

Ek, Claes

**Working Paper**

## Prosocial Behavior and Policy Spillovers: A Multi-Activity Approach

Working Paper, No. 2015:26

**Provided in Cooperation with:**

Department of Economics, School of Economics and Management, Lund University

*Suggested Citation:* Ek, Claes (2016) : Prosocial Behavior and Policy Spillovers: A Multi-Activity Approach, Working Paper, No. 2015:26, Lund University, School of Economics and Management, Department of Economics, Lund

This Version is available at:

<https://hdl.handle.net/10419/260164>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

Working Paper 2015:26

Department of Economics  
School of Economics and Management

# Prosocial Behavior and Policy Spillovers: A Multi-Activity Approach

Claes Ek

August 2015

Revised: September 2017



**LUND**  
UNIVERSITY

# Prosocial Behavior and Policy Spillovers: A Multi-Activity Approach<sup>☆</sup>

Claes Ek<sup>a,1</sup>

<sup>a</sup>*Centre for Collective Action Research, Department of Economics, University of Gothenburg, P.O. Box 640, SE-40530 Gothenburg, Sweden*

---

## Abstract

Observing that people who wish to engage in prosocial behavior are often presented with more than one means to the same end, we develop a model in which agents may contribute to a single public good through a range of different activities. We use this model to make two points. First, noting that effort on one activity has been argued to reduce (moral licensing) as well as increase (moral consistency) effort on other activities, we derive sufficient conditions under which policy to facilitate one activity partially crowds out effort on other activities. Second, we use an example to argue that a given single-activity model may be extended to multiple activities in several alternative ways, and that not all generalizations need reproduce the ideas and results of the original model; in general, careful thinking is needed to determine which multi-activity model is appropriate.

*Keywords:* public goods, prosocial behavior, moral licensing, self-image

*JEL classification:* D03, D11, H41

---

## 1 Introduction

Introspection, as well as research on ‘mental accounting’ (Henderson and Peterson, 1992; Thaler, 1999) suggests that people view commonly undertaken prosocial acts as belonging to broad categories. Recycling household waste, buying organic products, and donating to environmental NGOs are all examples of pro-environmental acts; voting, running for local office, or demonstrating are ways to affirm democratic values; and a multitude of charities (for which one may volunteer as well as donate) imply that there are many ways of aiding the less fortunate. In each domain, there is a range of alternatives. By contrast, the vast majority of economic studies on prosocial behavior assume that people may contribute through only

---

<sup>☆</sup>This version: September 2017. The first draft of this paper was completed while at the Lund University School of Economics and Management.

*Email address:* [claes.ek@economics.gu.se](mailto:claes.ek@economics.gu.se) (Claes Ek)

<sup>1</sup>Tel.: +46 31 786 52 49.

a single activity. Beyond explaining *why* people contribute, analyzing *how* people contribute requires a high-resolution model that includes more than one contributing activity. This paper develops such a model.

We focus on the situation where some policy intervention targets one of several prosocial activities. For example, many countries have introduced directed schemes (whether through facilitation, public information or economic incentives) to encourage specific environmental behaviors like recycling, leaving the car at home, or reducing energy use. To better understand such situations, multi-activity models are useful in at least two respects.

First, they allow for analysis of spillovers, that is, policy impacts on activities other than the one targeted. Such effects can be expected to arise whenever the efforts that people spend on a given activity reflect not merely its own characteristics, but those of similarly categorized acts as well. Spillovers can be seen as a manifestation of standard notions of (gross) substitutability: as policies alter the relative material incentives to contribute through the targeted activity, demand for other activities shifts up or down.

There are at present a few field studies on spillovers, e.g. on electricity use from a water-conservation campaign (Jacobsen et al., 2012; Lacetera et al., 2012; Tiefenbeck et al., 2013; Carlsson et al., 2016), as well as a small but rapidly growing literature on ‘expenditure substitution’ across charities in response to donation matching or other shocks (Reinstein, 2011; Cairns and Slonim, 2011; Null, 2011; Meer, 2016; Filiz-Ozbay and Uler, 2016; Ek, 2017). All of these studies estimate whether spillovers (or cross-price effects) are positive or negative, i.e. whether policies produce crowding in or crowding out.<sup>2</sup> While the bulk of the evidence points toward crowding out, exceptions are not uncommon.

In psychology, this fact is reflected in competing theoretical accounts (Merritt et al., 2010; Dolan and Galizzi, 2015). On the one hand, some theories stress consistency across separate decisions (Festinger, 1957; Bem, 1967), suggesting that one prosocial act can spur another. If activity-specific policy does increase effort on the targeted activity, we should then expect crowding in. On the other hand, a literature on ‘moral balancing’ (Cain et al., 2005; Sachdeva et al., 2009; Blanken et al., 2015) argues that people who have just behaved prosocially may feel they have a ‘moral license’ to subsequently relax their moral standards, while people who have just behaved badly may feel obliged to engage in subsequent ‘moral cleansing’ (Ploner

---

<sup>2</sup>Here, crowding out should not be confused with other uses of the term. First, government provision of public goods may crowd out private contributions (Andreoni, 1990; Nyborg and Rege, 2003), and spending by one charity may crowd out spending by other charities (Ribar and Wilhelm, 2002). Second, material incentives for prosocial behavior may crowd out intrinsic motivation, perversely reducing contributions (Frey and Jegen, 2001; Bénabou and Tirole, 2006; Bowles, 2008). To this we add a third layer of complexity, namely that interventions that target a particular activity may crowd out (or in) effort on *other* activities.

and Regner, 2013). Again, if the policy is effective, efforts on other activities should decrease, leading to crowding out.

Importantly, experimental and empirical methods predominate among these prior studies, with formal analysis arguably lagging behind. Our model represents a first step toward filling that gap. Building on the model of Brekke et al. (2003), we derive sufficient conditions under which activity-specific policy crowds *out* effort on all other activities. Thus, we establish negative spillovers as the benchmark in large-scale public-goods settings.<sup>3</sup> This result, however, is limited in two ways. First, we also explore a model variant where agents endogenously update the ‘ideal’ contribution with which they compare their own contribution. In this case, the sign of spillovers is found to be ambiguous. Second, we solve the model only for the case of two activities. Because of this, the activity which exhibits spillovers should be interpreted as an aggregate of all ‘other activities’ that are not directly impacted by an intervention, and negative spillovers need not apply to all (real-world) activities within that aggregate.

A second reason why multi-activity models are useful is that they allow us to re-evaluate previous results from single-activity models. Unlike our multi-activity framework, single-activity models obviously do not include activities that are only indirectly targeted by a policy, and depending on the manner in which this feature is added, results may alter *even with respect to the targeted activity*. This general point should not be too counter-intuitive: for example, compare the impact of a price increase in a single-good utility problem with a situation where we also add a close substitute not affected by the price increase.

As our model draws on that of Brekke et al. (2003), we illustrate this point by revisiting a variant of that model, namely Bruvold and Nyborg (2004). Our model, which at first glance seems a straightforward extension of Bruvold and Nyborg (2004) to multiple activities, does introduce new issues and accommodate results that are sometimes at odds with the ideas and conclusions of the original paper. However, several plausible ways of extending Bruvold and Nyborg (2004) exist, raising questions concerning which multi-activity model within the overall Brekke et al. (2003) framework is appropriate. We attempt to shed light on the issues involved by outlining some key alternatives.

Like Brekke et al. (2003), we focus on moral rather than social norms, and our model can be interpreted in terms of ‘duty-orientation’: individuals derive utility from maintaining a

---

<sup>3</sup>This paper also relates to a branch of the charitable-giving literature which examines the choice between different *modes* of contributing, i.e. between volunteering and money donations (Andreoni et al., 1996; Feldman, 2010; Lilley and Slonim, 2014; Brown et al., 2016). Since our model expresses all contributing activities in terms of time spent, it is not directly comparable to the theoretical analysis contained within these papers; yet the overall focus is similar. For example, volunteering and monetary donations are typically found to be net (if not necessarily gross) substitutes.

positive self-image of themselves as responsible citizens, but also *disutility* from failing to live up to some ‘ideal’ contribution level. Thus prosocial motivation derives from a process of self-image maintenance in which agents base their actions partly on the (perceived) characteristics of the public-good production function. Note that it is common for moral-licensing accounts (e.g. Sachdeva et al., 2009) to also appeal to some manner of self-concept maintenance process. We believe our model forms a useful framework for rigorous analysis of such processes, including the formulation of testable hypotheses related to moral balancing.

Compared to theories that emphasize for instance the signaling of one’s prosociality to others (e.g. Bénabou and Tirole, 2006; Andreoni and Bernheim, 2009), the moral-norms approach has the advantage of relative simplicity. Moreover, internalized norms are arguably the main drivers of behavior in many contexts. Common prosocial acts such as waste recycling, anonymous donations, or voting by mail are not readily observable by other people, and any explanation of them in terms of social image will be indirect at best.<sup>4</sup>

The structure of this paper is as follows. Section 2 outlines our multi-activity extension to the duty-orientation model. Section 3 introduces our main comparative-statics results by means of a relatively simple example. Section 4 analyzes a more general case, and section 5 considers an extension where the magnitude of the ‘ideal contribution’ is endogenously determined. Section 6 revisits earlier research on government centralization of an activity. Section 7 concludes.

## 2 The multi-activity model

Consider a community of  $N$  identical individuals. Each person may contribute to a single public good  $G$  by spending effort, measured in time units, on  $K$  different activities. As agents are identical, we denote the effort that each person spends on activity  $k$  simply as  $e_k \geq 0$ . Assuming that labor supply is fixed, the time available to each agent is some constant  $T > 0$ , and because any time not spent contributing is devoted to leisure instead, the choice facing each agent is to allocate time across leisure  $L \geq 0$  and all effort variables to maximize utility given the restriction  $L + \sum_{k=1}^K e_k = T$ .

We will examine the elements of each identical agent’s utility function in turn. As a generalization of Brekke et al. (2003), consider first

$$g = g(\mathbf{e}, \boldsymbol{\theta})$$

---

<sup>4</sup>Variants of the Brekke et al. (2003) moral-motivation model have been applied to corporate social responsibility and labor market screening (Brekke and Nyborg, 2008), hypothetical bias in stated-preference surveys (Johansson-Stenman and Svedsäter, 2012), green consumerism (Nyborg et al., 2006), and households’ recycling efforts (Bruvoll and Nyborg, 2004).

where  $g$  is the *personal* public-good production function utilized by each identical agent, and  $\mathbf{e}$  is the ‘effort vector’ (with typical element  $e_k \geq 0$ ). Finally,  $\boldsymbol{\theta}$  is the vector of parameters (with typical element  $\theta_k \geq 0$ ) regulating the productivity of all agents with respect to activity  $k$ . Because individual contributions are embedded in a wider institutional framework, these parameters reflect not only the quality of available technology but also issues of convenience and information, so long as these are relevant for how much is accomplished per unit of effort. For example, suppose the local government launches a campaign to facilitate household recycling by increasing the number of drop-off sites and by sending out leaflets on how to recycle. Both interventions would arguably cause the relevant  $\theta$  to increase. However, ‘nudging’ policies exploiting framing or peer effects would not.

Production is characterized by the following.  $g$  is twice continuously differentiable, strictly concave, and increasing in all  $e_k$  and  $\theta_k$ . Productivity is convex, with  $g''_{e_k\theta_k} > 0$ . We exclude direct productivity spillovers across activities, so for all  $l \neq k$ ,  $g''_{e_k\theta_l} = g''_{\theta_k\theta_l} = 0$ . While this last assumption may seem strong,  $\theta$ s are activity-specific by definition: for instance, a policy affecting more than one activity should be viewed as shifting *both* productivity parameters.

There are some additional technicalities. If productivity with respect to a particular activity is zero, then increasing the amount of effort put into that activity has no effect on production, so  $g'_{e_k} > 0$  if and only if  $\theta_k > 0$ . Unless stated otherwise, we will assume that productivity is indeed nonzero. Furthermore, if no effort is spent on a given activity, changing its productivity does not affect production:  $g'_{\theta_k} > 0$  if and only if  $e_k > 0$ . Hence, defining  $g^0 = g(\mathbf{0}, \boldsymbol{\theta})$ , we always have  $g(\mathbf{e}, \boldsymbol{\theta}) \geq g^0$ . In many cases, it will be natural to set  $g^0 = 0$ .<sup>5</sup>

Like Andreoni (1990), we assume that people derive utility from personally contributing to the public good. Such a ‘warm-glow’ component may be given various psychological interpretations, including a favorable sense of social image or self-image as a generous person (Bénabou and Tirole, 2006), or as the desire to help other contributors feel validated (Rotemberg, 2014). Brekke et al. (2003) interpret it as the motivation to adhere to a moral norm and formulate a ‘self-image function’ reflecting whether the agent views herself as a responsible citizen. Our multi-activity version is given by

$$I = I(g(\mathbf{e}, \boldsymbol{\theta}) - g^*) \tag{1}$$

where  $g(\mathbf{e}, \boldsymbol{\theta})$  is the amount that each agent actually contributes, while  $g^* \geq g^0$  is the ‘ideal’ amount that a truly responsible citizen would contribute; unless stated otherwise, we will

---

<sup>5</sup>Given all of the above assumptions, one can show that it must be possible to write the production function as  $g = f_0(e_1, \dots, e_K) + \sum_k f_k(e_k, \theta_k)$ . Hence, our framework includes the multiple-public-goods model where  $g = \sum_k g_k(e_k, \theta_k)$  as a special case. But as this expression explicitly states that each  $g_k$  uses only a single input activity, the converse is not true.

suppose  $g^* > g^0$ . In line with psychological theories of ‘self-discrepancy’ (e.g. Higgins, 1987), the utility of contributing is a function of the distance between the two.  $I$  is twice continuously differentiable, and maximized when  $g(\mathbf{e}, \boldsymbol{\theta}) = g^*$ ; furthermore,  $I' > 0$  whenever  $g < g^*$ ;  $I' = 0$  when  $g = g^*$ ;  $I' < 0$  when  $g > g^*$ ; and  $I'' \leq 0$  everywhere.

