Behavioral Spillovers across Prosocial Alternatives

Ek, Claes

2016

Document Version:
Publisher's PDF, also known as Version of record

Link to publication

Citation for published version (APA):
Ek, C. (2016). *Behavioral Spillovers across Prosocial Alternatives*. [Doctoral Thesis (compilation), Department of Economics].

Total number of authors:
1

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Behavioral Spillovers across Prosocial Alternatives

Claes Ek

DOCTORAL DISSERTATION
by due permission of the School of Economics and Management, Lund University, Sweden.

To be defended at the Holger Crafoord Centre, EC3:210, Lund, on Wednesday September 21, 2016 at 14:15.

Faculty opponent
Olof Johansson-Stenman, University of Gothenburg

Organization **LUND UNIVERSITY** Department of Economics Box 7082 SE–220 07 LUND Sweden	Document name **DOCTORAL DISSERTATION**
	Date of disputation 2016-09-21
	Sponsoring organization
Author(s) Claes Ek	

Title
Behavioral Spillovers across Prosocial Alternatives

Abstract
This thesis contributes to the economic literature on prosocial behavior. It includes three papers, all of which relate to the issue of policy-driven spillovers across prosocial alternatives. For example, the common fundraising practice of donation matching could affect contributions through other channels; similarly, recycling policy could affect household efforts on other environmental activities. Given that the performance of one activity has been argued to crowd out ('moral licensing') as well as in ('moral consistency') other activities, such spillovers could plausibly have either sign.

The first paper develops a model where agents may contribute to a single public good through several different activities. For a large set of plausible cases, we predict that policy facilitating one activity reduces effort on other activities, though overall public-good production still increases. Furthermore, the multi-activity framework admits new interpretations that are sometimes at odds with prominent results from single-activity variants of our model.

In the second paper, we run an experimental dictator game where subjects may donate to two different real-world charities. To simulate activity-specific interventions, we vary the relative productivity of those charities, and introduce several treatments to test whether spillovers occur even across (possibly very) dissimilar alternatives. We find that negative spillovers occur significantly in all cases, but that the effect is weaker, the more dissimilar are the charity alternatives.

The third paper estimates policy spillovers within the context of a natural experiment on food waste in Sweden. We use a difference-in-difference design to measure the causal impact of introducing food-waste collection on the sorting of packaging waste. Results suggest a positive spillover effect corresponding to 5-10% of the population average. Point estimates are relatively small, and sometimes insignificant, when we control for shifts in the waste-related incentives facing households. Although unable to control for all such incentive shifts, we argue that the remaining bias may be negligible.

Key words
Behavioral spillovers, prosocial behavior, public goods, moral licensing, charitable giving, environmental behavior, dictator game, recycling

Classification system and/or index terms (if any)
JEL classification: C91, D03, D04, D11, D12, H41, Q53

Supplementary bibliographical information	Language English
ISSN and key title 0460-0029 Lund Economic Studies no. 195	ISBN 978-91-7623-904-9

Recipient's notes	Number of pages 156	Price
	Security classification	

Signature Claes Ek   Date July 4, 2016

Behavioral Spillovers across Prosocial Alternatives

Claes Ek

Lund University

Lund Economic Studies Number 195

School of Economics and Management, Department of Economics

ISBN: 978-91-7623-904-9 (print)
ISBN: 0987654321 (pdf)
ISSN: 0460-0029 Lund Economic Studies no. 195

Printed in Sweden by Media-Tryck, Lund University, Lund 2016

Contents

Abstract

This thesis contributes to the economic literature on prosocial behavior. It includes three papers, all of which relate to the issue of policy-driven spillovers across prosocial alternatives. For example, the common fundraising practice of donation matching could affect contributions through other channels; similarly, recycling policy could affect household efforts on other environmental activities. Given that the performance of one activity has been argued to crowd out ('moral licensing') as well as in ('moral consistency') other activities, such spillovers could plausibly have either sign.

The first paper develops a model where agents may contribute to a single public good through several different activities. For a large set of plausible cases, we predict that policy facilitating one activity reduces effort on other activities, though overall public-good production still increases. Furthermore, the multi-activity framework admits new interpretations that are sometimes at odds with prominent results from single-activity variants of our model.

In the second paper, we run an experimental dictator game where subjects may donate to two different real-world charities. To simulate activity-specific interventions, we vary the relative productivity of those charities, and introduce several treatments to test whether spillovers occur even across (possibly very) dissimilar alternatives. We find that negative spillovers occur significantly in all cases, but that the effect is weaker, the more dissimilar are the charity alternatives.

The third paper estimates policy spillovers within the context of a natural experiment on food waste in Sweden. We use a difference-in-difference design to measure the causal impact of introducing food-waste collection on the sorting of packaging waste. Results suggest a positive spillover effect corresponding to 5-10% of the population average. Point estimates are relatively small, and sometimes insignificant, when we control for shifts in the waste-related incentives facing households. Although unable to control for all such incentive shifts, we argue that the remaining bias may be negligible.

Keywords: behavioral spillovers, prosocial behavior, public goods, moral licensing, charitable giving, environmental behavior, dictator game, recycling

JEL classification: C91, D03, D04, D11, D12, H41, Q53

Lund, July 2016

Acknowledgments

Q. Why would this applicant wish to join the PhD program?

A. At the risk of seeming a romantic, I will answer: to seek the truth.

I reproduce these, the opening lines of my application letter to the Lund PhD program, to serve as a bookend to the five years I have spent doing exactly that: seeking the truth. As stressed in the original letter (which presented an entire list of caveats immediately following the above passage), in general I do not believe in simple truths, nor do I believe that finding the truth is simple. Indeed, the academic approach to truth-seeking is slightly counter-intuitive: one best pursues truth by not immediately accepting what appears, on the face of it, to be true. Instead, the ideal is to criticize and closely scrutinize even relatively persuasive arguments and results, so as to constantly do better.

Achieving this ideal is very difficult. It also produces (by necessity, perhaps) an academic culture where people who, like me, are somewhat timid and prone to bouts of self-doubt sometimes struggle. Add to that the sheer amount of hard work involved, and doing science may seem a truly daunting task. And yet, in the end, I find it difficult to think of a more rewarding one. I never pictured myself the economist (in fact I still don't), yet here I am, five years later, PhD thesis all but finished. Clearly, thanks are due to all the people who aided me in getting this far. So, this section is for you.

First and foremost, I would like to thank my two supervisors Erik Wengström and Ola Andersson for their attentiveness, helpfulness, and always constructive feedback. Let me provide some feedback in return: it has truly been a pleasure working together, and I have learned a lot from you, as hopefully you would agree. I also wish to thank Håkan J. Holm for providing initial supervision and getting me on the right track in those confusing and confused early years. I am very grateful to Håkan for giving me an opportunity to run my first lab experiment by including me in his research project financed by the Jan Wallander foundation.

I am indebted to a number of other people at the department. Thanks go to Tommy, for persuading me to apply to the PhD program in the first place and for providing good and frank advice in the years since. To Pontus, for having such great faith in my potential as a lecturer and for helping me decide all kinds of tricky borderline cases. To Natalia, for helpful comments and advice on all of my papers. And to the administrative personnel: Azra, Jenny, Mariana, and Nathalie, for aiding me in navigating the many intricacies of everyday work at a university.

Moreover, interactions with the following people at conferences, workshops, and departmental visits have been instrumental in improving the thesis: Kjell-Arne Brekke, Toke Fosgaard, Lars Gårn Hansen, Alessandro Martinello, Karsten Neuhoff, Karine Nyborg, Amrish Patel, Marco Piovesan, Arthur Schram, and Lea Skræp Svenningsen. A big thank you also to my co-author Jurate, without whom I would have struggled to produce one last article on time. Finally, I want to thank Thomas Sterner and Wolfgang Habla for a very pleasant month at Gothenburg University.

Within the community of PhD candidates at the department, a number of people stand out. In no particular order: to Jens, for helping me get my bearings early on, and for writing excellent lecture notes, of which I stole probably half. To Albin, for lunch at the old Sparta restaurant and for being an early example that economists, contrary to expectations, can actually be pretty decent people. To Emma, for being an all-around excellent office mate and for sharing my passion for the Civilization series; it is impossibly unfair that you're gone. To Hilda, for co-founding the Brain Trust and for always having a cool research idea up your sleeve. To Sara, for exclaiming 'Oh, come on!', liberating junior researchers everywhere. To Lina, for the bitterness and the great parties. To Sanna, for hilariously throwing me off balance on my first lecture.

We are halfway through. To Anna, for too many things to count, but including: our mutual rants about life in academia and on Planet Earth, your patience as my go-to second opinion on all things research-related, and being great company in the old '268; plus, of course, ping-pong. To Karl, for all the jokes, and for your patience as my *second* go-to second opinion; and ping-pong. To Jim and Jörgen, for good discussions on all things theoretical; and ping-pong. To Margaret, for the long hours we spent cracking the math of that French duo; and ping-pong. Finally, to Pol; for ping-pong.

