utility problem:

$$\max_{T\geq 0} E[\Pi(T)] = \sum_{1}^{I} E[u_i] = \sum_{1}^{I} [pu_1 + (1-p)u_2], \text{s.t. } 0 \leq T \leq x.$$

This is equivalent to

$$\max_{T} E[\pi] = pu_1 + (1-p)u_2, \text{s.t. } 0 \leq T \leq x,$$

where

$$u_1 = u(\alpha^H, T) = V(g^*) - N(G^*) - c(\alpha^H, g^*) - xg^*,$$

$$u_2 = u(\alpha^L, T) = V(g^*) - N(G^*) - c(\alpha^L, g^*) - xg^*,$$

and $g^* = h(\alpha, T)$ is the solution to the following individual utility maximization problem based on the level of taxation:

$$\max_g s = V(g) - N(G) - c(\alpha, g) - Tg - xg,$$

subject to

$$g(x + T) \leq b, g \geq 0.$$

Write $\mathcal{L}(g, \lambda_1, \lambda_2) = s(g) - \lambda_1[g(x + T) - b] - \lambda_2(-g)$. In the case of inequality constraints, we solve the Kuhn-Tucker conditions in additions to the inequalities $g(x + T) \leq b$ and $g \geq 0$. The Kuhn-Tucker conditions for maximum consist of the first order condition

$$\frac{d\mathcal{L}}{dg} = 0 \Rightarrow \frac{ds}{dg} - \lambda_1(x + T) + \lambda_2 = 0 \Rightarrow \frac{ds}{dg} = \lambda_1(x + T) - \lambda_2$$

and the complementary slackness conditions given by

$\lambda_1 \geq 0$ and $\lambda_1 = 0$ whenever $g(x + T) = b$,

$\lambda_2 \geq 0$ and $\lambda_2 = 0$ whenever $g = 0$.

Similarly write $\mathcal{L}'(T, \delta_1, \delta_2) = pu_1 + (1 - p)u_2 - \delta_1(T - x) - \delta_2(-T)$, and solve for Kuhn-Tucker conditions in additions to the inequality constraints:

$$F.O.C. \Rightarrow \frac{dE[\pi]}{dT} = p\frac{du_1}{dT} + (1 - p)\frac{du_2}{dT} - \delta_1 + \delta_2 = F(T) = 0,$$

and

$\delta_1 \geq 0$ and $\delta_1 = 0$ whenever $T = x$

$\delta_2 \geq 0$ and $\delta_2 = 0$ whenever $T = 0$.

According to the Implicit Function Theorem,

$$\frac{dT^*}{dp} = -\frac{\partial F/\partial p}{\partial F/\partial T} = -\frac{\frac{du_1}{dT} - \frac{du_2}{dT}}{p\frac{d^2u_1}{dT^2} + (1-p)\frac{d^2u_2}{dT^2}} = (*).$$

The denominator of $(*)$ has the same sign as $\frac{d^2u}{dT^2}$ and the sign of the numerator of $(*)$ is determined by $\frac{d^2u}{d\alpha dT}$. If $\frac{d^2u}{d\alpha dT} > 0$, then $\frac{du_1}{dT} > \frac{du_2}{dT}$, and the numerator of $(*)$ is positive, vice versa. So

$$\text{sign of } \left(\frac{dT^*}{dp}\right) \text{ is opposite to that of } \frac{\left(\frac{d^2u}{d\alpha dT}\right)}{\left(\frac{d^2u}{dT^2}\right)}.$$

Now $\frac{d^2u}{dT^2} = \frac{d^2g^*}{dT^2}\cdot T + \frac{dg^*}{dT} = \frac{dg^*}{dT} < 0$, so the sign of $\frac{dT^*}{dp}$ is the same as the sign of $\frac{d^2u}{d\theta dT}$, which is the same as the sign of $\frac{d^2g^*}{d\theta dT}$ as we have shown in the proof of Proposition 8.

L. Proof of Proposition 11

The government maximizes the expected value of Π:

$$\max_T E[\Pi] = \sum_1^I E[V(g_i^*) - N(G^*) - c(\theta, \omega, g_i^*)],$$

Then by the first order condition, we have the following true at $T = T^*$:

$$\frac{dE[\Pi]}{dT} = I \cdot E[V_h h_T - N_h h_T - c_h h_T] = 0.$$

By Implicit Function Theorem,

$$\frac{dT^*}{d\theta} = -\frac{E[h_T \cdot (V_{h\theta} - N_{h\theta} - c_{h\theta}) + h_{T\theta} \cdot (V_h - N_h - c_h)]}{E[h_T \cdot (V_{hT} - N_{hT} - c_{hT}) + h_{TT} \cdot (V_h - N_h - c_h)]}$$

Since $\frac{dg^*}{dT} = \frac{1}{\frac{d^2V}{dg^2} - \frac{d^2N}{dg^2} - \frac{d^2c}{dg^2}} < 0$, $h_T < 0$ and $h_{TT} = 0$.

Assuming $I \geq 2$, note that the first order condition of the individual utility maximization problem

gives

$$V_h - N_h - c_h = T \geq 0,$$

and taking derivative respect to T gives

$$V_{hT} - N_{hT} - c_{hT} = 1, V_{h\theta} - N_{h\theta} - c_{h\theta} = 0.$$

Also $h_{T\theta} = \left(\frac{d^2V}{dg^2} - \frac{d^2N}{dg^2} - \frac{d^2c}{dg^2}\right)^{-2} \cdot \frac{d}{d\theta}\left(\frac{d^2c}{dg^2}\right) = \left(\frac{d^2V}{dg^2} - \frac{d^2N}{dg^2} - \frac{d^2c}{dg^2}\right)^{-2} \cdot \frac{d\tilde{c}}{d\theta} \cdot \frac{d^2\hat{c}}{dg^2} \geq 0$. Thus, we have

$$\frac{dT^*}{d\theta} = -\frac{E[h_{T\theta} \cdot T]}{E[h_T]} \geq 0.$$

A similar proof shows that $\frac{dT^*}{d\omega} \geq 0$. The equalities hold if and only if $\frac{\partial^2\hat{c}}{\partial g^2} = 0$.