The ideal  $g^*$  itself could be exogenous or endogenous. Throughout most of the paper we will suppose it exogenous, interpreting it as some standard of conduct existing in the social environment independently of individual agents. People form beliefs about it and may eventually internalize it as simply ‘the right thing to do’ in particular or general contexts. This interpretation squares well with Akerlof and Kranton (2000)’s incorporation of social identity theory (e.g. Tajfel and Turner, 1979) into economics. In their model, individuals first subscribe to a particular social identity, which then provides them with a number of behavioral dos and don’ts that might best be viewed as exogenous from the point of view of a single person. Section 5 considers endogenous ideal formation of the particular type explored in Brekke et al. (2003).

We can now formulate the utility function that each agent maximizes. It is assumed to be additively separable in its arguments; each agent solves

$$\max_{\mathbf{e}} U(\mathbf{e}, \boldsymbol{\theta}) = u\left(T - \sum_{k=1}^K e_k\right) + v(G_{-i}(\boldsymbol{\theta})) + I(g(\mathbf{e}, \boldsymbol{\theta}) - g^*) \quad (2)$$

subject to  $0 \leq \sum_k e_k \leq T$ : we have used the budget constraint to define utility strictly in terms of the agent’s effort vector. The utility-of-leisure function  $u$  is twice continuously differentiable, strictly increasing and concave, as is  $v(G)$ , the utility of consuming the public good.<sup>6</sup> We assume throughout most of this paper that  $N$  is large enough that each agent approximates own contributions by zero, so the argument of  $v$  is simply other people’s public good production  $G_{-i} = (N - 1)g$ . As is standard, each agent views  $G_{-i}$  as exogenous, so choices are entirely nonstrategic. Nevertheless, because  $v(G_{-i})$  may change as other agents adjust their efforts in response to productivity shocks, this term will prove important for analyzing changes in individual utility in section 6.

As  $U$  is continuous and the choice set is compact, the set of solutions to utility problem

---

<sup>6</sup>Since activities may directly impact the well-being of the person performing them,

$$U(\mathbf{e}, L, \boldsymbol{\theta}) = u(\mathbf{e}, L) + v(G_{-i}(\boldsymbol{\theta})) + I(g(\mathbf{e}, \boldsymbol{\theta}) - g^*)$$

might be a more general, though also significantly more complex, utility model. Here activities are arguments not only of the self-image function but also directly of  $u$ . Acts that entail discomfort beyond foregone leisure would then feature negatively in  $u$ , while those seen as valuable in their own right would feature positively.



(2) is nonempty. The necessary and sufficient Kuhn-Tucker conditions for a solution are

$$-u' + g'_{e_k} I' - \lambda \leq 0 \quad (3)$$

for all  $k$ , with equality if  $e_k > 0$ . Here  $\lambda \geq 0$  is the Lagrange multiplier associated with the restriction  $\sum_k e_k \leq T$ , and the usual complementary slackness conditions imply  $\lambda = 0$  if  $\sum_k e_k < T$ . It is then easy to demonstrate the following property of any solution (all proofs can be found in Appendix B at the end of this paper).

**Proposition 1.** *If  $\bar{e}$  is a solution to problem (2), we have  $g^0 \leq g(\bar{e}) \leq g^*$ . If also  $g^* > g^0$ , then  $g(\bar{e}) < g^*$ .*

Thus no agent will produce more than the ideal amount, and except in the trivial case where  $g^* = g^0$ , all agents will produce strictly less than the ideal. However, in general we cannot rule out corner solutions where it is optimal to set all effort variables to zero and produce the minimal amount  $g^0$ .

For  $k \neq l$ , we define productivity-led spillovers as derivatives  $d\bar{e}_l/d\theta_k$ , i.e. as the marginal impact of  $\theta_k$  on optimal effort element  $\bar{e}_l$ . If the production function is linear ( $g = \sum_k \theta_k e_k$ ), or more generally whenever we can write  $g = g(\theta_1 e_1, \dots, \theta_K e_K)$ , these effects correspond to standard notions of price-driven gross substitution. To see this, suppose we interpret activities as units of production rather than durations; that is, define  $\hat{e}_k = \theta_k e_k$  for all  $k$ . Then we have  $g = g(\hat{e}_1, \dots, \hat{e}_K)$  and the budget constraint  $0 \leq \sum_k \hat{e}_k/\theta_k \leq T$ , implying prices  $p_k = 1/\theta_k$ . It follows that e.g.  $d\hat{e}_2/dp_1 > 0$  if and only if  $d\hat{e}_2/d\theta_1 < 0$ , which is itself equivalent to  $de_2/d\theta_1 < 0$ . Thus crowding-out corresponds to the redefined alternatives being gross substitutes, and crowding-in similarly corresponds to gross complementarity.

### 3 An example: individual carbon footprints

In Section 4, we discuss the model in general terms. First, however, let us build intuition by a rather more specific example. Suppose we interpret the public good as absence of the ‘bad’  $F$  produced by emissions of carbon dioxide. Individuals can reduce per-capita emissions (their personal ‘carbon footprint’) by engaging in some number of activities. As our point of comparison, we will first assume that there is only a single climate-friendly action that can be undertaken, that its corresponding effort variable is  $e_1 = e$ , and that its productivity parameter is  $\theta_1 = \theta$ ; we will consider two activities shortly. For now, use the simple function

$$g = -F = -(F^0 - \theta\sqrt{e})$$

$F^0 > 0$  is the amount emitted if no action is taken. In the real world, we expect at any given time that some proportion of the population is indeed taking action; therefore this parameter is best interpreted as lying above the actual population average.

The ideal contribution  $g^*$ , which is exogenous, translates into the negative of some ideal emissions profile  $0 \leq F^* < F^0$ . This parameter might be, for instance, some globally equitable ‘fair share’, in per-capita emission terms, of the carbon budget needed to keep global warming below two degrees Celsius. For example, the most ambitious emissions scenario described in Annex II of IPCC (2013) requires that global annual CO<sub>2</sub> emissions drop to 3.50 gigatonnes by 2050. Assuming a population of ten billion people, this yields an equal share of 0.35 tonnes per capita and year.<sup>7</sup>

The utility function we will use is designed to be as simple as possible. If utility-of-leisure is linear with a marginal utility of one, and the self-image function is a quadratic function with a maximum of zero, individuals simply maximize

$$U(e) = T - e + G_{-i} - a(-F - (-F^*))^2 = T - e + G_{-i} - a(F^* - F^0 + \theta\sqrt{e})^2 \quad (4)$$

subject to  $0 \leq e \leq T$ . Note that  $v(G_{-i})$  has been dropped from (4), as it is irrelevant for behavior. The parameter  $a$  here reflects the strength of the self-image motive. In the following, we will concern ourselves only with the class of interior solutions for which  $0 < e < T$ ; yet if  $a = 0$ , the corner solution  $e = 0$  is trivially optimal. Thus, we require  $a > 0$ .<sup>8</sup>

We denote the interior maximum to (4) by  $\bar{e}$ . It is

$$\bar{e} = \left( \frac{a\theta(F^0 - F^*)}{1 + a\theta^2} \right)^2 \quad (5)$$

Let us also write

$$F(\bar{e}) = F^0 - \theta\sqrt{\bar{e}} = \frac{F^0 + a\theta^2 F^*}{1 + a\theta^2} \quad (6)$$

where it is easy to check that  $F^* < F(\bar{e}) < F^0$ , in confirmation of Proposition 1. Now suppose that some new government policy is implemented with the aim of making it easier for individuals to lower their emissions. In other words, suppose that  $\theta$  increases. Then

$$\frac{dF(\bar{e})}{d\theta} = -\frac{2a\theta(F^0 - F^*)}{(1 + a\theta^2)^2} < 0 \quad (7)$$

so all agents reduce their emissions in response; but

$$\frac{d\bar{e}}{d\theta} = \frac{2a^2\theta(F^0 - F^*)^2}{(1 + a\theta^2)^3} (1 - a\theta^2) \quad (8)$$

---

<sup>7</sup>Of course, our choice of the year 2050 arbitrary; also, since IPCC (2013) states that warming is largely determined by cumulative emissions, strictly positive per-capita emissions may not be viable in the long run.

<sup>8</sup>The marginal product of effort approaches infinity as  $e \rightarrow 0$ , so any optimum must have  $e > 0$  if  $a > 0$ . However, corner solutions for which  $e = T$  (so  $L = 0$ ) are possible in general, but are here assumed not to arise.

This expression is positive if and only if  $a \leq 1/\theta^2$ , a condition which if combined with (6) can be reformulated in emissions terms as  $F(\bar{e}) \geq (F^0 + F^*)/2$ . This tells us that efforts increase only if emissions were high enough initially.<sup>9</sup> Of course, if productivity keeps increasing, effort will always drop eventually. Indeed, in the limit as  $\theta \rightarrow \infty$ , it becomes possible to live up to the ideal with next to no effort:  $\bar{e} \rightarrow 0$  but, by (6),  $F(\bar{e}) \rightarrow F^*$ .

Such are the features of the single-activity version of the model. We now move on to its two-activity counterpart. The production function is now  $g = -F = -(F^0 - \theta_1\sqrt{e_1} - \theta_2\sqrt{e_2})$ , so contributions made through  $e_1$  and  $e_2$  are simply added up. Agents maximize

$$U(e_1, e_2) = T - e_1 - e_2 + G_{-i} - a(F^* - F^0 + \theta_1\sqrt{e_1} + \theta_2\sqrt{e_2})^2 \quad (9)$$

and similarly to before, we focus on the interior solutions where  $e_1 > 0, e_2 > 0$  (true for  $a > 0$ ) but  $e_1 + e_2 < T$  (need not be true). For  $k = 1, 2$ , solving for optimal efforts yields

$$\bar{e}_k = \left( \frac{a\theta_k(F^0 - F^*)}{1 + a(\theta_1^2 + \theta_2^2)} \right)^2 \quad (10)$$

and

$$F(\bar{e}_1, \bar{e}_2) = F^0 - \theta_1\sqrt{\bar{e}_1} - \theta_2\sqrt{\bar{e}_2} = \frac{F^0 + a(\theta_1^2 + \theta_2^2)F^*}{1 + a(\theta_1^2 + \theta_2^2)} \quad (11)$$

where again we have  $F^* < F(\bar{e}_1, \bar{e}_2) < F^0$ .<sup>10</sup>

Now consider an increase in  $\theta_1$  (the effect of a shift in  $\theta_2$  being completely analogous). Activity 2 should then be interpreted as an aggregate of all climate-related activities that are

---

<sup>9</sup>If one extends the model to include heterogeneity such that each agent  $i$  has some parameter value  $a_i \geq 0$ , the same conditions (with  $a$  replaced by  $a_i$ ) reveal that only the *less* image-concerned agents, for whom initial emissions were relatively high, increase their efforts. Consequently, as  $\theta$  increases, and emissions reductions can be carried out with greater ease, there is a gradual convergence of the effort put in by different types. If productivity is low, highly image-concerned agents make great efforts but unconcerned agents next to none; as productivity increases both groups, from the direction of each, approach the middle ground. If a decision maker's ambition is simply to minimize total emissions, it may thus be profitable to selectively disseminate information on productivity increases; that is, to inform only the previously unengaged, whose  $a_i$ , presumably, is small. It may be less advisable from a democratic perspective.

<sup>10</sup>It is simple to show that effort on any activity increases (implying that actual emissions drop) when  $F^*$  decreases. In the single-activity model, we have

$$\frac{d\bar{e}}{dF^*} = -\frac{2a^2\theta^2(F^0 - F^*)}{(1 + a\theta^2)^2} < 0$$

and similar results hold for each individual activity in the two-activity model. We may interpret the shift in  $F^*$  as new information becoming available, e.g. a high-profile scientific report concluding that the carbon budget deemed safe is smaller than previously expected, implying that everyone's 'fair share' of that budget diminishes.

not directly affected by the shift in  $\theta_1$ . Suppose  $e_1$  is the (extra) time invested when traveling by bicycle instead of taking the car,<sup>11</sup> and  $\theta_1$  shifts as a result of a policy to increase the number of bicycle lanes in the community. Since as a result less extra time  $e_1$  will be needed for any given trip, agents can travel by bicycle instead of by car on more trips and still hold  $e_1$  constant. Thus, the emissions reduction per unit of  $e_1$  increases. Differentiation of (11) with respect to  $\theta_1$  now reveals that

$$\frac{dF(\bar{e}_1, \bar{e}_2)}{d\theta_1} = -\frac{2a\theta_1 (F^0 - F^*)}{(1 + a(\theta_1^2 + \theta_2^2))^2} < 0 \quad (12)$$

so emissions always drop; while differentiation of  $\bar{e}_1 + \bar{e}_2$  yields

$$\frac{d\bar{e}_1}{d\theta_1} + \frac{d\bar{e}_2}{d\theta_1} = \frac{2a^2\theta_1 (F^0 - F^*)^2}{(1 + a(\theta_1^2 + \theta_2^2))^3} (1 - a(\theta_1^2 + \theta_2^2))$$

The effect on total efforts is positive iff  $a \leq 1/(\theta_1^2 + \theta_2^2)$  or, in emission terms,  $F(\bar{e}_1, \bar{e}_2) \geq (F^0 + F^*)/2$ . Again, we see that only if initial emissions are sufficiently large do total efforts increase. Clearly, aggregate behavior in the two-activity model is qualitatively identical to that of the single-activity model. However, the following important results arise when each effort variable is considered in isolation.

$$\frac{d\bar{e}_1}{d\theta_1} = \frac{2a^2\theta_1 (F^0 - F^*)^2}{(1 + a(\theta_1^2 + \theta_2^2))^3} (1 - a(\theta_1^2 - \theta_2^2))$$

$$\frac{d\bar{e}_2}{d\theta_1} = -\frac{4a^3\theta_1\theta_2^2 (F^0 - F^*)^2}{(1 + a(\theta_1^2 + \theta_2^2))^3} < 0$$

We see that the sign of  $d\bar{e}_1/d\theta_1$  is ambiguous; the more striking fact, however, is that  $d\bar{e}_2/d\theta_1$  is always strictly negative. Thus, overall the directed productivity shock drives negative spillovers across activities, with aggregate effort on other activities being crowded out. The interpretation could be the following. Because bicycling is now easier, the agent undertakes fewer car trips and, as a result, feels licensed to engage more in other carbon-intensive activities, such as consuming red meat. But the compensation is incomplete; (12) shows that in the net, total emissions still drop.