I am grateful to Ann-Charlotte, Måns and Sofia for assistance in running my experiment. Also Peter, whose cozy corner shall not be forgotten.

My dear friends also deserve thanks. When the load got heavy, I leaned on you. Above all, I am deeply grateful to Eskil, who once again (like it or not!) has probably been better informed about my work than anyone else, except perhaps my supervisors. In Uppsala, our weekly lunches will be sorely missed, but there are still worlds for us to conquer. Big thanks also go out to: Ulrika, Linus, Rebecca, Emily, Fredrik, Peter (again), Minna, Caroline, Nora, and Filippa. The Stockholm crowd: Binge, Carl, Mikael, My, Niklas, Per, and Ulf. And a very special thank you to Sofia.

And finally, as is appropriate, I wish to thank my family. To Thomas, for blazing the trail. And to Mum and Dad, for your endless, boundless support in all things; thank you. I hope this little book I wrote makes you proud:

Chapter 1

Introduction

1.1 Background

Much of economics boils down to a study of human behavior in what is perhaps tautologically termed 'economic situations'. Some of these may seem obvious: buying and selling goods or services, deciding how much labor to demand (for firms) or supply (for workers), and deciding how long to stay in school are all examples of situations where economic considerations are important. By contrast, consider 'prosocial' behavior: the ways that people do good, be it with respect to their fellow man, the environment, or some abstract ideal. Is this a valid subject for economics?

Arguably yes, for at least two reasons. First, the charitable sector represents quite a large share of GDP in many countries, not least in the United States, where giving from private sources is roughly $300 billion per year (Andreoni and Payne, 2013). Second, it is no great exaggeration to say that charitable donations, voluntary work, and everyday prosocial acts have a significant and lasting impact on society, including the economy.

In economic theory, prosocial behavior of various kinds is typically conceived of as contributions by individuals to some 'public good', the benefits of which are shared by everyone. For example, if people's actions help to create a cleaner environment, this benefit is enjoyed (all else being equal) by all members of society. Similarly we might think that even the well-fed, affluent, and healthy would enjoy living in a society freed from hunger, poverty, and disease.

The trouble, from the point of view of economists, is that public goods can be enjoyed even by those who did not contribute to them in the first place; in technical terms, public-good consumption is 'nonexcludable'. Because of this feature, there is a danger that some people will free ride: that is, although they appreciate the public good, they will try to contribute as little as possible in the hope that other people will still provide enough of the public good. They might also decide that a single individual's impact is too small to warrant contributions, even if no one else is supplying the public good. Of course, if a large proportion of the population approaches the problem in this way, too little of the public good will be supplied.

Note that this social dilemma of underprovision is not based on any assumption that people are selfish. Indeed, the dilemma arises in any situation where human beings possess a preference for having the public good supplied, regardless of whether that preference is driven by selfish concerns (because contributors themselves benefit from the public good) or not. Underprovision follows instead from the assumption (which we do make) that people do not particularly care who does the supplying of the public good, in which case different people's contributions become substitutes: if you provide more, then I will want to respond by providing *less*.

However, this story seems to paint an overly pessimistic picture of prosocial behavior. In the real world, not all people free ride all the time. Accordingly, Andreoni (1988, 1990) showed that the above model of behavior, which he termed the 'pure altruist' model, fails to predict many commonly observed behaviors, such as widespread charitable donations.

For example, it predicts that only the very wealthiest individuals (who are rich enough to single-handedly make a difference) will donate to charity at all. As an alternative, he launched a concept of 'impure altruism', where people are motivated not only by a preference for seeing the public good supplied, but for supplying it themselves: what Andreoni termed the 'warm glow' of giving. Although psychologically speaking this term is quite vague, it is probably deliberately so, and Andreoni (1990) did note that many different psychological mechanisms could give rise to impurely altruistic preferences.

Much of the subsequent research on prosocial behavior can in fact be viewed as attempts to flesh out the psychology of public-good contributions in more detail, so as to better predict behavior e.g. in response to policies aimed at increasing prosocial behavior, such as monetary incentives, information campaigns, and 'nudges' of various types. For example, Brekke et al. (2003) interpret warm glow in terms of self-image and as the desire to adhere to a moral norm. Bénabou and Tirole (2006) model it as the satisfaction derived from successfully signaling to a third party (which they argue could even be one's future self) that one is a prosocial person. In Rotemberg (2014), warm glow is the desire to help other contributors to the public good feel validated.

1.2 Contribution of the thesis

Most work in the literature on prosocial behavior, however, builds upon a partial analysis where it is assumed that people may contribute to some public good through only a single prosocial activity. This is not very realistic, as most causes offer several ways to contribute: for example, people may help the environment by recycling household waste, buying organic products, donating to an NGO, etc. The present doctoral thesis therefore considers issues of *how* (i.e. by which means) people choose to contribute, in addition to those of *why* they do so.

Specifically, we look at policy-driven 'behavioral spillovers'. Public policy sometimes targets just a single prosocial activity; for example, local governments often promote household recycling through information campaigns or economic incentives. Such policies could spill over onto *other* environmental acts, like buying organic groceries. Moreover, the effect may plausibly go either way. If the number of good deeds that people are willing to perform is roughly constant, policy may lead them to recycle more but also to feel that they have a 'moral license' to spend less on organic food. We might call this *negative spillovers.* Conversely, if for instance the policy raises overall environmental awareness, both recycling and organic purchases may rise, causing *positive spillovers.*

This thesis includes three papers on the spillover issue. The first is theoretical, the second performs a lab experiment, and the third looks for spillovers using real-world data.

Paper 1: Prosocial Behavior and Policy Spillovers: A Multi-Activity Approach

The first paper takes as its starting point the Brekke et al. (2003) model of moral motivation, extending it to include not one, but two prosocial activities that people can choose from. The aim of this extension is twofold. First, we check whether prominent results from the standard single-activity version of the Brekke et al. (2003) model hold also in the multi-activity case. If not, we argue, then there is a danger that current results in the literature are not as relevant for the real world as one might hope. Some results do become less general and more ambiguous in the multi-activity setting. We suggest that single-activity models are best understood as capturing behavior in the aggregate, and caution against using them to understand any one, disaggregated, prosocial activity.

Second, we derive results on policy spillovers. In the Brekke et al. (2003) model, the motivation to contribute derives from an 'ideal contribution' which reflects what a responsible citizen would ideally do. Consistent with the intuition given previously, if this ideal remains fixed in the face of a policy intervention, negative spillovers arise under plausible conditions. If instead the ideal is allowed to increase as a result of the policy, spillovers could go either way.

Paper 2: Some Causes are More Equal than Others? The Effect of Similarity on Spillovers in Charitable Giving

The second paper presents a lab experiment on the same theme. In addition to testing model predictions on the sign of spillovers, the purpose of the experiment is to make a first attempt at charting their 'reach': do significant spillovers arise even across (very) dissimilar activities? The design is simple. We use a so-called dictator game design where subjects can donate to two real-world charities using money supplied by the research budget. To donate, they first earn 'points' in a real-effort task where, as a stand-in for policy shifts, we systematically vary the relative productivity (number of donated Swedish crowns/point) of the charities. This permits estimation of spillover effects in the experimental setting. In addition, there are four between-subject treatments. In each, different charities are available for donations, thereby manipulating the similarity of charity pairs along two dimensions: geographical scope (local/global) and cause (welfare of children/environmental).

The following results are obtained. First, significant negative spillovers arise in all treatments: as the exchange rate of one charity rises, subjects allocate less points to the other charity. Second, the magnitude of spillovers for the most dissimilar charity pair is only around half of that for the most similar pair. There is no significant difference in how strongly the local/global and the cause dimensions drive spillovers. Finally, post-session questionnaire items asking subjects how similar they think each charity pair is are found to be highly predictive of negative spillovers within and across treatments, and may prove useful as a starting point for predicting spillovers from actual or planned policies.

Paper 3: Behavioral Spillovers from Food-Waste Collection in Swedish Municipalities

The third paper is an empirical application that is joint work with Jurate Miliute-Plepiene at the Lund University Centre for Sustainability Studies (LUCSUS). We attempt to identify real-world policy spillovers using yearly panel data on the amount of packaging waste collected in more than 200 Swedish municipalities. Over the past few decades, most municipalities have begun to collect food waste from households, and we match data on the timing of such collection with the packaging-waste data. A difference-in-difference design with municipality fixed effects, exploiting the staggered implementation of food-waste systems, may permit isolation of the causal spillover effect of food-waste collection on packaging amounts collected.

Results indicate a significant positive spillover effect which rises subsequent to formally stated adoption dates, possibly due to the fact that implementation is known to have been slow in many areas. When we account for certain known changes in the monetary and non-monetary recycling incentives facing households, the effect diminishes but is still on the order of 5-10% of the data average. In particular, estimates are smaller when the introduction of curbside collection of packaging from single-family households is controlled for. While we are unable to include all such incentive shifts (leaving open the possibility of remaining upward bias), we point to institutional factors and perform certain tests to argue that those that remain unaccounted for are less problematic than those that we do include.