M. Proof of Proposition 12

$$\frac{d^2T^*}{d\omega d\theta} = -\frac{E[h_{T\theta\omega} \cdot T]E[h_T] - E[h_{T\omega}]E[h_{T\theta} \cdot T]}{(E[h_T])^2}$$
$$= \frac{T \cdot \left[E[h_{T\omega}]E[h_{T\theta}] - E[h_{T\theta\omega}]E[h_T]\right]}{(E[h_T])^2}$$

So we have the sign of $\frac{d^2T^*}{d\omega d\theta}$ is the same as the sign of:

$$T \cdot \left[\underbrace{E[h_{T\omega}]E[h_{T\theta}]}_{\geq 0} - E[h_{T\theta\omega}]E\underbrace{[h_T]}_{<0}\right].$$

Note $h_{T\theta} \geq 0$. Similarly, we can derive $h_{T\omega} \geq 0$. Also, $h_T < 0$. Thus, the sign of $\frac{d^2T^*}{d\omega d\theta}$ depends on the sign of $E[h_{T\theta\omega}]$. The sign of $E[h_{T\theta\omega}]$ is in turn determined by the sign of $\frac{d^2\tilde{c}}{d\omega d\theta}$:

$$h_{T\theta\omega} = \frac{d}{d\omega}\left(\frac{d^2g^*}{d\theta dT}\right) = \frac{d}{d\omega}\left[\frac{\frac{d}{d\theta}\left(\frac{d^2c}{dg^2}\right)}{\left(\frac{d^2V}{dg^2} - \frac{d^2N}{dg^2} - \frac{d^2c}{dg^2}\right)^2}\right]$$

$$= \frac{\left(\frac{d^2V}{dg^2} - \frac{d^2N}{dg^2} - \frac{d^2c}{dg^2}\right)^2\left[\frac{d^2\left(\frac{d^2c}{dg^2}\right)}{d\omega d\theta}\right] - \left[\frac{d}{d\theta}\left(\frac{d^2c}{dg^2}\right)\right]\cdot 2\left(\frac{d^2V}{dg^2} - \frac{d^2N}{dg^2} - \frac{d^2c}{dg^2}\right)\cdot\left[-\frac{d}{d\omega}\left(\frac{d^2c}{dg^2}\right)\right]}{\left(\frac{d^2V}{dg^2} - \frac{d^2N}{dg^2} - \frac{d^2c}{dg^2}\right)^4}$$

$$= \frac{\left(\frac{d^2V}{dg^2} - \frac{d^2N}{dg^2} - \frac{d^2c}{dg^2}\right)^2\left(\frac{d^2\tilde{c}}{d\omega d\theta}\cdot\frac{d^2\hat{c}}{dg^2}\right) + 2\cdot\left[\frac{d}{d\theta}\left(\frac{d^2c}{dg^2}\right)\right]\cdot\left[\frac{d}{d\omega}\left(\frac{d^2c}{dg^2}\right)\right]\cdot\left(\frac{d^2V}{dg^2} - \frac{d^2N}{dg^2} - \frac{d^2c}{dg^2}\right)}{\left(\frac{d^2V}{dg^2} - \frac{d^2N}{dg^2} - \frac{d^2c}{dg^2}\right)^4}$$

Since $\frac{d^2\hat{c}}{dg^2} \geq 0, \frac{d}{d\theta}\left(\frac{d^2c}{dg^2}\right) = \frac{d\tilde{c}}{d\theta}\cdot\frac{d^2\hat{c}}{dg^2} \geq 0$, $\frac{d}{d\omega}\left(\frac{d^2c}{dg^2}\right) = \frac{d\tilde{c}}{d\omega}\cdot\frac{d^2\hat{c}}{dg^2} \geq 0$, the sign of $h_{T\theta\omega}$ is opposite to that of:

$$\underbrace{\left(\frac{d^2\hat{c}}{dg^2}\right)\left(\frac{d^2V}{dg^2} - \frac{d^2N}{dg^2} - \frac{d^2c}{dg^2}\right)}_{<0}\left(\frac{d^2\tilde{c}}{d\omega d\theta}\right) + 2\cdot\underbrace{\left[\frac{d}{d\theta}\left(\frac{d^2c}{dg^2}\right)\right]\cdot\left[\frac{d}{d\omega}\left(\frac{d^2c}{dg^2}\right)\right]}_{\geq 0}$$

Thus, the sign of $h_{T\theta\omega}$ is determined by $\frac{d^2\tilde{c}}{d\omega d\theta}$. If $\frac{d^2\tilde{c}}{d\omega d\theta} < 0$, then $h_{T\theta\omega} > 0$, and as a result, $\frac{d^2T^*}{d\omega d\theta} > 0$. If $\frac{d^2\tilde{c}}{d\omega d\theta} > 0$, then the sign of $h_{T\theta\omega}$ is ambiguous. For $\frac{d^2\tilde{c}}{d\omega d\theta}$ large enough, we will have $h_{T\theta\omega}$ really negative that can possibly make $\frac{d^2T^*}{d\omega d\theta} < 0$.