We see that to fully understand engagement with any one activity, we must consider the characteristics of other activities as well. Suppose that we were to evaluate a policy designed

---

<sup>11</sup>It is in the abstract nature of our model that this example is not entirely natural, because there are issues involved of not only time, but also money, convenience, etc. Also, in real life there are upper limits to what each activity, taken alone, can achieve. No such limit exists here.

to increase the productivity of a particular activity. In principle, all impacts of this policy should be considered in its evaluation. It is true that the single-activity model correctly associates increased productivity with reduced emissions for all agents; but comparison of (7) and (12) shows that this reduction is overestimated. Because the single-activity model considers effort only towards the particular activity targeted by the productivity shock, it misses the larger picture, which is that the policy drives lower efforts on other activities.

The next section will show that an important determinant of the sign of spillovers is the cross-partial derivative of the public-goods production function,  $g''_{e_1 e_2} = \partial^2 g / \partial e_1 \partial e_2$ , which in this example was equal to zero, as  $g$  was additive. We will also see that crowding-out of the same type as observed here obtains for a broad set of production technologies.

#### 4 Exogenous ideals

With  $g^*$  remaining exogenous, this section will extend most of the results of the previous example to a more general case by applying the theory of monotone comparative statics (Milgrom and Shannon, 1994). This is an ordinal theory that does not presuppose the solution to be unique or interior, or the parameter shift to be marginal.

To facilitate understanding of the results to follow, we begin by stating some fundamentals of lattice theory. A *lattice* is a particular type of partially ordered set  $X$  with the property that, for any two elements  $x, y \in X$ , both their join (or least upper bound)  $x \vee y$  and their meet (or greatest lower bound)  $x \wedge y$  exist as elements of  $X$ . In this paper, all partially ordered sets will be given by Euclidean space with the *coordinatewise* order, where for two vectors  $x, y \in \mathbb{R}^n$ ,  $x \geq y$  if  $x_i \geq y_i$  for each  $i = 1, \dots, n$ . In this case the definitions of join and meet are particularly simple, namely  $x \vee y = (\max\{x_1, y_1\}, \dots, \max\{x_n, y_n\}) \in \mathbb{R}^n$  and  $x \wedge y = (\min\{x_1, y_1\}, \dots, \min\{x_n, y_n\}) \in \mathbb{R}^n$ . It follows from the definition that  $\mathbb{R}^n$  is a lattice, as both join and meet of any two  $n$ -dimensional vectors are in that set.

A *sublattice* is some subset  $S$  of a lattice  $X$  for which both the join and meet (in  $X$ ) of any two points in  $S$  are elements of  $S$ .

Below we will be interested in seeing how the set of solutions to (2) changes when we shift some productivity parameter  $\theta_k$ . We will use the following comparative-statics result, which is slightly adapted from Theorem 4 in Milgrom and Shannon (1994). Since we will shift only one parameter  $r$  at a time, the parameter space  $R$  is simply the positive real line.

**Proposition 2.** *Let  $U : X \times R \rightarrow \mathbb{R}$ , where  $X$  is a lattice and  $R$  is a partially ordered set. If the constraint set  $S : R \rightarrow 2^X$  is nondecreasing in  $r$  and if  $U$  is quasimodular in  $\mathbf{e}$  and satisfies the single-crossing property in  $(\mathbf{e}; r)$ , then  $\arg \max_{\mathbf{e} \in S(r)} U(\mathbf{e}, r)$  is monotone nondecreasing in  $r$ .*

Here (monotone) nondecreasingness is defined in terms of the *strong set order*  $\geq_s$ : for  $X$  a lattice with some given relation  $\geq$  (such as the coordinatewise order on  $\mathbb{R}^n$ ), with  $Y$  and  $Z$  subsets of  $X$ , we say that  $Y \geq_s Z$  if for every  $y \in Y$  and  $z \in Z$ ,  $y \vee z \in Y$  and  $y \wedge z \in Z$ . If the stated conditions hold, Proposition 2 then implies that for  $r' \geq r$ ,  $\arg \max_{\mathbf{e} \in S(r')} U(\mathbf{e}, r') \geq_s \arg \max_{\mathbf{e} \in S(r)} U(\mathbf{e}, r)$ . In the special case where the solution sets of (2) are singleton, the set order  $\geq_s$  reduces to the coordinatewise order, with Proposition 2 simply stating that all choice variables are increasing in  $r$ .

For  $U$  twice continuously differentiable, quasisupermodularity holds if  $\partial^2 U / \partial e_k \partial e_l \geq 0$  for all  $k \neq l$ , and the single-crossing condition holds if  $U$  exhibits ‘increasing differences’ such that  $\partial^2 U / \partial e_k \partial r \geq 0$  for all  $k$ . Hence, quasisupermodularity represents a general type of complementarity (in utility terms) between the choice variables, while the single-crossing property similarly represents complementarity between the choice variables and the parameter being shifted.

Under what conditions can Proposition 2 be applied to utility problem (2), with  $r = \theta_k$ ? While checking for quasisupermodularity and the single crossing property is straightforward, nondecreasingness of the constraint set  $S = \{(e_1, \dots, e_K) | 0 \leq \sum_k e_k \leq T\}$  is trickier. This set does not depend on productivity, so for any  $\theta_k$  we will have  $S(\theta_k) = S$ . We therefore obtain nondecreasingness in  $\theta_k$  (as required by the theorem) if and only if  $S$  is a sublattice, as it is then easy to verify that  $S \geq_s S$ . However, when  $K > 1$ , our constraint set  $S$  is *not* a sublattice of the lattice  $X = \mathbb{R}_+^K$ . To see why, take effort vectors  $e' = (T, 0, \dots, 0, 0) \in S$  and  $e'' = (0, 0, \dots, 0, T) \in S$ . Under the coordinatewise order,  $e' \vee e'' = (T, 0, \dots, 0, T)$ , which is not in  $S$ . Hence,  $S$  is not a sublattice.

Fortunately, for  $K = 2$ , there is a workaround. For  $k \neq l$ , redefine (2) as<sup>12</sup>

$$\max_{e_k, \tilde{e}_l} U(e_k, \tilde{e}_l, \theta_k) = u(T - e_k + \tilde{e}_l) + v(G_{-i}(\theta_k)) + I(g(e_k, -\tilde{e}_l, \theta_k) - g^*) \quad (13)$$

with  $\tilde{e}_l = -e_l$ . Then for  $X = \{(e_k, \tilde{e}_l) | e_k \geq 0, \tilde{e}_l \leq 0\}$  and  $S = \{(e_k, \tilde{e}_l) \in X | 0 \leq e_k - \tilde{e}_l \leq T\}$ ,  $S$  is in fact a sublattice of  $X$  in the coordinatewise order. Hence, as long as  $K = 2$ , there is no immediate obstacle to applying Proposition 2.<sup>13</sup> When  $e_l$  is redefined in the above manner, Proposition 2 implies a tendency for negative spillovers. In particular, if all solution

<sup>12</sup>For notational brevity, we suppress  $\theta_l$  as an argument of  $U$  throughout this section.

<sup>13</sup>This method only works if  $K = 2$ . For example, for  $K = 3$ , define  $\tilde{e}_2 = -e_2$  and  $\tilde{e}_3 = -e_3$  and let  $S = \{(e_1, \tilde{e}_2, \tilde{e}_3) | e_1 - \tilde{e}_2 - \tilde{e}_3 \leq T, e_1 \geq 0, \tilde{e}_2 \leq 0, \tilde{e}_3 \leq 0\}$ . Then, for  $e' = (0, -T, 0) \in S$  and  $e'' = (0, 0, -T) \in S$ , under the coordinatewise order  $e' \wedge e'' = (0, -T, -T) \notin S$ . Using some simple order on  $X$  other than the coordinatewise one does not solve this problem; for instance, letting  $x \geq y$  if  $x_1 \geq y_1$ ,  $x_2 \leq y_2$ , and  $x_3 \leq y_3$  is exactly equivalent to using  $\tilde{e}_2 = -e_2$  and  $\tilde{e}_3 = -e_3$  in the usual coordinate order: for  $e' = (0, T, 0) \in S$  and  $e'' = (0, 0, T) \in S$ , the alternative order implies  $x \wedge y = (0, T, T) \notin S$ .

sets are singleton,  $e_k$  is increasing in  $\theta_k$  but  $e_l = -\tilde{e}_l$  is decreasing. This is the ‘crowding-out’ effect observed in the previous section. As before, in most settings activity  $l \neq k$  should be interpreted as an aggregate of all activities that are only indirectly affected by the productivity shock. Although we will be able to describe the movement of the aggregate, we can draw no conclusions about particular activities within it.

#### 4.1. The sign of policy spillovers

With theoretical foundations now in place, we will proceed to analyze the effect of activity-specific policy. It may be helpful to begin by briefly considering the single-activity model

$$U(e, \theta) = u(T - e) + v(G_{-i}(\theta)) + I(g(e, \theta) - g^*)$$

Note that the interval  $0 \leq e \leq T$  is a sublattice of  $\mathbb{R}_+$ , so analyzing a shift in  $\theta$  using Proposition 2 is feasible. Since there is only one effort variable, we only need to check the single-crossing property. Using the notation  $g' = dg/de$ ,  $g'_\theta = dg/d\theta$ , and  $g'' = d^2g/de^2$ , we find effort increasing in line with productivity if

$$\frac{\partial^2 U(e, \theta)}{\partial e \partial \theta} = I'' g' g'_\theta + I' g''_{e\theta} \geq 0 \quad (14)$$

This condition is analogous to that required for (8) to be positive; indeed, for  $I = -a(F^* - F)^2$  and  $g = -F = \theta\sqrt{e} - F^0$ , it translates into  $F(e) \geq (F^0 + F^*)/2$ . It specifies when  $e$  and  $\theta$  are complements in a utility sense. When  $\theta$  increases, on the one hand the marginal productivity of additional effort grows since  $g''_{e\theta} > 0$ . On the other hand,  $I$  is concave. As there is a ‘first-order’ effect  $g'_\theta$  (from shifting  $\theta$  but hypothetically holding effort constant) on production, the marginal benefit in utility terms of any extra effort decreases. Complementarity obtains if and only if the former effect dominates.<sup>14</sup>

Now, suppose  $K = 2$  and consider a shift in activity-specific productivity parameter  $\theta_k$ . We differentiate (13) twice to find sufficient conditions for Proposition 2 to hold:

$$\frac{\partial^2 U(e_k, \tilde{e}_l, \theta_k)}{\partial e_k \partial \theta_k} = I'' g'_{e_k} g'_{\theta_k} + I' g''_{e_k \theta_k} \geq 0 \quad (15)$$

$$\frac{\partial^2 U(e_k, \tilde{e}_l, \theta_k)}{\partial \tilde{e}_l \partial \theta_k} = -I'' g'_{e_l} g'_{\theta_k} - I' g''_{e_l \theta_k} \geq 0 \quad (16)$$

$$\frac{\partial^2 U(e_k, \tilde{e}_l, \theta_k)}{\partial e_k \partial \tilde{e}_l} = -u'' - I'' g'_{e_k} g'_{e_l} - I' g''_{e_1 e_2} \geq 0 \quad (17)$$

---

<sup>14</sup>Since in general both terms in (14) depend on  $e$ , the condition may be satisfied only for some initial values of  $e$ . We saw in footnote 12 that if agents are heterogeneous the idea may be extended to people, with different types reacting differently.

Here conditions (15) and (16) check for increasing differences, while (17) is a quasisupermodularity condition. If all three hold, the set of optimal redefined efforts  $(e_k, \tilde{e}_l)$  is monotone nondecreasing in  $\theta_k$ ; in other words, spillovers will tend to be negative.

Condition (15) is exactly the same type of within-activity complementarity condition we observed in the single-variable case.<sup>15</sup> Condition (16) states that  $\theta_k$  and  $e_l$  should be substitutes in utility terms; since there are no productivity spillovers ( $g''_{e_l\theta_k} = 0$ ), their only interaction is through the first-order effect  $g'_{\theta_k}$ , so it is immediately true.