References

J. Andreoni. Privately provided public goods in a large economy: the limits of altruism. *Journal of Public Economics*, 35(1), 1988.

J. Andreoni. Impure Altruism and Donations to Public Goods: A Theory of Warm-glow Giving. *The Economic Journal*, 100(401), 1990.

J. Andreoni and A. Payne. Charitable giving. In A. Auerbach, R. Chetty, M. Feldstein, and E. Saez, editors, *Handbook of Public Economics.* 2013.

R. Bénabou and J. Tirole. Incentives and Prosocial Behavior. *American Economic Review*, 95(5), 2006.

K.A. Brekke, S. Kverndokk, and K. Nyborg. An economic model of moral motivation. *Journal of Public Economics*, 87(9), 2003.

J.J. Rotemberg. Charitable giving when altruism and similarity are linked. *Journal of Public Economics*, 114, 2014.

Chapter 2

Prosocial Behavior and Policy Spillovers: A Multi-Activity Approach

2.1 Introduction

Do people view each commonly undertaken prosocial act as completely distinct, or rather as belonging to one of several categories? Research on 'mental accounting' (Henderson and Peterson, 1992; Thaler, 1999; Read et al., 1999) suggests the latter; for example, recycling household waste, buying organic products, donating to environmental NGOs, *et cetera* are all examples of 'environmentally conscious' acts; voting, running for local office, writing letters to the editor of a newspaper, or demonstrating are ways of 'strengthening democratic society'; and contributing to particular charities by donations or voluntary work 'aids the less fortunate'. Accordingly, an individual who decides to contribute to these causes will find that there are many ways to do so. By contrast, the vast majority of economic research on prosocial behavior assumes that people may contribute through only a single activity. While explaining *why* people contribute is clearly of primary importance, there are also interesting issues concerning *how* they choose to do so; this paper thus forms a systematic attempt at re-analyzing some aspects of current theory using more than one contributing activity.

Our aim with the multi-activity extension is twofold. First, we deliver positive results on policy-driven 'spillovers' in prosocial behavior. Public intervention sometimes focuses on only one activity within a given prosocial category; for example, many countries have introduced directed schemes to encourage (whether through facilitation, public information or economic incentives) specific environmental behaviors like recycling, leaving the car at home, or reducing energy use. Yet if the efforts that people spend on a given activity reflect not merely its own characteristics, but those of similarly categorized acts as well, such policies may have significant behavioral spillover effects: to encourage recycling may affect e.g. sales of organic products. While a few field studies exist (Jacobsen et al., 2012; Tiefenbeck et al., 2013), this point deserves further theoretical analysis.

Second, we subject existing single-activity modelling approaches to a kind of robustness check. The point we wish to make is the following. While a one-dimensional model whose single activity is understood as an aggregate of all existing activities is likely to align well with aggregate outcomes in a multi-activity model, these aggregate results need not apply to each individual activity. Note that this idea is quite standard: for instance, even if overall consumption is normal, individual goods may be inferior. Despite this, single-activity models are frequently taken to have (policy) relevance for individual real-world prosocial activities (e.g. recycling), even when that activity exists among a plausible set of others. We would argue that this practice is problematic and that, in certain cases at least, a multi-activity model is much better suited to the task.

The treatment in this paper draws on the public-goods model introduced by Brekke et al. (2003). Our robustness check, then, amounts to revisiting prominent results from single-activity variants of that model, showing that the multi-activity framework admits new interpretations which sometimes conflict with those of models that are framed in terms of a single, non-aggregate activity. This is because, loosely speaking, single-activity models implicitly assume that agents use narrow bracketing to categorize prosocial activities and decisions, while with multiple activities, it becomes possible to model either narrow or broad bracketing. Our model imposes the latter, and we argue that it is preferable to an otherwise equivalent narrow-bracket model capable of replicating earlier results.

The Brekke et al. (2003) model focuses on moral rather than social norms, and is usually interpreted as mimicking 'duty-orientation': individuals derive utility from maintaining a positive self-image of themselves as responsible citizens, but also *dis*utility from failing to live up to some 'ideal' level of effort. 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 Brekke et al. (2003) 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, so any explanation of them in terms of social image will be indirect at best.[1]

We model policy interventions as an activity-specific productivity shift.[2] Indeed, the kind of spillovers we examine can be seen as a manifestation of standard notions of substitutability: as policies alter material incentives to contribute through the targeted activity, relative-price effects shift demand for other activities up or down. Similarly, population-level spillovers sometimes manifest themselves as the 'crowding-out' of

[1] Variants of the Brekke et al. (2003) model have been applied to quite a broad set of issues, including 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).

[2] Because our focus is on productivity, we do not consider spillovers arising due to priming or 'nudges', e.g. by providing feedback on peer behavior.

other activities, and sometimes as 'crowding-in'.[3] However, behavioral economists and psychologists have sought more detailed predictions, and have struggled to explain the inconsistent sign of spillovers (Merritt et al., 2010; Dolan and Galizzi, 2015).

The mixed evidence is reflected in competing theoretical accounts. On the one hand, several psychological models stress consistency across separate decisions (Festinger, 1957; Bem, 1967), suggesting that one prosocial act can spur another. Hence, if activity-specific policy is effective in the sense of increasing effort on the targeted activity, we should expect positive-sign spillovers.[4] On the other hand, a competing 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 justified in subsequently relaxing their moral standards ('moral licensing'), while people who have just behaved badly may feel obliged to engage in subsequent 'cleansing'. Again, if the policy is effective, efforts on other activities should decrease, so this hypothesis corresponds to negative-sign spillovers.[5]

[3]What we refer to here differs from the types of 'crowding' already examined by economists. It has long been argued that government provision of public goods may crowd out voluntary contributions by individuals (Andreoni, 1990; Nyborg and Rege, 2003); using a multi-activity model, Ribar and Wilhelm (2002) extend the argument, showing that spending by one charity (incompletely) crowds out spending by other charities. In addition, the related but distinct literature on 'motivation crowding theory' argues that under certain circumstances economic incentives to contribute (such as subsidies) may *decrease*, rather than increase, contributions (Frey and Jegen, 2001; Bowles, 2008). To this we add a third layer of complexity, namely that public policy with respect to a particular activity may crowd out (or in) effort on *other* activities.

[4]Gneezy et al. (2012) argue that people base their view of themselves on observing their own behavior; thus, if the prosocial behavior that an individual engages in is sufficiently costly (and so sufficiently informative), she is led to conclude that she is a prosocial type and will subsequently try to act in accordance. They present laboratory and field experiment data in support of this idea. Similarly, Brown et al. (2012) found that the more people donated to the 2004 Indian Ocean tsunami, the more likely they were to donate *more*, rather than less, to charity in the future. Also, Greenberg (2014) presents data indicating that tipping in restaurants increases during the holiday season, when people are presumably more charitable in other ways as well.

[5]Moral balancing is illustrated by Ploner and Regner (2013), whose results seem

In our model, agents base their actions partly on the (perceived) characteristics of the public-good production function, and we are able to show that for a range of realistic functions, activity-specific policy crowds *out* effort on other activities, though public-good production and total effort still increases. Thus, we establish negative spillovers as the benchmark in anonymous public-goods settings. This result, however, is limited in two ways. First, in the latter half of the paper, we explore a model variant where agents endogenously update the 'ideal' contribution that they hold as a benchmark. In this case, spillovers are found to be theoretically 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; in line with the overall theme of this paper, negative spillovers need not apply to all (real-world) activities within that aggregate.

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. Sections 4 considers a more general case, and 5 revisits earlier research on government centralization of a particular contributing activity. Sections 6 considers an extension where the magnitude of the 'ideal contribution' is endogenously determined; section 7 then forms a critical re-examination of an earlier result from such an endogenous-ideal model. Section 8 discusses why earlier results prove difficult to replicate. Finally, section 9 concludes.

to contradict the conjecture of Gneezy et al. (2012) that costly prosocial behavior necessarily promotes consistency. At an initial stage, subjects could donate to charity by participating in a real-effort task; this was followed by a dictator game. Ploner and Regner (2013) found that subjects who contributed relatively little in the initial game tended to reverse their behavior and be especially generous in the dictator game. There are caveats, however. The dictator-game endowment was determined by the roll of a die, the outcome of which (in one treatment) could be misrepresented by subjects; but, contrary to expectations, cheating in such a manner did not significantly affect subsequent play in the dictator game.

2.2 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 (nonstrategic) 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

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

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 θ 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 θ_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, θs 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 $g^0 = 0$.[6]

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

$$I = I(g(\mathbf{e}, \boldsymbol{\theta}) - g^*) \tag{2.1}$$

where $g(\mathbf{e}, \theta)$ 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 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 $g(\mathbf{e}, \boldsymbol{\theta}) = g^*$;

[6] Given all of the above assumptions, one can show that it must be possible to write the production function as $g = f_0(e_1, ..., e_K) + \sum_k f_k(e_k, \theta_k)$. 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.