Finally, (17) states that  $e_k$  and  $e_l$  should be substitutes in a utility sense. It is notable that for this to hold, it is sufficient that the two activities are substitutes in the production of the public good, i.e. that  $g''_{e_1e_2} \leq 0$ ; in particular, we do not need to assume that  $g$  is concave or quasiconcave. Public goods to which people commonly contribute arguably tend to satisfy this substitute condition, if only because many production functions likely have  $g''_{e_1e_2} = 0$ . Also, consider an example. An agent has to decide how to split a donation (or time spent doing voluntary work) across two charities working with the same social issues; but if the benefit of this type of social work is a (one-dimensional) concave function, any amount which the agent gives to one charity will weakly reduce the marginal benefit of also donating to the other.<sup>16</sup>

#### 4.2. Do negative spillovers reverse the effect of policy?

We have just seen that when activities are substitutes in public-good production, we might expect a qualitative crowding-out effect to dominate. What are the quantitative properties of that effect? In particular, is crowding-out incomplete in the sense that public-good production increases with  $\theta_k$ , as it did in the previous section? To answer this question, we will redefine the utility problem as a single-variable one where agents choose their sum of efforts  $\hat{e} = e_1 + e_2$ .

We begin by stating a useful result. It guarantees that for any utility maximum with associated total efforts  $\hat{e}$ , public-goods production can be understood as the value function  $\hat{g}(\hat{e})$  of a ‘production-maximization problem’ (PMP) with the restriction that total effort be exactly  $\hat{e}$ . The situation is illustrated in Figure 1. Utility maximization becomes a two-stage process; we will focus on the agent’s choice of total efforts, which are then allocated in an

---

<sup>15</sup>Conditions (15)-(17) are sufficient but not necessary. For example, again given  $I = -a(F^* - F)^2$  and  $g = -F = \theta_1\sqrt{e_1} + \theta_2\sqrt{e_2} - F^0$ , condition (15) translates into  $2F(e_1, e_2) \geq (F^0 + F^*) - \theta_l\sqrt{e_l}$ . Combining with (10) and (11), we obtain  $a < 1/\theta_k^2$ , which is a stronger requirement than  $a < 1/(\theta_k^2 - \theta_l^2)$ , the necessary and sufficient condition for  $d\theta_k/de_k$  to be positive.

<sup>16</sup>The alternative, where  $g''_{e_1e_2} > 0$  (so activities are complements in production), would reflect a ‘weakest link’ structure. For example, coral reefs across the world face multiple threats including unsustainable fishing, pollution, and climate change (Burke et al., 2011). If each of these alone represents an existential threat to the reefs, the marginal benefit of addressing one cause will *increase*, the more action is taken on the others.



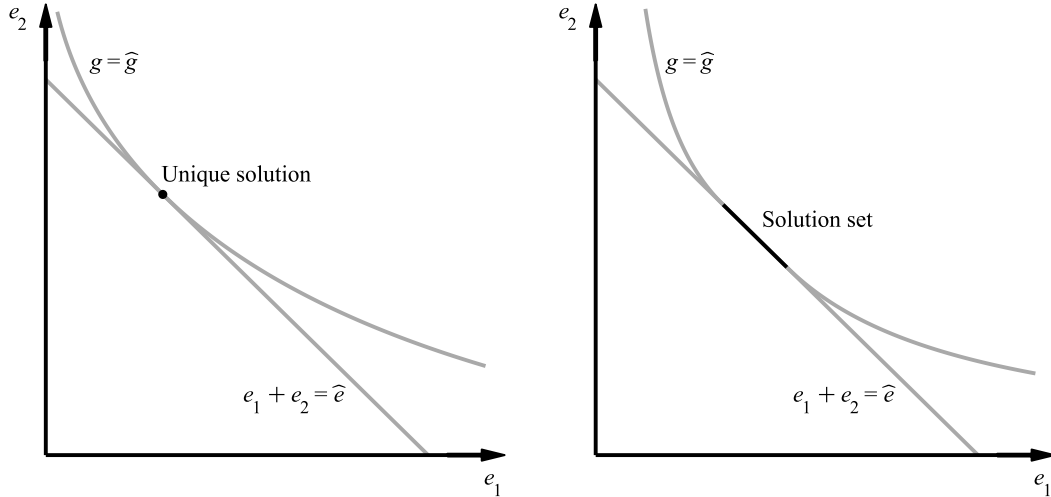
optimal fashion across specific activities. While our assumption that  $g$  is strictly concave implies that the PMP has a unique solution, the right panel of Figure 1 shows that the result holds under multiple optima as well. Lemma 1 also provides an increasingness result which is a necessary prerequisite for drawing comparative-statics conclusions about production levels.

**Lemma 1.** *Each  $\hat{e}$  that results from maximizing (2) corresponds to a unique optimal production level*

$$\hat{g}(\hat{e}, \theta_k) = \max_{e_1, e_2} g(e_1, e_2, \theta_k) \quad (18)$$

subject to  $e_1 \geq 0$ ,  $e_2 \geq 0$ , and  $e_1 + e_2 = \hat{e}$ .

Moreover,  $\hat{g}$  is strictly increasing in  $\hat{e}$ .



**Figure 1:** *The production-maximization problem. Left panel: the PMP has a unique solution. Right panel: the PMP has multiple solutions.*

It follows from Lemma 1 that a given effort vector  $\mathbf{e} = (e_1, e_2)$  will solve (2) if and only if it solves the PMP given some  $\hat{e} = e_1 + e_2$  which solves

$$\max_{\hat{e}} U_{\hat{e}}(\hat{e}, \theta_k) = u(T - \hat{e}) + v(G_{-i}(\theta_k)) + I(\hat{g}(\hat{e}, \theta_k) - g^*) \quad (19)$$

subject to  $0 \leq \hat{e} \leq T$ . The equivalence arises because any solution to either (2) or (19) will solve the PMP; the former, because of Lemma 1; the latter, by construction.<sup>17</sup>

The key to showing that production levels are monotone nondecreasing is to note how Lemma 1 implies that  $\hat{g}$ , being bijective, has an inverse  $\hat{e} = \hat{g}^{-1}(\hat{g}, \theta_k)$ . Thus, reformulate

<sup>17</sup>In Appendix A, we derive sufficient conditions for total efforts  $\hat{e}$  to increase in line with activity-specific productivity and the ideal contribution.

(19) yet again as

$$\max_{\hat{g}} U_{\hat{g}}(\hat{g}, \theta_k) = u(T - \hat{g}^{-1}(\hat{g}, \theta_k)) + v(G_{-i}(\theta_k)) + I(\hat{g} - g^*) \quad (20)$$

subject to  $0 \leq \hat{g}^{-1}(\hat{g}, \theta_k) \leq T$ . Note that although the constraint set now depends on the parameter  $\theta_k$ , Proposition 2 still applies, as the set is nondecreasing. That is, for  $\theta'_k \geq \theta_k$ ,  $\{\hat{g} | 0 \leq \hat{g}^{-1}(\hat{g}, \theta'_k) \leq T\} \geq_S \{\hat{g} | 0 \leq \hat{g}^{-1}(\hat{g}, \theta_k) \leq T\}$ , or equivalently  $[g^0, \hat{g}(T, \theta'_k)] \geq_S [g^0, \hat{g}(T, \theta_k)]$ .

The single-crossing condition corresponding to (20) is

$$\frac{\partial^2 U_{\hat{g}}}{\partial \hat{g} \partial \theta_k} = u'' (\hat{g}^{-1})'_{\theta_k} (\hat{g}^{-1})'_{\hat{g}} - u' (\hat{g}^{-1})''_{\hat{g} \theta_k} \quad (21)$$

Twice totally differentiating the identity  $\hat{g}^{-1}(\hat{g}(\hat{e}, \theta_k), \theta_k) = \hat{e}$  with respect to  $\hat{e}$  and/or  $\theta_k$  reveals that  $(\hat{g}^{-1})'_{\hat{g}} > 0$ ,  $(\hat{g}^{-1})'_{\theta_k} < 0$ , and  $(\hat{g}^{-1})''_{\hat{g} \theta_k} < 0$ , so (21) is strictly positive, and production is monotone nondecreasing in  $\theta_k$ .

The following proposition summarizes what we have learned regarding policy spillovers.<sup>18</sup>

**Proposition 3.** *Suppose  $K = 2$  and consider a shift in parameter  $\theta_k > 0$ . Then the set of optimal production levels  $\arg \max_{\hat{g}} U_{\hat{g}}(\hat{g}, \theta_k)$  is monotone nondecreasing in  $\theta_k$ . Furthermore, if*

$$\begin{aligned} I'' g'_{e_k} g'_{\theta_k} + I' g''_{e_k \theta_k} &\geq 0 \\ -u'' - I'' g'_{e_k} g'_{e_l} - I' g''_{e_1 e_2} &\geq 0 \end{aligned}$$

*then there is a crowding-out effect:  $\arg \max_{(e_k, -e_l)} U(e_k, -e_l, \theta_k)$  is monotone nondecreasing in  $\theta_k$ .*

We conclude that results on policy-driven behavioral spillovers are broadly similar to those of the previous section. For a range of plausible production functions, activity-specific policy tends to drive a crowding-out effect which is incomplete in the sense that overall public-goods production increases: spillovers weaken, but do not reverse the effect of the policy. It is important to note that none of the results of Proposition 3 require that  $g$  be either concave or

---

<sup>18</sup>Regarding the ideal contribution, it is very easy to show that optimal production is monotone nondecreasing in  $g^* > g^0$ . The one-dimensional constraint set of (20) is nondecreasing in  $g^*$ , as it does not depend on this parameter and is a sublattice. It is then enough to check the single sufficient increasing-differences condition

$$\frac{\partial^2 U_{\hat{g}}}{\partial \hat{g} \partial g^*} = -I'' > 0$$

which is obviously true.

quasiconcave.<sup>19</sup> Thus, when the ideal contribution is exogenous, the pattern of negative-sign spillovers seems relatively clear-cut.

## 5 Endogenous ideals

Thus far, we have assumed that the ideal contribution is exogenous, but there are at least two reasons to suppose that people sometimes update their beliefs about the proper course of action in response to activity-specific policy. First, what one ideally *should* do is arguably dependent in part on what one *can* do. Thus, as capabilities increase, people may feel that ‘I could have done more’. Second, activity-specific policy may signal that either a specific activity or the overall public good is more important than previously thought.

Brekke et al. (2003) attempt to capture both these concerns by letting an activity-specific ideal be formed endogenously through introspection and Kantian (categorical imperative-type) moral reasoning. We will generalize their approach to a two-activity setting. Throughout this section, we suppose that all solutions are interior and unique and that  $g''_{e_k e_k} \leq g''_{e_1 e_2} \leq 0$  for  $k = 1, 2$ . Finally, for tractability, we assume that utility is quasilinear such that  $v'' = 0$ ,  $v' = 1$ , noting that this permits us to consider situations where each agent does *not* approximate own contributions by zero.

In the endogenous-ideal model, there is a preliminary stage before utility maximization is performed. At this initial stage, all agents ask themselves: ‘What would be the outcome if everyone acted like me?’ Then, starting from that thought experiment, they maximize a social welfare function which aggregates all (identical) individual utilities  $U$ . If the welfare function is utilitarian (all  $N$  utilities are given equal weight), each agent considers maximization of

$$W(e_1, e_2) = NU(e_1, e_2) = N \left( u(T - e_1 - e_2) + Ng(e_1, e_2) \right) \quad (22)$$

subject to  $0 \leq e_1 + e_2 \leq T$ . Call (22) the agent’s *welfare* maximization problem. The effort vector  $(e_1, e_2)$  is associated with and chosen by the particular individual performing the thought experiment. No matter which effort vector is chosen, it will result in  $g = g^*$ , so  $I$  does not appear in (22).

Denote a welfare solution as  $(e_1^*, e_2^*)$ ; in addition  $g^* = g(e_1^*, e_2^*)$  by construction. The above analysis accommodates the impact of new information (the policy signaling effect), as agents’ view of the utility function  $U$  is not necessarily correct. In particular, information campaigns

---

<sup>19</sup>Also, since Proposition 3 does not depend on the magnitude of the ideal contribution, it applies without modification to more typical warm-glow functions of the type  $I(g(e_1, e_2, \theta_1, \theta_2))$ , where  $I$  is everywhere increasing in  $g$  and concave.

may cause agents to revise their idea of the shape of the production function  $g$ , possibly leading to a new welfare optimum (Nyborg, 2011).

In any case, once the welfare optimum has been calculated, to determine actual efforts, each individual maximizes *utility* as given by

$$U(e_1, e_2) = u(T - e_1 - e_2) + G_{-i} + g(e_1, e_2) + I(g(e_1, e_2) - g^*) \quad (23)$$

subject to  $0 \leq e_1 + e_2 \leq T$ . At this point, the solution from the welfare problem is again treated as a fixed ideal. Thus, the magnitude of  $g^*$  is the only information from the welfare problem that is retained in utility maximization.

We will now examine whether our previous results on policy-driven spillovers still hold for the above model. As a brief indication that they do not, let us return to the carbon-dioxide example of Section 3. Welfare optimization in this setting entails maximizing

$$W(e_1, e_2) = N(T - e_1 - e_2 - N(F^0 - \theta_1\sqrt{e_1} - \theta_2\sqrt{e_2}))$$

and produces interior solution  $(e_1^*, e_2^*) = (\theta_1^2 N^2/4, \theta_2^2 N^2/4)$ . This implies  $F^* = F^0 - (N/2)(\theta_1^2 + \theta_2^2)$ , which clearly depends on both productivity parameters. Inserting this ideal into utility maximum (10) yields<sup>20</sup>

$$e_k = \left( \frac{aN\theta_k(\theta_1^2 + \theta_2^2)}{2(1 + a(\theta_1^2 + \theta_2^2))} \right)^2$$

for  $k = 1, 2$ . Differentiating, we find that

$$\frac{de_2}{d\theta_1} = \frac{a^2 N^2 \theta_1 \theta_2^2 (\theta_1^2 + \theta_2^2)}{(1 + a(\theta_1^2 + \theta_2^2))^3} > 0$$

with an analogous expression for  $de_1/d\theta_2$ . Thus, we have crowding-in. The two models — exogenous and endogenous ideal formation — deliver opposing predictions on the sign of spillovers in this particular case.

However, returning to problems (22) and (23), it will become apparent that while the general endogenous-ideal model does not rule out crowding-in, it also does not guarantee it. The analysis to follow is based on implicit-function methods. The monotone comparative-statics method used in Section 4 is of limited value here, as it is unable to demonstrate positive spillovers: recall from the discussion on sublattices that one of the activities needs to be redefined as  $\tilde{e}_l = -e_l$ .<sup>21</sup>

---

<sup>20</sup>Recall that unlike (23), problem (9) had agents approximating their own contribution to overall emissions by zero. In the present example, dropping the own-contributions term  $g(e_1, e_2)$  does not affect the sign of spillovers.

<sup>21</sup>Using implicit-function methods to solve the exogenous-ideal model under the assumptions used in this section produces unambiguous crowding-out.

At an interior solution to the welfare maximization problem, first-order conditions are  $-u' + Ng'_k = 0$ , for  $k = 1, 2$ . Implicit differentiation with respect to  $\theta_k$  produces the following system of equations:

$$(\mathbf{u}'' + N\mathbf{H}_g) \mathbf{e}_\theta^* = -N\mathbf{g}_{e\theta}'' \quad (24)$$

Here  $\mathbf{u}''$  is a  $2 \times 2$  matrix for which every element is  $u''$ ,  $\mathbf{H}_g$  is the Hessian matrix of  $g$ , and  $\mathbf{g}_{e\theta}''$  is a diagonal matrix where, for  $k = 1, 2$ , element  $(k, k)$  is  $g''_{e_k\theta_k}$ . We are interested in  $\mathbf{e}_\theta^*$ ; element  $(k, l)$  of this matrix is  $de_k^*/d\theta_l$ . System (24) has a solution if and only if the matrix  $\mathbf{u}'' + N\mathbf{H}_g$  is invertible; and

$$|\mathbf{u}'' + N\mathbf{H}_g| = Nu''(g''_{e_1e_1} + g''_{e_2e_2} - 2g''_{e_1e_2}) + N^2(g''_{e_1e_1}g''_{e_2e_2} - (g''_{e_1e_2})^2) \geq 0 \quad (25)$$

by the concavity of  $g$ .<sup>22</sup> In any reasonably well-behaved problem, this determinant is nonzero.<sup>23</sup> Then, once we have  $\mathbf{e}_\theta^*$ , we may use it to calculate the pair  $\mathbf{g}_\theta^*$ , for which each element  $k = 1, 2$  is

$$\frac{dg^*}{d\theta_k} = g'_{e_1} \frac{de_1^*}{d\theta_k} + g'_{e_2} \frac{de_2^*}{d\theta_k} + g'_{\theta_k}$$

Finally, we consider the impact of increased productivity on the utility-maximizing solution, taking the changed ideal into account. The first-order conditions of the utility-maximization problem (23) are  $-u' + (1 + I')g'_{e_k} = 0$  for  $k = 1, 2$ . Noting that  $g^* = g^*(\theta_k)$ , we differentiate these conditions with respect to  $\theta_k$  to produce the implicit-function system

$$(\mathbf{u}'' + (1 + I')\mathbf{H}_g + I''\nabla_{\mathbf{e}}\mathbf{g}(\nabla_{\mathbf{e}}\mathbf{g})') \mathbf{e}_\theta = -((1 + I')\mathbf{g}_{e\theta}'' + I''\nabla_{\mathbf{e}}\mathbf{g}(\nabla_{\theta}\mathbf{g} - \mathbf{g}_\theta^*))' \quad (26)$$

where  $\nabla_{\mathbf{e}}\mathbf{g}$  is the gradient of  $g$  with respect to the effort variables, and  $\nabla_{\theta}\mathbf{g} = (g'_{\theta_1}, g'_{\theta_2})'$  is the pair of ‘first-order’ effects of the productivity parameters on production.  $\mathbf{e}_\theta$  is the sought-after matrix of derivatives, as element  $(l, k)$  of  $\mathbf{e}_\theta$  is  $de_l/d\theta_k$ .

System (26) has a solution if and only if the matrix  $\mathbf{u}'' + (1 + I')\mathbf{H}_g + I''\nabla_{\mathbf{e}}\mathbf{g}(\nabla_{\mathbf{e}}\mathbf{g})'$  is invertible. Again, it will be in most problems, since at an interior solution the determinant

<sup>22</sup>Strict concavity implies not only  $g''_{e_1e_1}g''_{e_2e_2} - (g''_{e_1e_2})^2 \geq 0$ , but  $g''_{e_1e_1} + g''_{e_2e_2} - 2g''_{e_1e_2} \leq 0$  as well. To see why, note that the former condition implies

$$g''_{e_1e_2} \geq -\sqrt{g''_{e_1e_1}g''_{e_2e_2}} = -\sqrt{(-g''_{e_1e_1})(-g''_{e_2e_2})} \geq -\frac{g''_{e_1e_1} - g''_{e_2e_2}}{2} = \frac{g''_{e_1e_1} + g''_{e_2e_2}}{2}$$

where the last inequality is due to the relationship between arithmetic and geometric means. Thus  $g''_{e_1e_1} + g''_{e_2e_2} - 2g''_{e_1e_2} \leq 0$ .

<sup>23</sup>The textbook example of a function for which both terms in parentheses could equal zero is a higher-order polynomial like  $-x_1^4 - x_2^4$ , whose Hessian determinant is zero in exactly one point. This function is not strictly increasing, however, and while it is probably possible to construct a different counterexample which satisfies all our assumptions about  $g$  (for instance, in one dimension, the primitive function of  $10 - \tan^{-1}(x - 1)^3$  does), clearly it would need to be very specific.

of this matrix is equal to

$$(1 + I') (g''_{e_1 e_1} + g''_{e_2 e_2} - 2g''_{e_1 e_2}) (u'' + (g'_{e_1})^2 I'') + (1 + I')^2 (g''_{e_1 e_1} g''_{e_2 e_2} - (g''_{e_1 e_2})^2) \geq 0 \quad (27)$$

The qualitative properties of the solution depend on the sign of the relevant element of  $\nabla_{\theta} \mathbf{g} - \mathbf{g}_{\theta}^*$ . This vector consists of first-order productivity effects on the absolute distance between actual and ideal effort. For exogenous  $g^*$ , the first-order change in self-ideal discrepancy is simply  $g'_{\theta_k} > 0$ , and increased productivity instantaneously make individuals better off in the sense that they enjoy improved self-image even without adjusting their efforts. If, however, the ideal is endogenous, the first-order effect is  $g'_{\theta_k} - dg^*/d\theta_k$ , which may be negative; in other words, the ideal may not only rise in response to a productivity increase, but rise by *more* than the first-order effect on production,  $g'_{\theta_k}$ . If so, people are instantaneously made worse, rather than better, off. Indeed, in the welfare problem, this turns out to be our benchmark case.

**Lemma 2.** *Suppose a unique and interior solution to (22) exists. Consider a small shift in productivity parameter  $\theta_k$  at that solution and suppose that  $g$  is strictly quasiconcave, with  $g''_{e_k e_k} \leq g''_{e_1 e_2} \leq 0$  for  $k = 1, 2$ . If a solution to system (24) exists, then*

$$g'_{\theta_k} \leq \frac{dg^*}{d\theta_k}.$$

for  $k = 1, 2$ .

In analyzing the utility-maximization problem, it is important to bear in mind that its optimum is likely to be a different point in the ‘effort plane’ from the welfare optimum. Hence, there is no guarantee that Lemma 2 will still hold at that point, so we need to learn more about the magnitude of  $g'_{\theta_k}$  at the utility-maximizing point. We will now show that given our assumptions on  $g''_{e_1 e_2}$ , Lemma 2 applies to the utility-maximizing point as well.

Recall that by Lemma 1, each utility-maximizing sum of efforts  $\hat{e}$  corresponds to only a single production level  $\hat{g}$ . But because of strict quasiconcavity, each optimal production level then also corresponds to a unique optimal effort vector. Hence, we may define a vector-valued ‘conditional effort function’  $\mathbf{e}(\hat{g}(\hat{e})) = \mathbf{e}(\hat{e})$ . Moreover, if only interior solutions are allowed, casual inspection of the first-order conditions for a utility maximum shows that  $\mathbf{e}(\hat{e})$  must map on to the line  $g'_{e_1} = g'_{e_2}$  in effort space. Indeed, by Proposition 1, the segment of this line where  $g^0 \leq g \leq g^*$  traces out the set of all possible interior utility maxima.

Now, differentiate the condition  $g'_{e_1} = g'_{e_2}$  implicitly with respect to  $\hat{e}$ . That is, define (for  $k = 1, 2$ )

$$g'_{e_1}(e_k(\hat{e}), \hat{e} - e_k(\hat{e})) - g'_{e_2}(e_k(\hat{e}), \hat{e} - e_k(\hat{e})) = 0$$

and differentiate both sides with respect to  $\hat{e}$  to produce (for  $k \neq l$ )

$$\frac{de_k}{d\hat{e}} = \frac{g''_{e_1 e_1} - g''_{e_1 e_2}}{g''_{e_1 e_1} + g''_{e_2 e_2} - 2g''_{e_1 e_2}} \geq 0$$

so long as  $g''_{e_1 e_1} + g''_{e_2 e_2} - 2g''_{e_1 e_2} \neq 0$ , since  $g$  is concave and by assumption  $g''_{e_1 e_2} \geq g''_{e_1 e_1}$ . Thus, both activities are ‘normal’ in the production-maximization problem. Finally, applying the chain rule to the conditional effort function gives

$$\frac{de_k}{d\hat{g}} = \frac{\frac{de_k}{d\hat{e}}}{\frac{d\hat{g}}{d\hat{e}}}$$

By Lemma 1 we have  $d\hat{g}/d\hat{e} > 0$ , so  $de_k/d\hat{g} \geq 0$  at any point where  $g'_{e_1} = g'_{e_2}$ . It follows that, if  $\bar{\mathbf{e}} = (\bar{e}_1, \bar{e}_2)$  denotes the utility maximum, we must have  $\bar{e}_k \leq e_k^*$  for  $k = 1, 2$ . Finally, by Lemma 2 and since  $g''_{e_k \theta_k} > 0$ ,

$$g'_{\theta_k}(\bar{\mathbf{e}}) \leq g'_{\theta_k}(\mathbf{e}^*) \leq \frac{dg^*}{d\theta_k} \quad (28)$$

With these technicalities out of the way, we can go on to analyze the end result of a productivity shift on actual efforts. Our results are summarized in the following proposition.

**Proposition 4.** *Suppose a unique and interior solution to (22) exists, as well as to (23). Consider a small shift in productivity parameter  $\theta_k > 0$  at these points and suppose  $g$  is strictly quasiconcave, with  $g_{kk} \leq g_{12} \leq 0$  for  $k = 1, 2$ . If systems (24) and (26) both have a solution, then for  $l \neq k$ ,*

- A.  $\frac{de_k}{d\theta_k} > 0$ .
- B.  $\frac{de_l}{d\theta_k} \begin{matrix} \geq \\ < \end{matrix} 0$ .
- C. *production strictly increases:*  $\frac{dg}{d\theta_k} > 0$ .

Unlike Proposition 3, the above result does not guarantee crowding-out, although activity-specific policy still drives increased public-good production. In fact, we can draw no general conclusions about the sign of spillovers. The intuition is the following. Agents revise their view of the ideal contribution upwards to match productivity increases, fueling further expansion of effort. While the productivity shift itself tends to drive increased effort only on activity  $k$ , the increase in ideal contributions affects both activities, and may or may not be sufficiently pronounced to produce effort increases also for  $e_l$ , depending on complex interactions between e.g. the marginal product of  $e_1$  and  $e_2$ .

Under what conditions are predictions from the endogenous-ideal model close to those of the exogenous-ideal one? The proof of Proposition 4 reveals that the difference in spillovers

between the exogenous and the endogenous case is determined by  $dg^*/d\theta_k$ , or more precisely by

$$\frac{g'_1 (g''_{e_k e_k} - g''_{e_1 e_2}) (1 + I') I''}{|\mathbf{u}'' + (1 + I') \mathbf{H}_g + I'' \nabla_{\mathbf{e}} \mathbf{g} (\nabla_{\mathbf{e}} \mathbf{g})'|} \cdot \frac{dg^*}{d\theta_k} \quad (29)$$

where the determinant is given by (27), and

$$\frac{dg^*}{d\theta_k} = g'_{e_1} \frac{de_1^*}{d\theta_k} + g'_{e_2} \frac{de_2^*}{d\theta_k} + g'_{\theta_k} = \frac{N^2 g'_{e_1} g''_{e_k \theta_k} (g''_{e_1 e_2} - g''_{e_1 e_l})}{|\mathbf{u}'' + N \mathbf{H}_g|} + g'_{\theta_k} \quad (30)$$

with the determinant given by (25). All else equal, differences between the models grow larger (and, by (B.5), spillovers more likely to be positive) as  $dg^*/d\theta_k$  increases, but since the ideal contribution is endogenous, all else is typically *not* equal. Indeed, it turns out that there are few clear-cut patterns to be found. For example, differentiating (29) and (30) reveals only ambiguous effects of increasing  $g'_1 = g'_2$  or  $g''_{e_1 e_2}$  in the welfare as well as the utility maximum.