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 6 considers endogenous ideal formation of the particular type explored in Brekke et al. (2003); Section 7 (re)analyzes an application of that model.

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

$$\max_{\mathbf{e}} U(\mathbf{e}, \boldsymbol{\theta}) = u\left(T - \sum_{k=1}^{K} e_k\right) + v(G_{-i}(\boldsymbol{\theta})) + I\left(g(\mathbf{e}, \boldsymbol{\theta}) - g^*\right) \quad (2.2)$$

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.[7] We assume throughout most of this paper that N is large enough that each

[7] 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.

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, though its magnitude may change as other agents adjust their efforts in response to productivity shocks. Hence, although irrelevant to individual behavior, $v(G_{-i})$ will prove important for analyzing changes in individual utility, e.g. in Section 4.[8]

As U is continuous and the choice set is compact, the set of solutions to utility problem (2.2) is nonempty. The necessary and sufficient Kuhn-Tucker conditions for a solution are

$$-u' + g'_{e_k} I' - \lambda \leq 0 \tag{2.3}$$

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* $\bar{\mathbf{e}}$ *is a solution to problem (2.2), we have* $g^0 \leq g(\bar{\mathbf{e}}) \leq g^*$. *If also* $g^* > g^0$, *then* $g(\bar{\mathbf{e}}) < g^*$.

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.

[8] It is worth noting here that when the production function is linear ($g = \sum_k \theta_k e_k$), or more generally whenever we can write $g = g(\theta_1 e_1, ..., \theta_K e_K)$, productivity-led spillovers correspond to standard notions of price-driven gross substitution. To see this, we interpret activities not as durations but as units of production; that is, define $\hat{e}_k = \theta_k e_k$ for all k. Then we have $g = g(\hat{e}_1, ..., \hat{e}_K)$ 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 (holding θ_2 constant) 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.

2.3 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_2 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.[9]

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,

[9] Of course, our choice of the year 2050 arbitrary; since IPCC (2013) also states that warming is largely determined by cumulative emissions, the only viable long-run level of per-capita emissions is zero.

individuals simply maximize

$$\begin{aligned} U(e) &= T - e + G_{-i} - a\left(-F - (-F^*)\right)^2 \\ &= T - e + G_{-i} - a\left(F^* - F^0 + \theta\sqrt{e}\right)^2 \end{aligned} \tag{2.4}$$

subject to $0 \leq e \leq T$. Note that $v(G_{-i})$ has been dropped from (2.4), 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 $a > 0$.[10]

We denote the interior maximum to (2.4) by $\bar{e}$. It is

$$\bar{e} = \left(\frac{a\theta\left(F^0 - F^*\right)}{1 + a\theta^2}\right)^2 \tag{2.5}$$

Let us also write

$$F(\bar{e}) = F^0 - \theta\sqrt{\bar{e}} = \frac{F^0 + a\theta^2 F^*}{1 + a\theta^2} \tag{2.6}$$

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 θ increases. Then

$$\frac{dF(\bar{e})}{d\theta} = -\frac{2a\theta\left(F^0 - F^*\right)}{\left(1 + a\theta^2\right)^2} < 0 \tag{2.7}$$

so all agents reduce their emissions in response; but

$$\frac{d\bar{e}}{d\theta} = \frac{2a^2\theta\left(F^0 - F^*\right)^2}{\left(1 + a\theta^2\right)^3}\left(1 - a\theta^2\right) \tag{2.8}$$

[10]The marginal product of effort approaches infinity as $e \to 0$, 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 (2.6) 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.[11] Of course, if productivity keeps increasing, effort will always drop eventually. Indeed, in the limit as $\theta \to \infty$, it becomes possible to live up to the ideal with next to no effort: $\bar{e} \to 0$ but, by (2.6), $F(\bar{e}) \to F^*$.

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

$$U(e_1, e_2) = T - e_1 - e_2 + G_{-i} - a\left(F^* - F^0 + \theta_1\sqrt{e_1} + \theta_2\sqrt{e_2}\right)^2 \quad (2.9)$$

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

$$\bar{e}_k = \left(\frac{a\theta_k\left(F^0 - F^*\right)}{1 + a\left(\theta_1^2 + \theta_2^2\right)}\right)^2 \quad (2.10)$$

and

$$F(\bar{e}_1, \bar{e}_2) = F^0 - \theta_1\sqrt{\bar{e}_1} - \theta_2\sqrt{\bar{e}_2} = \frac{F^0 + a\left(\theta_1^2 + \theta_2^2\right)F^*}{1 + a\left(\theta_1^2 + \theta_2^2\right)} \quad (2.11)$$

[11] 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 θ 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.

where again we have $F^* < F(\bar{e}_1, \bar{e}_2) < F^0$.[12]

Now consider an increase in θ_1 (the effect of a shift in θ_2 being completely analogous). Activity 2 should then be interpreted as an aggregate of all climate-related activities that are not directly affected by the shift in θ_1. Suppose e_1 is the (extra) time invested when traveling by bicycle instead of taking the car,[13] and θ_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 (2.11) with respect to θ_1 now reveals that

$$\frac{dF(\bar{e}_1, \bar{e}_2)}{d\theta_1} = -\frac{2a\theta_1 \left(F^0 - F^*\right)}{\left(1 + a\left(\theta_1^2 + \theta_2^2\right)\right)^2} < 0 \qquad (2.12)$$

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

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

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

[12] 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

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

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.

[13] 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.

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.

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

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

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; (2.12) 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 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 (2.7) and (2.12) shows that this reduction is overestimated. Becuase 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, $g''_{e_1e_2} = \partial^2 g/\partial e_1\partial e_2$, 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.

2.4 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 *co-ordinatewise* order, where for two vectors $x, y \in \mathbb{R}^n$, $x \geq y$ if $x_i \geq y_i$ for each $i = 1, ..., n$. In this case the definitions of join and meet are particularly simple, namely $x \vee y = (\max\{x_1, y_1\}, ..., \max\{x_n, y_n\}) \in \mathbb{R}^n$ and $x \wedge y = (\min\{x_1, y_1\}, ..., \min\{x_n, y_n\}) \in \mathbb{R}^n$. 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 (2.2) changes when we shift some productivity parameter θ_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 $U : X \times R \to \mathbb{R}$, where X is a lattice and R is a* *partially ordered set. If the constraint set* $S : R \to 2^X$ *is nondecreasing in* r *and if* U *is quasisupermodular 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 (2.2) 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 (2.2), with $r = \theta_k$? While checking for quasisupermodularity and the single crossing property is straightforward, nondecreasingness of the constraint set $S = \{(e_1, ..., e_K)| \sum_k e_k \leq T\}$ is trickier. This set does not depend on productivity, so for any θ_k we will have $S(\theta_k) = S$. We therefore obtain nondecreasingness in θ_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 $e' = (T, 0, ..., 0, 0) \in S$ and $e'' = (0, 0, ..., 0, T) \in S$. Under the coordinatewise order, $e' \vee e'' = (T, 0, ..., 0, T)$, which is not in S. Hence, S is not a sublattice.

Fortunately, for $K = 2$ a workaround exists. Redefine (2.2) as[14]

$$\max_{e_k, \tilde{e}_l} U(e_k, \tilde{e}_l, \theta_k) = u(T - e_k + \tilde{e}_l) + v(G_{-i}(\theta_k)) + I(g(e_k, -\tilde{e}_l, \theta_k) - g^*) \quad (2.13)$$

for $k \neq l$, with $\tilde{e}_l = -e_l$. Then for $X = \{(e_k, \tilde{e}_l) | e_k \geq 0, \tilde{e}_l \leq 0\}$ and $S = \{(e_k, \tilde{e}_l) \in X | e_k - \tilde{e}_l \leq T\}$, 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.[15] As before, in most settings activity $l \neq k$ should be interpreted as an aggregate of all the 'other activities' that are being only indirectly impacted by the productivity shock. Thus, although we will be able to describe the movement of the aggregate, we can draw no conclusions about the particular activities within it.

Note that when e_l is redefined in the above manner, Proposition 2 implies a tendency for negative spillovers. In particular, if all solution sets are singleton, e_k is increasing in θ_k but $e_l = -\tilde{e}_l$ is decreasing. This is the 'crowding-out' effect observed in the previous section.

2.4.1 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

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

[14] For notational brevity, we suppress θ_l as an argument of U throughout this section.

[15] 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 $x \wedge y = (0, T, T) \notin S$.

Note that the interval $0 \leq e \leq T$ is a sublattice of $\mathbb{R}_+$, so analyzing a shift in θ 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

$$\frac{\partial^2 U(e,\theta)}{\partial e \partial \theta} = I'' g' g'_\theta + I' g''_{e\theta} \geq 0 \tag{2.14}$$

This condition is analogous to that required for (2.8) to be positive; indeed, for $I = -a(F^* - F)^2$ and $g = \theta\sqrt{e} - F^0$, it translates into $F(e) \geq (F^0 + F^*)/2$. It specifies when e and θ are complements in a utility sense. When θ 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'_θ (from shifting θ 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.[16]

Now, suppose $K = 2$ and consider a shift in activity-specific productivity parameter θ_k. We differentiate (2.13) twice to find sufficient conditions for Proposition 2 to hold:

$$\frac{\partial^2 U(e_k, \tilde{e}_l, \theta_k)}{\partial e_k \partial \theta_k} = I'' g'_{e_k} g'_{\theta_k} + I' g''_{e_k\theta_k} \geq 0 \tag{2.15}$$

$$\frac{\partial^2 U(e_k, \tilde{e}_l, \theta_k)}{\partial \tilde{e}_l \partial \theta_k} = - I'' g'_{e_l} g'_{\theta_k} - I' g''_{e_l\theta_k} \geq 0 \tag{2.16}$$

$$\frac{\partial^2 U(e_k, \tilde{e}_l, \theta_k)}{\partial e_k \partial \tilde{e}_l} = - u'' - I'' g'_{e_k} g'_{e_l} - I' g''_{e_1 e_2} \geq 0 \tag{2.17}$$

Here conditions (2.15) and (2.16) check for increasing differences, while

[16] Since in general both terms in (2.14) depend on e, the condition may be satisfied only for some initial values of e. We saw in footnote 11 that if agents are heterogeneous the idea may be extended to people, with different types reacting differently.

(2.17) is a quasisupermodularity condition. If all three hold, the set of optimal redefined efforts $(e_k, \tilde{e}_l)$ is monotone nondecreasing in θ_k; in other words, spillovers will tend to be negative.

Condition (2.15) is exactly the same type of within-activity complementarity condition we observed in the single-variable case. Condition (2.16) states that θ_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'_{θ_k}, so it is immediately true.

Finally, (2.17) 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.[17]

2.4.2 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

[17] 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.

increases with θ_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 where choose their sum of efforts $\hat{e} = e_1 + e_2$.

We begin by stating a useful result. The following lemma guarantees, first, 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 2.1. Utility maximization becomes a two-stage process; we will focus on the agent's choice of total efforts, which are then allocated in an optimal fashion across specific activities according to $\hat{g}(\hat{e})$. Second, Lemma 1 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 (2.2) corresponds to a unique optimal production level*

$$\hat{g}(\hat{e}, \theta_k) = \max_{e_1, e_2} g(e_1, e_2, \theta_k) \tag{2.18}$$

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

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

It follows from Lemma 1 that a given effort vector $\mathbf{e} = (e_1, e_2)$ will solve problem (2.2) if and only if $\hat{e} = e_1 + e_2$ solves

$$\max_{\hat{e}} U_{\hat{e}}(\hat{e}, \theta_k) = u(T - \hat{e}) + v(G_{-i}(\theta_k)) + I(\hat{g}(\hat{e}, \theta_k) - g^*) \tag{2.19}$$

subject to $0 \leq \hat{e} \leq T$. The equivalence arises because any solution to either (2.2) or (2.19) will solve the PMP; the former, because of Lemma 1; the latter, by construction.[18]

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

[18] In Appendix 2.9, we derive sufficient conditions for total efforts $\hat{e}$ to increase in line with activity-specific productivity and the ideal contribution.

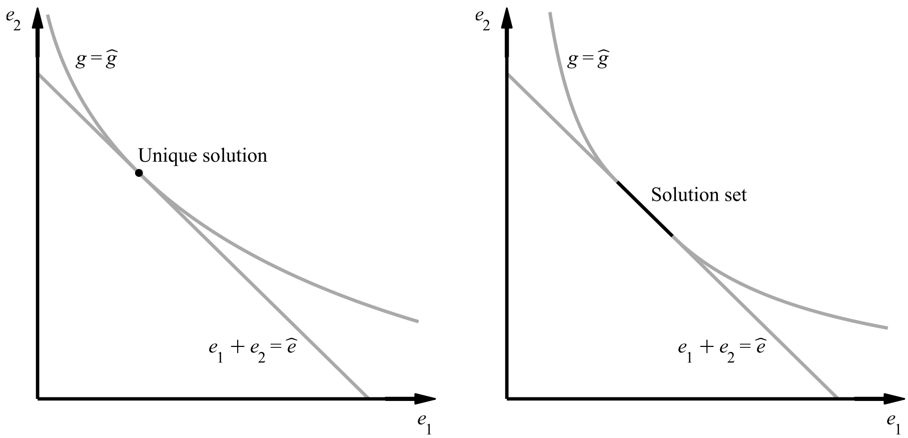


Figure 2.1: *The production-maximization problem. Left panel: the PMP has a unique solution. Right panel: the PMP has multiple solutions.*

$\hat{e} = \hat{g}^{-1}(\hat{g}, \theta_k)$. Thus, reformulate (2.19) yet again as

$$\max_{\hat{g}} U_{\hat{g}}(\hat{g}, \theta_k) = u(T - \hat{g}^{-1}(\hat{g}, \theta_k)) + v(G_{-i}(\theta_k)) + I(\hat{g} - g^*) \qquad (2.20)$$

subject to $0 \leq \hat{g}^{-1}(\hat{g}, \theta_k) \leq T$. Note that although the constraint set now depends on the parameter θ_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 (2.20) is

$$\frac{\partial^2 U_{\hat{g}}}{\partial \hat{g} \partial \theta_k} = u'' \left(\hat{g}^{-1}\right)'_{\theta_k} \left(\hat{g}^{-1}\right)'_{\hat{g}} - u' \left(\hat{g}^{-1}\right)''_{\hat{g}\theta_k} \qquad (2.21)$$

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 θ_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 (2.21) is strictly positive, and production is monotone nondecreasing in θ_k.

The following proposition summarizes what we have learned regarding policy spillovers.[19]

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* θ_k. *Furthermore, if*

$$I'' g'_{e_k} g'_{\theta_k} + I' g''_{e_k \theta_k} \geq 0$$

$$- u'' - I'' g'_{e_k} g'_{e_l} - I' g''_{e_1 e_2} \geq 0$$

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

Although allowing for multiple utility maxima complicate the picture somewhat, we can safely 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 quasiconcave.[20] Thus, when the ideal contribution is exogenous, the pattern of negative-sign spillovers seems relatively clear-cut.

[19] 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 (2.20) 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.

[20] 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 increasing in g and concave.

2.5 Earlier results: centralizing an activity

We now revisit a paper by Bruvoll and Nyborg (2004) in which they analyze the aggregate utility consequences of having government centralize a single contributing activity. Using a single-activity model with an exogenous ideal, they ask what would happen if recyclable municipal solid waste were to be sorted at a centralized plant rather than by households. Such a policy would obviously render individual recycling efforts superfluous; Bruvoll 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, the opportunity cost of time with respect to household recycling should be strictly positive. What Bruvoll and Nyborg (2004) are saying is that, while activities such as household waste sorting may *appear* voluntary because usually 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.

Their interpretation of centralization is simply to exogenously set g^* equal to $g^0 = 0$; because only one contributing activity exists, this eliminates the need to perform *any* activity. It is not difficult to see why this might raise utility. But to some extent the result appears predicated on treating the single centralized activity as if it represented an aggregate of all activities. In trying to extend their argument to our multi-activity model, issues arise that are not discussed by Bruvoll and Nyborg (2004).

First, it is not clear that g^* should change. Certainly, we cannot simply set it to zero if only one of many activities is centralized. In fact g^*, if exogenous, may well reflect concepts that are independent of the number of available activities. In the carbon-footprint example of Section 3, ideal per-capita emissions were derived from the atmospheric CO_2 concentration considered safe, a quantity that does not change if an activity is centralized.[21] To reflect this fact, we will shift not g^* but the relevant productivity parameter θ_k.

Now arises a second difficulty. Even if g^* remains constant, centralization effectively guarantees that some level of effort will be provided on the targeted activity. Absent 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.[22] Let us make this idea a bit more precise. Suppose, as before, that N identical agents maximize

$$U(e_1, e_2) = u\left(T - e_1 - e_2\right) + v(G_{-i}) + I\left(g(e_1, e_2) - g^*\right) \tag{2.22}$$

which is just (2.2) with $K = 2$. In what follows, we will assume that solutions are unique. 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. If activity k is fully centralized, agents subsequently choose the boundary solution $e_k^1 = 0$ for all agents. To see why, note that under centralization $\theta_k^1 = 0$, implying $g'_{e_k} = 0$ everywhere. The left-hand side of the kth Kuhn-Tucker condition (2.3) for maximizing (2.22) then becomes $-u' - \lambda < 0$, which is incompatible with $e_k^1 > 0$.