Nevertheless, some conclusions can be drawn. For example, (30) is increasing in  $g'_1 = g'_2$ , implying that  $g^*$  increases relatively strongly if its level, based on introspective moral reasoning, would have been low to begin with. Consider a hypothetical example. In each of two different collective households, having a clean kitchen is a public good. Cleaning can be done by vacuuming ( $e_k$ ) and/or mopping the floor ( $e_l$ ). Suppose the tenants of the first household place a higher marginal value on leisure than those of the second household, implying that initially the Kantian ideal level of cleanliness would be lower in the first household than in the second. (Actual cleaning efforts need not be lower in the first household, as these are partly determined by image motivation.) Now suppose that both households are given a more efficient model of vacuum cleaner ( $\theta_k$  increases). Where should we expect the perceived ideal to increase the most? According to (30), the answer is the first household, because in the second household, cleaning the kitchen very thoroughly is already considered ideal. As a result, if we are willing to make the *ceteris paribus* assumption, we might expect the first household to mop more when receiving the new vacuum cleaner, and the second to mop less.

More importantly, we can also conclude from (29) and (30) that the differences between the models increase (and spillovers shift towards crowding-in) as  $u''$  increases, i.e. as the utility of leisure becomes more approximately linear. The intuition is that  $-u''$ , in reflecting the marginal value of leisure time, also reflects the benefits of engaging in moral licensing. When  $\theta_k$  increases, agents will tend to increase  $e_k$  in response. The resulting drop in leisure is costlier for smaller (more negative)  $u''$ , and the temptation to let the increase in  $e_k$  license decreased  $e_l$  will be correspondingly greater. Since the pattern emerges even in the hypothetical welfare problem, the ideal contribution is more responsive to productivity shocks for large  $u''$ , further increasing the divide between spillover results in the exogenous and endogenous model. This may be one reason why the quasilinear example of Section 3 implies crowding-out when  $g^*$  is



exogenous, and crowding-in when it is endogenous.

## 6 Revisiting an earlier result: centralizing an activity

In this section, we argue that there may be several plausible ways of extending a given single-activity model to the multi-activity case, and that the mode of extension is likely to make a difference. In particular, not all generalizations need reproduce the ideas of the original model; one needs to think carefully about which multi-activity model is appropriate to a given situation.

We illustrate these points by revisiting Bruvold and Nyborg (2004), who analyze the aggregate utility consequences of centralization, i.e. of having government perform a single contributing activity directly. They use a self-image function with an exogenous production-based ideal, making their model analogous to<sup>24</sup>

$$U = u(T - e) + v(G_{-i}(\theta)) + I(g(e, \theta) - g^*) \quad (31)$$

which is utility function (2) with  $K = 1$ .

Bruvold and Nyborg (2004) interpret  $g$  as the environmental improvement from each agent's efforts on recycling household waste, and ask what would happen if that waste were to be sorted at a centralized plant rather than by households. Such a policy would obviously render individual recycling efforts superfluous; Bruvold and Nyborg (2004) argue that it would also ease the burden of having to secure a positive self-image. Thus, holding the amount of waste sorted constant across regimes, individual utility would increase. This in turn suggests that in cost-benefit analyses, household willingness to pay to avoid having to recycle should be strictly positive. The argument is that, while activities such as source separation of waste by households may appear voluntary because individuals are not, strictly speaking, obligated to engage in them, they really are not: the need to improve one's self image forms a behavioral constraint that individuals would be better off without.

The effect of centralization in Bruvold and Nyborg (2004) is to exogenously set  $g^*$  equal to  $g^0 = 0$ . Because only one contributing activity exists, doing so eliminates the need to perform any activity, and it is not difficult to see that this should raise utility. The clarity of this result,

---

<sup>24</sup>The Bruvold and Nyborg (2004) utility model is specified as

$$U = u(c, T - e, G_{-i} + g(e), I(g(e) - g^*))$$

where  $c$  is individual consumption. Bruvold and Nyborg (2004) do not consider productivity shifts, so  $\theta$  is not explicitly an argument of  $G_{-i}$  or  $g$ . Compared to the original model, equation (31) imposes linear separability, abstracts from consumption  $c$ , and assumes that individuals disregard their own contribution  $g(e)$ , except in the self-image term  $I$ . None of these changes alter the central conclusions of the original paper.

however, relies on the implicit assumption that the single centralized activity can be analyzed separately from other activities, e.g. other environmentally friendly acts. This assumption is not necessarily present in multi-activity models, where at least one non-centralized activity will be included. To make this point more clearly, suppose  $N$  identical agents maximize

$$U(e_1, e_2) = u(T - e_1 - e_2) + v(G_{-i}(\theta_1, \theta_2)) + I(g(e_1, e_2, \theta_1, \theta_2) - g^*) \quad (32)$$

From a purely technical standpoint, this may seem like a straightforward generalization of model (31); indeed, it is model (2) with  $K = 2$ . Yet (32) does not examine the centralized activity in isolation, and therefore raises new issues that are not apparent in the single-activity case.

First, it is no longer clear whether or how  $g^*$  should change as a result of centralization. Certainly, as the ideal is based on overall production  $g(e_1, e_2, \cdot)$ , it seems unreasonable to set it to zero if only one of possibly many activities is centralized. An exogenous  $g^*$  may well reflect concepts that are independent of the number of available activities: in the example of Section 3, ideal per-capita emissions were derived from the atmospheric CO<sub>2</sub> concentration considered safe, which does not change if an activity is centralized.<sup>25</sup> In what follows, therefore, we will set the relevant productivity parameter  $\theta_k$  to zero, while  $g^*$  will remain fixed at its initial level. Supposing that solutions to (32) are unique, we denote an identical individual's optimal allocations before and after centralization by  $\mathbf{e}^0 = (e_1^0, e_2^0)$  and  $\mathbf{e}^1 = (e_1^1, e_2^1)$ , respectively. When activity  $k$  is fully centralized, agents subsequently choose the boundary solution  $e_k^1 = 0$  for all agents. To see why, note that  $\theta_k^1 = 0$  implies  $g'_{e_k} = 0$  everywhere. The left-hand side of the  $k$ th Kuhn-Tucker condition (3) for maximizing (32) then becomes  $-u' - \lambda < 0$ , which is incompatible with  $e_k^1 > 0$ .<sup>26</sup>

Second, in Bruvoll and Nyborg (2004), full centralization is taken to mean that total production  $G = Ng(\cdot)$  is held fixed at its initial level. While a reasonable approach in a single-activity setting, holding overall production constant is not really possible here; agents will, for instance, increase  $e_l$  ( $l \neq k$ ) in response to activity  $k$  being centralized if Proposition 3 applies. It seems rather more natural to fix efforts. Let centralization be equivalent to all agents not only making efforts  $e_l^1$  (as they actually do), but also  $e_k^0$  at initial productivity  $\theta_k^0$ .

---

<sup>25</sup>If, like Bruvoll and Nyborg (2004), we interpret the public good as 'environmental quality', another example is the concept of an individual ecological footprint (Wackernagel and Rees, 1996).

<sup>26</sup>A reviewer suggested that centralization should not be understood as  $\theta_k = 0$ , but rather as making recycling infinitely easy, since waste is then sorted without individual effort. Indeed, we noted in the single-activity example of section 3 that effort also approaches zero when productivity becomes very large. In response, we would argue that  $\theta_k$  is associated with each agent's *own* productivity; our interpretation is that centralization makes recycling morally irrelevant ( $g^* = 0$ ) in the Bruvoll and Nyborg (2004) model precisely because agents are rendered unable to directly affect the share of sorted waste at the margin (so  $\theta_k = 0$ ).

Under the notational convention  $g = g(e_k, e_l; \theta_k)$ , the ex-post production of others will then be  $G_{-i} = (N - 1)g(e_k^0, e_l^1; \theta_k^0)$ .

Third, centralization effectively guarantees that some level of effort will be provided on the targeted activity. Disregarding additional tax payments, agents may view that guarantee as an effort-free windfall to their self-image, making them better off; then again, they might not. If windfalls are fully image-relevant, the production level fed into the self-image function should be  $g(e_k^0, e_l^1; \theta_k^0)$ , which is a single identical agent's share of total ex-post production. Then the conclusion of Bruvold and Nyborg (2004) — that centralization increases each agent's utility — is unsurprisingly borne out. We follow the original treatment in excluding public-sector costs of performing the centralized activity.

**Proposition 5.** *Suppose that activity  $k$  is fully centralized. Ex-post production is given by  $G_{-i} = (N - 1)g(e_k^0, e_l^1; \theta_k^0)$ , and is fully counted as a windfall, so  $I = I(g(e_k^0, e_l^1; \theta_k^0) - g^*)$  for all agents. Then  $U(0, e_l^1) \geq U(e_k^0, e_l^0)$ .*

However, it can also be argued that windfalls should *not* be image-relevant. Recall that  $g$  is a personal production function. In truth, agents only put in  $(e_k, e_l) = (0, e_l^1)$ , while facing productivity  $\theta_k = 0$ . If they are impure altruists in the sense of Andreoni (1990), they will be especially concerned with the effects of their own actions, and may be unimpressed by outcomes that they have not ‘earned’ themselves. If so, we should feed  $g(0, e_l^1; 0)$  into the self-image function.<sup>27</sup>

In this case, centralization has an ambiguous effect, which is especially illuminating to examine in incremental terms. Suppose that regardless of how much effort agents put in on activity  $k$ , public policy adds just enough to keep effort constant at  $e_k^0$  in  $G_{-i} = (N - 1)g(e_k, e_l; \theta_k^0)$ .  $\theta_k$  is the productivity faced solely by individuals and can be interpreted as the degree of centralization, as overall effort on activity  $k$  is always defined by  $e_k^0$  and  $\theta_k^0$ . If windfalls are not image-relevant, the resulting utility function is formally identical to (32),

---

<sup>27</sup>The windfall issue is also recognized by Johansson-Stenman and Svedsäter (2012), who lean toward this position. The image-relevance of windfalls likely depends on the framing adopted by agents. If an agent views production in terms of individual ‘shares’, then her share remains her own even if it is centralized, and what is done in her name may spill over onto her personal self-image. This might be the case, for example, when the CO<sub>2</sub> problem is expressed in terms of individual carbon footprints. Other frames may not be share-oriented, such as the view that pro-environmental acts are a way, not to ease personal culpability, but to support a societal transition toward ‘sustainability’. More broadly, we consider windfalls likely to be disregarded whenever  $g^*$  is framed as improvements *by* some constant amount rather than improvements *up to* some amount. This is similar to warm-glow models, where doing more is always desirable; indeed, provided windfalls are image-irrelevant, the results to follow apply to increasing image functions  $I(g(\cdot))$  as well.

and at an interior solution the envelope theorem gives

$$\frac{dU}{d\theta_k} = v' \frac{dG_{-i}}{d\theta_k} + I' g'_{\theta_k} = (N-1)v' g'_{e_l} \frac{de_l}{d\theta_k} + I' g'_{\theta_k} \quad (33)$$

If Proposition 3 holds, the sign of this expression is indeterminate. Thus, centralization (lowering  $\theta_k$ ) may leave agents worse off. This is due to a trade-off. On the one hand,  $e_k^0$  is now guaranteed to be supplied at productivity  $\theta_k^0$ , leaving individuals free to substitute to other activities, increasing total production. On the other hand, people are worse off because  $g^*$  is unchanged and centralized efforts do nothing to improve their self-image.<sup>28</sup>

In this setting, then, a second-best problem may arise. While all else equal it would increase aggregate utility to relieve agents of their moral motivation to contribute, such an objective is beyond reach, and incremental steps toward it — by centralizing a single activity — may lower aggregate utility instead.<sup>29</sup>

In conclusion, we have shown that extending the Bruvold and Nyborg (2004) framework to model (32) introduces some additional issues, and moreover accommodates the idea that centralization may not entail a utility gain. More broadly, however, it is a simple matter to formulate a different multidimensional model which clearly reproduces both the intuitions and the results of Bruvold and Nyborg (2004). Consider a plausible alternative to ‘self-image’ function (1), namely

$$S(\mathbf{e}) = I(e_1 - e_1^*, \dots, e_K - e_K^*) \quad (34)$$

where for  $k = 1, \dots, K$ ,  $e_k^* \geq 0$  are *activity-specific* ideal contributions. In most other ways the self-image function is similar to (1): it is twice continuously differentiable and maximized when, for all  $k$ ,  $e_k = e_k^*$ . We have  $I'_k > 0$  when  $e_k < e_k^*$ ;  $I'_k = 0$  when  $e_k = e_k^*$ ;  $I'_k \leq 0$  when  $e_k > e_k^*$ . We could also assume that  $I$  is strictly concave in effort.

---

<sup>28</sup>If we allow the ideal contribution to be endogenously determined as in section 5, the counterpart to (33) is

$$\frac{dU}{d\theta_k} = (N-1)v' g'_{e_l} \frac{de_l}{d\theta_k} + I' \left( g'_{\theta_k} - \frac{dg^*}{d\theta_k} \right)$$

where condition (28) implies that the second term is negative and Proposition 4 implies that the first term may be positive. Thus, compared to the case of exogenous ideals, both terms may switch sign, and it remains unclear whether a given effect necessarily dominates.

<sup>29</sup>As this suggests that some agents may be willing to pay to avoid decreasing individual responsibility for recycling, it may also be premature to conclude that the aggregate opportunity cost of time spent recycling must be strictly positive. Several survey studies (Bartelings and Sterner, 1999; Bruvold et al., 2002; Berglund, 2006) estimate this cost by asking whether people would prefer waste sorting to be done by households or by others. For example, in Bruvold et al. (2002) 27% of respondents stated a preference for sorting the waste themselves, while 72% did not. In all studies, however, subsequent willingness-to-pay questions were phrased in a way that ruled out negative WTP values; we suggest it may be worthwhile to drop that restriction.