Bruvoll and Nyborg (2004) interpret full centralization to mean that total production $G = Ng(e_1, e_2)$ is held fixed at its initial level. With multiple activities, however, holding overall production constant is not really possible, because by Proposition 3 individuals will increase e_l ($l \neq k$) in response to activity k being centralized. 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 θ_k^0. Under the notational convention $g = g(e_k, e_l; \theta_k)$, ex-post production will then be $G_{-i} = (N-1)g(e_k^0, e_l^1; \theta_k^0)$.

The issue at hand concerns what production level to feed into the self-

[21] 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).

[22] This issue is also recognized by Johansson-Stenman and Svedsäter (2012); like us, they lean towards the latter position.

image function. If windfalls are fully image-relevant, this should be $g(e_k^0, e_l^1; \theta_k^0)$, which is a single identical agent's share of total production. Then the conclusion of Bruvoll and Nyborg (2004) — that centralization increases each agent's utility — is unsurprisingly borne out:

Proposition 4. *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 $U(0, e_l^1) \geq U(e_k^0, e_l^0)$.*

We would argue, however, that there is more to the story. 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.[23]

There are implications for utility. Centralization now has an ambiguous effect, which is especially illuminating to view 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 $G_{-i} = (N-1)g(e_k, e_l; \theta_k^0)$. θ_k is now 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 θ_k^0. If windfalls are not image-relevant, the resulting utility function is formally identical to (2.22), and

[23] We suggest that the image-relevance of windfalls depends on how agents themselves frame the issues. 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 by others in her name may spill over into her personal self-image. For example, when we express the CO_2 problem in terms of individual carbon footprints, windfalls might seem intuitively image-relevant. But other frames are less individual-centred; for example, the view that environmental activities are a way, not to ease personal culpability, but to assist in the task for society as a whole to transition towards 'sustainability'. In essence, that a given public good can be viewed in 'share' terms does not imply that people do.

at an interior solution the envelope theorem gives

$$\frac{dU}{d\theta_k} = v'\frac{dG_{-i}}{d\theta_k} + I'g'_{\theta_k} = (N-1)v'g'_{e_l}\frac{de_l}{d\theta_k} + I'g'_{\theta_k} \tag{2.23}$$

If Proposition 3 holds, the sign of this expression is indeterminate, and it is not difficult to construct examples confirming that (2.23) may indeed have either sign.[24] Thus, centralization (lowering θ_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 θ_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.

Thus, with multiple activities, we may conceivably be dealing with a kind of second-best problem. 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 towards it — by centralizing a single activity — may lower aggregate utility instead.[25] This conclusion is further strengthened by our having excluded public-sector costs of performing the centralized activity; had we included those costs as additional tax payments, utility in (2.23) would be further reduced.

[24]Starting from the simple utility function (2.9), expression (2.23) translates into

$$\frac{\partial U}{\partial \theta_k} = 1 - (N-1) * \frac{2a^2\theta_k\theta_l^2 g^*}{(1 + a(\theta_1^2 + \theta_2^2))^2}$$

If for instance $a = 0.07$, $\theta_1 = 3$, $\theta_2 = 1$, and $g^* = 1$, this expression is positive for $N < 100$ and negative otherwise.

[25]As our analysis suggests that some agents may be willing to pay to increase individual responsibility for recycling, it also seems premature to conclude that the opportunity cost of time spent recycling should be strictly positive. A number of survey studies (Bartelings and Sterner, 1999; Bruvoll et al., 2002; Berglund, 2006) have attempted to estimate this cost by asking whether people would prefer for waste sorting to be done by households themselves, or by other people. For example, in Bruvoll 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. In light of our analysis, we suggest it may be worthwhile to drop that restriction.

2.6 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, that g is strictly quasiconcave, 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

$$W(e_1, e_2) = NU(e_1, e_2) = N\Big(u\big(T - e_1 - e_2\big) + Ng(e_1, e_2)\Big) \qquad (2.24)$$

subject to $0 \leq e_1 + e_2 \leq T$. Call (2.24) 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. Clearly, solutions to (2.24) are pairs of effort levels. We denote such a solution

(e_1^*, e_2^*); in addition $g^* = g(e_1^*, e_2^*)$ by construction. No matter which effort vector is chosen, it will result in $g = g^*$, so I does not appear in (2.24). 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 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

$$U(e_1, e_2) = u\big(T - e_1 - e_2\big) + G_{-i} + g(e_1, e_2) \\ + I(g(e_1, e_2) - g^*) \quad (2.25)$$

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

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

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 (2.10) yields[26]

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

[26]Recall that unlike (2.25), problem (2.9) had agents approximating their own contribution to overall emissions by zero. Relaxing this assumption does not affect the sign of spillovers in the present example.

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

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

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 (2.24) and (2.25), 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 $\tilde{e}_l = -e_l$.[27]

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 θ_k produces the following system of equations:

$$\left(\mathbf{u}'' + N\mathbf{H_g}\right)\mathbf{e}_\theta^* = -N\mathbf{g}_{\mathbf{e}\theta}'' \tag{2.26}$$

Here $\mathbf{u}''$ is a 2×2 matrix for which every element is u'', $\mathbf{H_g}$ is the Hessian matrix of g, and $\mathbf{g}_{\mathbf{e}\theta}''$ 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 (2.26) has a solution if and only if the matrix $\mathbf{u}''+N\mathbf{H_g}$ is invertible; and

$$\begin{aligned} |\mathbf{u}'' + N\mathbf{H_g}| = Nu''\left(g_{e_1e_1}'' + g_{e_2e_2}'' - 2g_{e_1e_2}''\right) \\ + N^2\left(g_{e_1e_1}''g_{e_2e_2}'' - \left(g_{e_1e_2}''\right)^2\right) \geq 0 \end{aligned} \tag{2.27}$$

by the concavity of g.[28] In any reasonably well-behaved problem, this

[27]Using implicit-function methods to solve the exogenous-ideal model under the assumptions used in this section produces unambiguous crowding-out.

[28]Strict concavity implies not only $g_{e_1e_1}''g_{e_2e_2}'' - (g_{e_1e_2}'')^2 \geq 0$, but $g_{e_1e_1}'' + g_{e_2e_2}'' -$

determinant is nonzero.[29] 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 (2.25) 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 θ_k to produce the implicit-function system

$$\begin{aligned}\left(\mathbf{u}'' + (1 + I')\mathbf{H}_\mathbf{g} + I''\nabla_\mathbf{e}\mathbf{g}(\nabla_\mathbf{e}\mathbf{g})'\right)\mathbf{e}_\theta = -((1 + I')&\mathbf{g}''_{\mathbf{e}\theta} \\ &+ I''\nabla_\mathbf{e}\mathbf{g}\left(\nabla_\theta\mathbf{g} - \mathbf{g}^*_\theta\right)')\end{aligned} \quad (2.28)$$

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

$2g''_{e_1e_2} \leq 0$ as well. To see why, note that the former condition implies

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

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

[29] 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.

an interior solution the determinant of this matrix is equal to

$$(1+I')\left(g''_{e_1e_1}+g''_{e_2e_2}-2g''_{e_1e_2}\right)\left(u''+(g'_{e_1})^2I''\right) \\ +(1+I')^2\left(g''_{e_1e_1}g''_{e_2e_2}-(g''_{e_1e_2})^2\right)\geq 0 \quad (2.29)$$

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'_{θ_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 (2.24) exists. Consider a small shift in productivity parameter θ_k at that solution and suppose that g is strictly quasiconcave, with $g''_{e_ke_k}\leq g''_{e_1e_2}\leq 0$ for $k=1,2$. If a solution to system (2.26) 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'_{θ_k} at the utility-maximizing point. We will now show that given our assumptions on $g''_{e_1e_2}$, Lemma 2 applies to the utility-maximizing point as well.[30]

[30]Because of this, the endogenous-ideals counterpart to (2.23) in Section 5 is neg-

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 $k = 1, 2$)

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

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

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

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 $g''_{e_1 e_2} \geq g''_{e_l e_l}$. 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

ative:

$$\frac{dU}{d\theta_k} = (N-1)v' g'_{e_l} \frac{de_l}{d\theta_k} + I'\left(g'_{\theta_k} - \frac{dg^*}{d\theta_k}\right) < 0$$

Thus, the conclusion that centralization may decrease welfare applies only to exogenous ideals.

$g''_{e_k\theta_k} > 0$,

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

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 5. *Suppose a unique and interior solution to (2.24) exists, as well as to (2.25). 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 (2.26) and (2.28) both have a solution, then for* $l \neq k$,

A. $\frac{de_k}{d\theta_k} > 0$.

B. $\frac{de_l}{d\theta_k} \gtreqless 0$.

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.