An obvious difference between this model and model (1) lies in whether best efforts are good enough. Function (1) is a consequentialist’s model: people ask, ‘Am I making enough of a difference?’ Only productive results are compared, so it is enough to approach any point in  $\mathbb{R}^K$  which lies on the level curve  $g = g^*$ . By comparison, (34) is a deontological image function, where the question is, ‘Am I trying hard enough?’ Here, to improve one’s self-image is to approach a single point in  $\mathbb{R}^K$ .<sup>30</sup>

More importantly, the models can be thought of in terms of different mental accounting strategies. Read et al. (1999) document a range of broad versus narrow modes of ‘choice partitioning’, including ‘choice bracketing’ (whether a sequence of choices are made jointly or piecemeal) and ‘joint/separate evaluation’ (whether each alternative is judged solely on its own merits, or actively compared with other options). Differences between the two multi-dimensional self-image functions arguably capture something of the flavor of both, since contributions are often made as separate choices (joint/separate evaluation) and at different points in time (choice bracketing). Interpreting the models in these terms, function (1) imposes broad partitioning, while (34) assumes narrow partitioning.

Read et al. (1999) present a great deal of evidence documenting the often striking difference that different modes of choice partitioning make for behavior. Similarly, the ‘separate’ model (34) does much better at replicating the arguments of Bruvold and Nyborg (2004) than the ‘joint’ model (32). In the latter, it is not obvious that setting  $g^* = 0$  is appropriate when only activity  $k$  is being centralized, but we have no objection to setting  $e_k^* = 0$  within the former. The fit with previous results is better because narrow partitioning is imposed by construction within a one-dimensional model that is interpreted in terms of recycling or some other single non-composite activity. Whether the self-image function is production-based (e.g. Bruvold and Nyborg, 2004) or effort-based (Brekke et al., 2003), there exists only one activity, so the ideal will be general to all activities by definition, while comparisons between activities are omitted from the analysis.

As a result, the intuitions of most papers in the ‘duty-orientation’ literature differ from those that underlie equation (33). These papers often stress that moral responsibility, including for single non-aggregate activities, is a burden which individuals would prefer not to bear (Bruvold and Nyborg, 2004; Brekke et al., 2010). For example, Nyborg (2011) argues that a single-activity version of the self-image model is consistent with evidence (Dana et al., 2007)

---

<sup>30</sup>An experiment by Cornelissen et al. (2013) suggests that people’s ethical mind-set (consequentialist versus rule-based) may make a difference for the direction of spillovers. Subjects that were primed to think of decisions in terms of outcomes exhibited negative spillovers, while those primed to think in terms of rules exhibited positive spillovers. We have shown in this paper that spillovers tend to be negative for the consequentialist model (1); showing that (34) leads to positive spillovers requires additional assumptions on  $I$ .

that when people are offered the chance to remain ignorant of how their choices affects the other player’s payoff in a binary dictator game, many will choose to do so and then pick the selfish option. In equation (33), by contrast, an overarching motivation to contribute is taken for granted; while individuals would be better off without that general feeling of moral responsibility, they need not feel the same about any given activity. In effect, agents ask: ‘Given that (unfortunately) I ought to make this much of a difference, how should I allocate effort across activities?’ Here, to the extent that activities expand the choice set for approaching the overall ideal, they make agents better off, not worse.

As for which approach is the right one, model (1) does have the somewhat attractive property that substitutability across activities derives from the public-good production technology, which may be observable. By contrast, in a model based on (34), the sign of spillovers depends on mixed-partial derivatives of  $I(e_1 - e_1^*, \dots, e_K - e_K^*)$  that are difficult to interpret. Nevertheless, the competing ideas presented in this section should be tested: is moral behavior best described as consistent or as piecemeal, with each moral act considered in isolation? While it is clear that people often do engage in narrow partitioning, there is also evidence that efforts on different activities may be driven by the same general motivation (Thøgersen and Ölander, 2006). In our view, careful research on this question is warranted.

## 7 Concluding remarks

This paper has extended variants of the Brekke et al. (2003) duty-orientation model to the case of several contributing activities. This extension has been fruitful in two respects. First, we have used the multi-activity model to analyze productivity-led spillover effects from a policy which targets a particular prosocial activity. Here we have considered two cases: one where the ‘ideal contribution’ considered by agents is exogenous, and one where it is endogenously determined through moral introspection. In the former, our results point to crowding-out of effort on other activities, similar to the ‘moral licensing effect’(Blanken et al., 2015; Jacobsen et al., 2012; Tiefenbeck et al., 2013), as the benchmark case. In the latter, the model fails to provide a clear prediction on the sign of spillovers.

Second, we have argued that the manner of generalization to multiple activities matters: a given single-activity model may (but need not) be extended in such a way that previous results and intuitions are lost. In the setting we consider, results depend crucially on whether the self-image value of contributing to a public good is considered separately for each prosocial activity, or (as in our model) is aggregated across all activities. In our view, this dichotomy is important, but tends not to be readily apparent in single-activity settings.

With respect to both of the above types of results, we consider the present paper only a starting point for further research. In particular, we would like to see models emphasizing

social and strategic interaction (such as that of Bénabou and Tirole, 2006) extended to the case of two or more activities. It is likely that this would produce new results and intuitions with respect both to spillovers and direct effects. Our hope is that this paper has at least demonstrated the usefulness of explicitly considering non-targeted activities in research on policies that aim to encourage prosocial behavior.

## References

- G.A. Akerlof and R.E. Kranton. Economics and identity. *The Quarterly Journal of Economics*, 115(3), 2000.
- J. Andreoni. Impure Altruism and Donations to Public Goods: A Theory of Warm-glow Giving. *The Economic Journal*, 100(401), 1990.
- J. Andreoni and B.D. Bernheim. Social Image and the 50-50 Norm: A Theoretical and Experimental Analysis of Audience Effects. *Econometrica*, 77(5), 2009.
- J. Andreoni, W.G. Gale, and J.K. Scholz. Charitable Contributions of Time and Money, 1996. Working paper.
- H. Bartelings and T. Sterner. Household Waste Management in a Swedish Municipality: Determinants of Waste Disposal, Recycling and Composting. *Environmental and Resource Economics*, 13(4), 1999.
- D.J. Bem. Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological Review*, 74(3), 1967.
- R. Bénabou and J. Tirole. Incentives and Prosocial Behavior. *American Economic Review*, 95(5), 2006.
- C. Berglund. The assessment of households' recycling costs: The role of personal motives. *Ecological Economics*, 56(4), 2006.
- I. Blanken, N. van de Ven, and M. Zeelenberg. A Meta-Analytic Review of Moral Licensing. *Personality and Social Psychology Bulletin*, 41(4), 2015.
- S. Bowles. Policies Designed for Self-Interested Citizens May Undermine 'The Moral Sentiments': Evidence from Economic Experiments. *Science*, 320(5883), 2008.
- K.A. Brekke and K. Nyborg. Attracting responsible employees: Green production as labor market screening. *Resource and Energy Economics*, 30(4), 2008.

- K.A. Brekke, S. Kverndokk, and K. Nyborg. An economic model of moral motivation. *Journal of Public Economics*, 87(9), 2003.
- K.A. Brekke, G. Kipperberg, and K. Nyborg. Social Interaction in Responsibility Ascription: The Case of Household Recycling. *Land Economics*, 86(4), 2010.
- A.L. Brown, J. Meer, and J.F. Williams. Why Do People Volunteer? An Experimental Analysis of Preferences for Time Donations. 2016. NBER Working Paper No. 19066.
- A. Bruvold and K. Nyborg. The Cold Shiver of Not Giving Enough: On the Social Cost of Recycling Campaigns. *Land Economics*, 80(4), 2004.
- A. Bruvold, B. Halvorsen, and K. Nyborg. Households' recycling efforts. *Resources, Conservation and Recycling*, 2002.
- L. Burke, K. Reyntar, M. Spalding, and A. Perry. Reefs At Risk Revisited. Technical report, World Resources Institute, 2011.
- D.M. Cain, G. Loewenstein, and D.A. Moore. The Dirt on Coming Clean: Perverse Effects of Disclosing Conflicts of Interest. *The Journal of Legal Studies*, 34(1), 2005.
- J. Cairns and R. Slonim. Substitution effects across charitable donations. *Economics Letters*, (2), 2011.
- F. Carlsson, M.M. Jaime Torres, and C.I. Villegas. Spillover Effects from a Social Information Campaign. 2016. Environment for Development Discussion Paper No. 16-06.
- G. Cornelissen, M.R. Bashshur, J. Rode, and M. Le Menestrel. Rules or Consequences? the Role of Ethical Mind-Sets in Moral Dynamics. *Psychological Science*, 24(4), 2013.
- J. Dana, R.A. Weber, and J.X. Kuang. Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory*, 33(1), 2007.
- P. Dolan and M.M. Galizzi. Like ripples on a pond: Behavioral spillovers and their implications for research and policy. *Journal of Economic Psychology*, 47, 2015.
- C. Ek. Some Causes are More Equal than Others? The Effect of Similarity on Substitution in Charitable Giving. *Journal of Economic Behavior & Organization*, 136, 2017.
- N.E. Feldman. Time is Money: Choosing between Charitable Activities. *American Economic Journal: Economic Policy*, 2(1), 2010.
- L. Festinger. *A theory of cognitive dissonance*. Stanford University Press, 1957.



- E. Filiz-Ozbay and N. Uler. Demand for Giving to Multiple Charities: Theory and Experiments. 2016. Working paper.
- B.S. Frey and R. Jegen. Motivation Crowding Theory. *Journal of Economic Surveys*, 15(5), 2001.
- P.W. Henderson and R.A. Peterson. Mental Accounting and Categorization. *Organizational Behavior and Human Decision Processes*, 51(1), 1992.
- E.T. Higgins. Self-Discrepancy: A Theory Relating Self and Affect. *Psychological Review*, 94(3), 1987.
- IPCC. *Climate Change 2013: The Physical Science Basis*. Cambridge University Press, New York, NY, USA, 2013.
- G.D. Jacobsen, M.J. Kotchen, and M.P. Vandenbergh. The behavioral response to voluntary provision of an environmental public good: Evidence from residential electricity demand. *European Economic Review*, 56(5), 2012.
- O. Johansson-Stenman and H. Svedsäter. Self-image and valuation of moral goods: Stated versus actual willingness to pay. *Journal of Economic Behavior & Organization*, 84(3), 2012.
- N. Lacetera, M. Macis, and R. Slonim. Will There Be Blood? Incentives and Displacement Effects in Pro-Social Behavior. *American Economic Journal: Economic Policy*, 4(1), 2012.
- A. Lilley and R. Slonim. The price of warm glow. *Journal of Public Economics*, 114, 2014.
- J. Meer. Does Fundraising Create New Giving? 2016. Working paper.
- A.C. Merritt, D.A. Effron, and B. Monin. Moral Self-Licensing: When Being Good Frees Us to Be Bad. *Social and Personality Psychology Compass*, 4(5), 2010.
- P. Milgrom and C. Shannon. Monotone Comparative Statics. *Econometrica*, 62(1), 1994.
- C. Null. Warm glow, information, and inefficient charitable giving. *Journal of Public Economics*, 95(5-6), 2011.
- K. Nyborg. I don't want to hear about it: Rational ignorance among duty-oriented consumers. *Journal of Economic Behavior & Organization*, 79(3), 2011.
- K. Nyborg and M. Rege. Does public policy crowd out private contributions to public goods? *Public Choice*, 115(3), 2003.

- K. Nyborg, R.B. Howarth, and K.A. Brekke. Green consumers and public policy: On socially contingent moral motivation. *Resource and Energy Economics*, 28(4), 2006.
- M. Ploner and T. Regner. Self-image and moral balancing: An experimental analysis. *Journal of Economic Behavior & Organization*, 93, 2013.
- D. Read, G. Loewenstein, and M. Rabin. Choice Bracketing. *Journal of Risk and Uncertainty*, 19(1-3), 1999.
- D. Reinstein. Does One Charitable Contribution Come at the Expense of Another? *The B.E. Journal of Economic Analysis & Policy*, 11(1), 2011.
- D.C. Ribar and M.O. Wilhelm. Altruistic and Joy-of-Giving Motivations in Charitable Behavior. *Journal of Political Economy*, 110(2), 2002.
- J.J. Rotemberg. Charitable giving when altruism and similarity are linked. *Journal of Public Economics*, 114, 2014.
- S. Sachdeva, R. Iliev, and D.L. Medin. Sinning Saints and Saintly Sinners. *Psychological Science*, 20(4), 2009.
- H. Tajfel and J.C. Turner. An integrative theory of intergroup conflict. In W.G. Austin and S. Worchel, editors, *The social psychology of intergroup relations*. 1979.
- R.H. Thaler. Mental Accounting Matters. *Journal of Behavioral Decision Making*, 12(3), 1999.
- J. Thøgersen and F. Ölander. To what degree are environmentally beneficial choices reflective of a general conservation stance? *Environment and Behavior*, 38(4), 2006.
- V. Tiefenbeck, T. Staake, K. Roth, and O. Sachs. For better or for worse? Empirical evidence of moral licensing in a behavioral energy conservation campaign. *Energy Policy*, 57, 2013.
- M. Wackernagel and W. Rees. *Our Ecological Footprint: Reducing Human Impact on the Earth*. New Society Publishers, Gabriola Island, BC, Canada, 1996.