When should we expect spillover results from the endogenous, rather than the exogenous, model to apply? It may be fruitful to look for answers to this question within the theory itself because, as the proof of Proposition 5 reveals, the difference in spillovers between the exogenous and the endogenous case is given by $dg^*/d\theta_k$, or more precisely by

$$\frac{g_1'\left(g_{e_k e_k}'' - g_{e_1 e_2}''\right)(1+I')I''}{|\mathbf{u}'' + (1+I')\mathbf{H_g} + I''\nabla_{\mathbf{e}}\mathbf{g}(\nabla_{\mathbf{e}}\mathbf{g})'|} \cdot \frac{dg^*}{d\theta_k} \tag{2.30}$$

where the determinant is given by (2.29), and

$$\frac{dg^*}{d\theta_k} = g_{e_1}' \frac{de_1^*}{d\theta_k} + g_{e_2}' \frac{de_2^*}{d\theta_k} + g_{\theta_k}' = \frac{N^2 g_{e_1}' g_{e_k \theta_k}'' \left(g_{e_1 e_2}'' - g_{e_l e_l}''\right)}{|\mathbf{u}'' + N\mathbf{H_g}|} + g_{\theta_k}' \tag{2.31}$$

with the determinant given by (2.27). All else equal, differences between the models grow larger (and, by (2.37), 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 (2.30) and (2.31) 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, (2.31) 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 (θ_k increases). Where should we expect the perceived ideal to increase the most? According to (2.31), 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 (2.30) and (2.31) 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 θ_k increases, agents will tend to increase e_k in response. The resulting drop in leisure is costlier for larger (less 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.

2.7 Earlier results: economic incentives

Brekke et al. (2003) argue that their single-activity model with endogenous ideals can accommodate evidence (summarized in Frey and Jegen, 2001) on so-called 'motivation crowding', namely that monetary incentives towards activities that are driven by immaterial concerns may sometimes *decrease*, rather than increase, participation. They model the introduction of a fee which is paid (only) by non-contributors, and show that if it is sufficiently large to buy the relevant amount of the public good on the market, agents may prefer that solution and cease making efforts of their own.

As in Section 5, however, the argument appears in part to be predicated on interpreting a single activity as an aggregate of all activities. Brekke et al. (2003) focus on the case where introduction of the fee causes the relevant activity-specific ideal to drop to zero, but the ideal contribution in our model is common to all activities, and with many activities it seems unlikely that imposing a fee to finance a single activity should remove *all* motivation to contribute. Thus, the applicability of their argument to our multi-activity framework may be limited, as we will now show.

Suppose there are two activities A and B, each with strictly positive productivity; we write individual production as $g = g(e_A, e_B)$. In the original treatment, the single activity was interpreted as volunteer work by members of a sports club, such as building a new soccer field. Here, suppose that members can volunteer to build the new field but can also participate in renovating the club house, which will be done on a different date in the same month. Hence, members can volunteer to participate in either, both, or none of these activities.

To keep things as simple as possible, we follow Brekke et al. (2003) in making effort variables binary, so $(e_A, e_B) \in \{0,1\}^2$. It will be convenient to introduce some notation, namely $g^{AB} = g(1,1)$, $g^A = g(1,0)$, and $g^B = g(0,1)$. Let us suppose that $g^A \neq g^B$ and that, without loss of generality, $g^A > g^B$. For example, all else being equal, a day spent renovating the club house may benefit members more than a day spent building the soccer field. Finally, suppose for simplicity that $g^0 = g(0,0) = 0$. Then, because g is strictly increasing, we have $g^{AB} > g^A > g^B > 0$.

As in the previous section, the ideal g^* is determined through hypothetical welfare maximization, in this case of

$$W(e_A, e_B) = N\left(u(x, T - e_A - e_B) + Ng(e_A, e_B)\right)$$

where we suppose $T \geq 2$. Note that each identical agent's embedded utility function u now depends not only on leisure, but on wealth x, in which it is strictly increasing. In the absence of a fee, wealth is assumed fixed at some $m > 0$. The subsequent ideal may take three values: g^{AB}, g^A, and zero. $g^* = g^B$ can never maximize welfare, because by definition $W(0,1) = N\left(u(m, T-1) + Ng^B\right) < N\left(u(m, T-1) + Ng^A\right) = W(1,0)$. Consequently, any of the above three candidates needs only to do better than the other two to be optimal. For example, $g^* = g^{AB}$ if

$$\begin{cases} u(m, T-2) + Ng^{AB} > u(m, T-1) + Ng^A \\ u(m, T-2) + Ng^{AB} > u(m, T) \end{cases}$$

where the right-hand sides of the inequalities correspond to $g^* = g^A$ and $g^* = 0$, respectively; similar pairs of inequalities need to hold for these other candidates to be optimal. It should be clear that the three welfare maxima are mutually exclusive.

As usual, once an ideal has been determined, agents maximize utility, here given by

$$U = u(x, T - e_A - e_B) + G_{-i} + g(e_A, e_B) + I(g(e_A, e_B) - g^*)$$

We may then observe the following.

Proposition 6. *Suppose that the ideal is set at some level $g^k \in \{g^{AB}, g^A, 0\}$, and that in the resulting equilibrium all agents choose production level $g^l \neq g^k$. In the absence of a fee, no such equilibrium candidate can have*

A. $g^l > g^k$.

B. $g^l = g^B$.

Part A, which is in the spirit of Proposition 1, states that no one will end up making more effort than is ideal. Part B states that the outcome where everyone produces g^B can never be an equilibrium, no matter what the ideal contribution is. Careful examination reveals that all six equilibria not ruled out by Proposition 6 are in fact feasible.[31] Also,

[31] This was checked numerically. As all conditions on welfare and utility are linear, existence can be said to depend on whether in a linear-programming problem the feasible set corresponding to a given equilibrium is nonempty. We used the MATLAB linear-programming solver to identify feasible points for any integer N between 2 and 10000, while adding a total of 33 conditions reflecting our assumptions on u, g and I. With $\Delta I_{g-g^*} > 0$ denoting the self-image loss from producing g instead of ideal g^*, these conditions included non-negativity constraints as well as e.g. $u(m,T) >$

for a given ideal contribution, it can be shown that the equilibrium is always unique.

Now, suppose an *activity-specific* fee c_A is introduced with the aim of financing effort on activity A (e.g. renovating the club house). The magnitude and purpose of this fee is known to all agents, and it is 'sufficient' in the sense that if an agent pays it, production will be as if that agent had chosen $e_A = 1$. Because all agents who do not voluntarily set $e_A = 1$ pay the fee, its introduction guarantees that at least g^A will be provided.

Any of the four production levels may now become ideal; specifically, we may now have $g^* = g^B$, because $W(0,1) = u(m - c_A, T-1) + Ng^{AB} > u(m, T-1) + Ng^A = W(1,0)$ is quite possible (recall that g^* is defined as the amount which is ideal for agents to provide themselves, i.e. without relying on the fee).[32] As a result, any candidate for the ideal is defined by *three* inequalities. For example, $g^* = g^{AB}$ in the presence of a fee if

$$\begin{cases} u(m, T-2) + Ng^{AB} > u(m, T-1) + Ng^A \\ u(m, T-2) + Ng^{AB} > u(m - c_A, T-1) + Ng^{AB} \\ u(m, T-2) + Ng^{AB} > u(m - c_A, T) + Ng^A \end{cases}$$

where the right-hand side of these inequalities correspond to $g^* = g^A$, $g^* = g^B$, and $g^* = 0$, respectively. There are corresponding triplets of inequalities for these other candidates; again, it should be clear that the cases are mutually exclusive.

We then have the following.

$u(m,T-1) > u(m,T-2)$; $g^{AB} > g^A$; $\Delta I_{0-AB} > \Delta I_{B-AB} > \Delta I_{A-AB}$; and $\Delta I_{0-AB} > \Delta I_{0-A} > \Delta I_{0-B}$. For all equilibria not ruled out by Propositions 6 and 7, and all values of N, the MATLAB solver was able to produce a feasible point. Further details are available on request.

[32] Also note, with respect to the discussion of windfalls in Section 5, that we will now assume that whether one pays the fee makes no difference for self-image. Thus, an agent who chooses $e_A = 0$ and $e_B = 1$ will produce g^{AB}, but enjoy self-image corresponding only to g^B. The same assumption is made implicitly in Brekke et al. (2003).

Proposition 7. *Suppose that in the presence of a fee, the ideal is to voluntarily contribute some* $g^k \in \{g^{AB}, g^A, g^B, 0\}$*, with corresponding effort vector* $\mathbf{e^k} = (e_A^k, e_B^k)$*, and that in the resulting equilibrium all agents choose some effort vector* $\mathbf{e^l} = (e_A^l, e_B^l) \neq \mathbf{e^k}$*. If then, for any effort vector* $\mathbf{e}$*,* $\hat{g}(\mathbf{e})$ *is the amount produced by each agent when the effect of the fee is included, no equilibrium candidate can have*

$$\hat{g}(\mathbf{e^l}) \geq \hat{g}(\mathbf{e^k})$$

Application of Proposition 7 shows that, for example, when $g^* = g^A$ we cannot have $g = 0$ because $\hat{g}(1,0) = \hat{g}(0,0) = g^A$. There are, however, eight feasible equilibria that are not ruled out. Again it can be shown that for a given ideal, all equilibria are mutually exclusive.