# Appendices

## A Monotone comparative statics and total efforts

Here we derive sufficient conditions for optimal total efforts to be monotone nondecreasing in  $\theta_k$  and  $g^*$ . Our starting point is the problem (19). The constraint set of this problem is simply a constant interval on the real line, hence a sublattice. Let us therefore proceed to apply Proposition 2. Using differentiability of  $\hat{g}$  produces the sufficient single-crossing condition

$$\frac{\partial^2 U_{\hat{e}}}{\partial \hat{e} \partial \theta_k} = I'' \left( \frac{\partial \hat{g}}{\partial \theta_k} \frac{\partial \hat{g}}{\partial \hat{e}} \right) + I' \frac{\partial}{\partial \theta_k} \left( \frac{\partial \hat{g}}{\partial \hat{e}} \right) \geq 0 \quad (\text{A.1})$$

Now, note that both  $\hat{e}$  and  $\theta_k$  are parameters in the PMP. Thus, we may use an envelope theorem to differentiate  $\hat{g}$ . To make its application feasible, it is enough to suppose that the initial solution to the PMP is unique, with  $e_k > 0$ . The PMP is formally equivalent to the standard utility-maximization problem with an equality constraint and a price vector of ones; thus, as usual, strict quasiconcavity (which follows from strict concavity) ensures uniqueness. Applying the envelope theorem produces  $\partial \hat{g} / \partial \theta_k = g'_{\theta_k}$  and  $\partial \hat{g} / \partial \hat{e} = \lambda$ , where  $\lambda$  is the Lagrange multiplier associated with the restriction  $e_1 + e_2 - \hat{e} = 0$ . If the initial solution to (19) had  $e_k > 0$ , the corresponding necessary Kuhn-Tucker conditions for the PMP imply  $\lambda = g'_{e_k}$ .<sup>31</sup> When differentiating a second time, we can no longer ignore effects through effort variables, so

$$\frac{\partial}{\partial \theta_k} \left( \frac{\partial \hat{g}}{\partial \hat{e}} \right) = \frac{\partial}{\partial \theta_k} (g'_{e_k}(e_k(\theta_k), e_l(\theta_k), \theta_k)) = g''_{e_k e_k} \frac{de_k}{d\theta_k} + g''_{e_k e_l} \frac{de_l}{d\theta_k} + g''_{e_k \theta_k}$$

where  $de_k/d\theta_k$  and  $de_l/d\theta_k$  are shifts in optimal efforts within the PMP, i.e. while holding  $\hat{e}$  constant. For this reason,  $de_l/d\theta_k = -de_k/d\theta_k$ , and condition (A.1) is transformed into

$$\frac{\partial^2 U_{\hat{e}}}{\partial \hat{e} \partial \theta_k} = I'' g'_{e_k} g'_{\theta_k} + I' \left( (g''_{e_k e_k} - g''_{e_k e_l}) \frac{de_k}{d\theta_k} + g''_{e_k \theta_k} \right) \geq 0$$

Applying monotone comparative statics methods to the PMP, we see that  $de_k/d\theta_k$  will be nonnegative in the above expression whenever  $g''_{e_k e_l} \leq 0$  (as  $g''_{e_k \theta_k} > 0$  and  $g''_{e_l \theta_k} = 0$ ). If also  $g''_{e_k e_k} < g''_{e_k e_l}$ , the sufficient condition for total efforts to increase with productivity ends up being more restrictive than condition (15), which guarantees that direct effects from policy are positive. This is the case, for example, in the simple example of Section 3, where  $g''_{e_k e_k} < g''_{e_k e_l} = 0$ .

---

<sup>31</sup>The constraint qualification required for the Kuhn-Tucker conditions to be necessary holds, as, since  $e_k > 0$  by assumption, binding constraints in the PMP are  $e_1 + e_2 = \hat{e}$  and, possibly,  $e_l = 0$ . Even if both are active, the associated gradient vectors are linearly independent.

As for the relationship between  $g^*$  and total efforts, we obtain the following from (19):

$$\frac{\partial^2 \hat{U}}{\partial \hat{e} \partial g^*} = -I'' \frac{\partial \hat{g}}{\partial \hat{e}} = -I'' \lambda \geq 0$$

so the set of total efforts is then monotone nondecreasing in  $g^*$ .

## B Mathematical proofs

### Proof of Proposition 1

*Proof.*  $g(\mathbf{e}) \geq g^0$  by definition. The proof that  $g(\mathbf{e}) \leq g^*$  is by contradiction. Suppose  $g(\mathbf{e}) > g^*$ . Since  $g(\mathbf{e}) = g^0$  otherwise, there must exist some  $k$  for which optimal  $e_k > 0$ , implying that the corresponding Kuhn-Tucker condition (3) is satisfied with equality. But this is a contradiction as, for  $g(\mathbf{e}) > g^*$ , we have  $-u' + g'_{e_k} I' - \lambda < 0$ . Hence  $\mathbf{e}$  cannot be an optimum. If  $g^* > g^0$ , we suppose instead that  $g(\mathbf{e}) \geq g^*$ . The rest of the argument is identical.  $\square$

### Proof of Lemma 1

*Proof.* If  $\hat{e} = 0$ , the result is trivial. For  $\hat{e} > 0$ , suppose  $\mathbf{e}' = (e'_1, \dots, e'_K)$  solves (2) but that  $g(\mathbf{e}')$  does not satisfy (18) at effort level  $\hat{e}' = \sum_k e'_k$ . Because Proposition 1 implies  $g^* > g(\mathbf{e}')$  and  $g$  is continuous, there must then exist some other vector  $\mathbf{e}''$  with  $\hat{e}'' = \sum_k e''_k = \hat{e}'$  and  $g^* \geq g(\mathbf{e}'') > g(\mathbf{e}')$ , implying

$$u(T - \hat{e}') + v(G_{-i}) + I(g(\mathbf{e}') - g^*) < u(T - \hat{e}'') + v(G_{-i}) + I(g(\mathbf{e}'') - g^*)$$

so  $\mathbf{e}'$  cannot solve (2).

For strict increasingness, consider vectors  $\mathbf{e}' \neq \mathbf{e}''$ , with  $\hat{e}'' > \hat{e}'$  but  $g(\mathbf{e}'') \leq g(\mathbf{e}')$ . Also, suppose that both vectors satisfy (18) for their respective effort levels. Now, consider the effort vector  $(e'_1 + \hat{e}'' - \hat{e}', e'_2, \dots, e'_K)$ . Its sum of efforts is  $e'_1 + \hat{e}'' - \hat{e}' + e'_2 + \dots + e'_K = \hat{e}''$ , yet if  $g$  is strictly increasing,

$$g(e'_1 + \hat{e}'' - \hat{e}', e'_2, \dots, e'_K) > g(\mathbf{e}') \geq g(\mathbf{e}'')$$

so  $\mathbf{e}''$  cannot satisfy (18) after all.  $\square$

## Proof of Proposition 5

*Proof.* Note that

$$\begin{aligned} U(0, e_l^1) - U(e_k^0, e_l^0) &= u(T - e_l^1) + (N - 1)g(e_k^0, e_l^1; \theta_k^0) + I(g(e_k^0, e_l^1; \theta_k^0) - g^*) \\ &\quad - u(T - e_k^0 - e_l^0) - (N - 1)g(e_k^0, e_l^0; \theta_k^0) + I(g(e_k^0, e_l^0; \theta_k^0) - g^*) \end{aligned} \quad (\text{B.1})$$

We claim that any agent will have  $e_l^0 \leq e_l^1 \leq e_k^0 + e_l^0$ ; then, because both  $g$  and  $I$  are increasing functions, (B.1) is clearly nonnegative. The proof of  $e_l^0 \leq e_l^1 \leq e_k^0 + e_l^0$  is by contradiction.

First, to show that  $e_l^1 \leq e_k^0 + e_l^0$ , suppose instead that  $e_l^1 > e_k^0 + e_l^0 \geq 0$ . This is possible only if  $e_k^0 + e_l^0 < T$ . Recall from the discussion of Kuhn-Tucker conditions (3) that  $\lambda \geq 0$ , with  $\lambda = 0$  if  $e_k + e_l < T$  (complementary slackness). Thus, if  $e_l^1 = e_k^0 + e_l^0 < T$ , ex-ante and ex-post Lagrange multipliers (denoted  $\lambda^0, \lambda^1$ ) are both equal to zero, and if  $e_l^1 = T$ , we have  $\lambda^0 = 0$  but  $\lambda^1 \geq 0$ ; hence, in either case,  $\lambda^1 \geq \lambda^0$ . Then we have

$$\begin{aligned} I'(g(e_k^0, e_l^1; \theta_k^0) - g^*)g'_{e_l}(e_k^0, e_l^1) - \lambda^1 &< I'(g(e_k^0, e_l^0; \theta_k^0) - g^*)g'_{e_l}(e_k^0, e_l^0) - \lambda^0 \\ &\leq u'(T - e_k^0 - e_l^0) \leq u'(T - e_l^1) \end{aligned}$$

where the first (strict) inequality is due to strict concavity of  $g$ , and the second is the Kuhn-Tucker condition number  $l$  in the ex-ante problem. By Kuhn-Tucker condition number  $l$  in the ex-post problem, this chain of inequalities implies that  $e_l^1 = 0$ , so the ex-post solution cannot have  $e_l^1 > e_k^0 + e_l^0$ .

Next, to show that  $e_l^0 \leq e_l^1$ , suppose instead that  $e_l^0 > e_l^1 \geq 0$ . Then, regardless if  $e_k^0 + e_l^0 = T$  or not, we must certainly have  $e_l^1 < T$ , so  $\lambda^1 \leq \lambda^0$ . But then, similarly to above, we have

$$\begin{aligned} I'(g(e_k^0, e_l^1; \theta_k^0) - g^*)g'_{e_l}(e_k^0, e_l^1) - \lambda^1 &> I'(g(e_k^0, e_l^0; \theta_k^0) - g^*)g'_{e_l}(e_k^0, e_l^0) - \lambda^0 \\ &= u'(T - e_k^0 - e_l^0) \geq u'(T - e_l^1) \end{aligned}$$

so the ex-post utility maximum cannot have  $e_l^1 < e_l^0$ .  $\square$

## Proof of Lemma 2

*Proof.* The solution to system (24), if it exists, is

$$\mathbf{e}_\theta^* = -N(\mathbf{u}'' + N\mathbf{H}_g)^{-1} \mathbf{g}_{e\theta}'' \quad (\text{B.2})$$

For  $k = 1, 2$  and  $l \neq k$ , we use this to calculate

$$\frac{dg^*}{d\theta_k} - g'_{\theta_k} = g'_{e_1} \frac{de_1^*}{d\theta_k} + g'_{e_2} \frac{de_2^*}{d\theta_k} = \frac{N^2 g'_{e_1} g''_{e_k \theta_k} (g''_{e_1 e_2} - g''_{e_l e_l})}{|\mathbf{u}'' + N\mathbf{H}_g|}$$

where the second equality uses the first-order conditions for an interior optimum. This expression is non-negative for  $g''_{e_1 e_2} \geq g''_{e_l e_l}$ .  $\square$

## Proof of Proposition 4

*Proof.* The solution to system (26), if it exists, is

$$\mathbf{e}_\theta = - \left( \mathbf{u}'' + (1 + I') \mathbf{H}_g + I'' \nabla_{\mathbf{e}g} (\nabla_{\mathbf{e}g})' \right)^{-1} \left( (1 + I') \mathbf{g}_{e_\theta}'' + I'' \nabla_{\mathbf{e}g} (\nabla_{\theta} g - \mathbf{g}_\theta^*)' \right) \quad (\text{B.3})$$

For part A, we use (B.3) to find

$$\frac{de_k}{d\theta_k} = \frac{(1 + I') \left( g_{e_k \theta_k}'' \left( -u'' - g_{e_1 e_1}'' (1 + I') - (g_{e_1}')^2 I'' \right) + g_{e_1}' \left( \frac{dg^*}{d\theta_k} - g_{\theta_k}' \right) (g_{e_1 e_1}'' - g_{e_1 e_2}'') I'' \right)}{|\mathbf{u}'' + (1 + I') \mathbf{H}_g + I'' \nabla_{\mathbf{e}g} (\nabla_{\mathbf{e}g})'|} > 0 \quad (\text{B.4})$$

because  $g_{e_1 e_2}'' \geq g_{e_1 e_1}''$ .

For part B, we use (B.3) to also calculate

$$\frac{de_l}{d\theta_k} = \frac{(1 + I') \left( g_{e_k \theta_k}'' \left( u'' + g_{e_1 e_2}'' (1 + I') + (g_{e_1}')^2 I'' \right) + g_{e_1}' \left( \frac{dg^*}{d\theta_k} - g_{\theta_k}' \right) (g_{e_k e_k}'' - g_{e_1 e_2}'') I'' \right)}{|\mathbf{u}'' + (1 + I') \mathbf{H}_g + I'' \nabla_{\mathbf{e}g} (\nabla_{\mathbf{e}g})'|} \leq 0 \quad (\text{B.5})$$

Finally, for part C, since the FOC implies  $g_{e_1}' = g_{e_2}'$ , total differentiation at an interior utility maximum produces

$$\frac{dg}{d\theta_k} = g_{e_1}' \left( \frac{de_1}{d\theta_k} + \frac{de_2}{d\theta_k} \right) + g_{\theta_k}'$$

but using expressions (B.4) and (B.5) we find

$$\frac{de_1}{d\theta_k} + \frac{de_2}{d\theta_k} = \frac{(1 + I') \left( g_{e_k \theta_k}'' (g_{e_1 e_2}'' - g_{e_1 e_1}'') (1 + I') + g_{e_1}' \left( \frac{dg^*}{d\theta_k} - g_{\theta_k}' \right) (g_{e_1 e_1}'' + g_{e_2 e_2}'' - 2g_{e_1 e_2}'') I'' \right)}{|\mathbf{u}'' + (1 + I') \mathbf{H}_g + I'' \nabla_{\mathbf{e}g} (\nabla_{\mathbf{e}g})'|} \geq 0$$

because  $g$  is concave. Hence,  $\frac{dg}{d\theta_k} > 0$ .  $\square$