Proposition 7 has the important implication that one of the post-fee equilibria involves members producing more than the ideal amount themselves: when $g^* = g^B$ it is quite possible that all agents produce g^A themselves. Intuitively, even if agents collectively think that it is better to perform activity B themselves when A can be financed by a fee, individual agents may have a strong aversion towards paying the fee and may consider g^A close enough to g^B to prefer activity A in practice. A kind of free-riding results, with agents producing $\hat{g} = g^A$ rather than $\hat{g} = g^{AB}$.

We will now reinterpret, for the multi-activity case, the 'weakly decreasing participation' concept that Brekke et al. (2003) use. Being interested in the activity A for which the fee is introduced, we write efforts on A in the pre-fee equilibrium as e_A^0, and e_A^1 for the post-fee equilibrium. Then, participation in A weakly decreases if there is a pair of ideals (one before, and one after the fee is introduced) such that all pairs of equilibria that are possible given those ideals have $e_A^0 \geq e_A^1$. For example, if initially $g^* = g^{AB}$ and subsequently $g^* = g^A$, it is quite possible that $g = 0$ initially and $g = g^A$ with the fee (so $e_A^0 = 0 < 1 = e_A^1$); thus, this pair of ideals does *not* lead to weakly decreasing effort.

Now, note that all initial levels of the ideal feature $g = 0$ as a possible equilibrium. It follows that weakly decreasing efforts in A arise for those post-fee ideals that admit only $g = 0$ and/or $g = g^B$ as equilibria (so $e^1_A = 0$). Yet since we have seen that $g = g^A$ is possible for $g^* = g^B$, a necessary condition for weakly decreasing efforts in A is that we have $g^* = 0$ after the fee is introduced, leaving three possibilities.

One: we have $g^* = 0$ both before and after the fee is introduced. This case is not very interesting, because it simply means that the representative club member finds none of the activities worthwhile, regardless of how they are financed.

Two: $g^* = g^A$ before the fee, and $g^* = 0$ subsequently. This case closely resembles the single-activity treatment in Brekke et al. (2003), because with or without the fee, agents do not find activity B welfare-improving. But then B is irrelevant, and there is no real reason to include it in the model.

Three: $g^* = g^{AB}$ initially and $g^* = 0$ with the fee. Here the fee has significant side effects: not only does the agent prefer activity A to be financed by a fee, but its introduction actually removes the image motivation to perform activity B as well. This situation grows less plausible the more activities there are. Suppose N activities exist and that before the fee, the ideal is to perform all of them. If now we introduce a fee to finance some activity, this case requires that $g^* = 0$ afterwards. It seems very unrealistic that all motivation to contribute should vanish because of a single activity-specific fee.

One particularly interesting combination of ideals is absent from the above list: if the fee on A affects *only* the motivation to perform A, we could have $g^* = g^{AB}$ before, and $g^* = g^B$ after it is introduced. This is arguably the closest multi-activity counterpart to the situation that Brekke et al. (2003) are interested in. But we have seen that it does not rule out $e^0_A = 0$ and $e^1_A = 1$: even if activity A is fee-financed, and only producing g^B is ideal, there may exist an equilibrium where everyone instead participates in A. Hence, this situation does not drive weakly decreasing participation in A, making it doubtful whether the prediction that a sufficient fee may weakly lower participation can be said to apply to our multi-activity setting.[33]

2.8 Discussion: why do results differ?

Sections 5 and 7 have presented examples indicating that certain results that are clear-cut in single-activity models become less so in the multi-activity setting. The point should not be overstated, however, because the multidimensionality of our model is not the only factor at work here. Consider a plausible alternative to 'self-image' function (2.1), namely

$$S(\mathbf{e}) = I(e_1 - e_1^*, ..., e_K - e_K^*) \tag{2.32}$$

where for $k = 1, ..., K$, $e_k^* \geq 0$ are *activity-specific* ideal contributions. In most other ways the self-image function is similar to (2.1): 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 $e_k > e_k^*$; and $I_{kk}'' \leq 0$ everywhere. The core difference lies in whether best efforts are good enough. Function (2.1) is a consequentalist'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, (2.32) 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 $\mathbb{R}^K$.

These self-image models can also be thought of as representations 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

[33] This conclusion is not affected if the fee is financing activity B. Then, again eight post-fee equilibria are not ruled out by Proposition 7, and in particular if $g^* = g^A$ we may have $g = g^B$. The discussion of weakly decreasing efforts will be identical to that above.

judged solely on its own merits, or actively compared with other options). The differences between the two multi-dimensional self-image functions arguably capture something of the flavor of both, since real-world contributions are often made as separate choices (joint/separate evaluation) and at different points in time (choice bracketing). Function (2.1), then, imposes broad partitioning, while (2.32) is based on narrow partitioning. However, while both strategies can be represented in a multi-dimensional framework, narrow partitioning is imposed by construction within a one-dimensional model that is interpreted in terms of a single non-aggregate activity. This is because, whether the self-image function is production-based (e.g. Bruvoll and Nyborg, 2004) or effort-based (Brekke et al., 2003), there exists only one activity, so comparisons between activities are omitted from the analysis.

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, a model based on (2.32) would do much better at replicating the previous results we have revisited in this paper. With respect to Section 5, while for model (2.1) it appeared not to make sense to set $g^* = 0$ when only activity is k being centralized, we have no objection to setting $e_k^* = 0$ within a model based on (2.32). Similarly, in Section 7, we argued that an activity-specific fee on activity k seemed unable to guarantee reduced effort on that activity; but with activity-specific ideals we could have $e_k^* = 1$ initially and $e_k^* = 0$ with the fee, in which case there would certainly be weakly decreasing efforts made on k.

Indeed, the outlook we adopt arguably differs from that of most papers in the 'duty-orientation' literature. These often stress that moral responsibility, even for single non-aggregate activities, should be seen as a burden which individuals would prefer not to bear (Bruvoll and Nyborg, 2004; Brekke et al., 2010).[34] In our model, by contrast, an overarch-

[34]For example, Nyborg (2011) shows that a single-activity version of the self-image model is consistent with evidence Dana et al. (2007) 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.

ing motivation to contribute is taken for granted; and while individuals would indeed 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?' Because each individual activity extends the choice set available for approaching the overall ideal, they make agents better off, not worse.

There are solid semi-technical reasons to prefer the function (2.1) over (2.32). For example, if the effort-based ideals in (2.32) are exogenous, agents have no reason to change their contribution in response to changes in (perceived) productivity. Also, in a model based on (2.32), the sign of behavioral spillovers depends on the sign of the mixed-partial derivatives of $I(e_1 - e_1^*, ..., e_K - e_K^*)$, which are much more difficult to interpret in a meaningful way than $g''_{e_1 e_2}$.[35] The real issue, however, concerns whether moral behavior is 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.

2.9 Concluding remarks

Much of this paper has been concerned with analysis of productivity-led spillover effects from policy interventions that target a particular prosocial activity. We solve the model for two cases: one where the 'ideal contribution' considered by agents is exogenous, and one where it is en-

[35] An experiment by Cornelissen et al. (2013) suggests that people's ethical mindset (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 (2.1); showing that (2.32) leads to positive spillovers requires additional assumptions on I.

dogenously determined through moral introspection. In the former case, our results point to crowding-out of effort on other activities, similar to the 'moral licensing effect', as the benchmark case. In the latter, the model fails to provide a clear prediction on the sign of spillovers.

We have also revisited certain results from published papers where the single-activity model is not interpreted in terms of aggregate contributions but rather in terms of a particular, non-aggregate activity. In such cases, results are sometimes difficult to reproduce in our multidimensional framework, and this appears to be an artifact of the particular manner in which we generalize the single-activity model. Nevertheless, as broad partitioning of choice options can only be modelled in a multi-activity framework, we would argue that there is some need to explicitly consider multiple activities in future research on 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.

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.

S. Brown, M.N. Harris, and K. Taylor. Modelling charitable donations to an unexpected natural disaster: Evidence from the U.S. Panel Study of Income Dynamics. *Journal of Economic Behavior & Organization*, 84(1), 2012.

A. Bruvoll and K. Nyborg. The Cold Shiver of Not Giving Enough: On the Social Cost of Recycling Campaigns. *Land Economics*, 80(4), 2004.

A. Bruvoll, B. Halvorsen, and K. Nyborg. Households' recycling efforts. *Resources, Conservation and Recycling*, 2002.

L. Burke, K. Reytar, 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.

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.

L. Festinger. *A theory of cognitive dissonance.* Stanford University Press, 1957.

B.S. Frey and R. Jegen. Motivation Crowding Theory. *Journal of Economic Surveys*, 15(5), 2001.

A. Gneezy, A. Imas, A. Brown, L.D. Nelson, and M.I. Norton. Paying to Be Nice: Consistency and Costly Prosocial Behavior. *Management Science*, 58(1), 2012.

A.E. Greenberg. On the complementarity of prosocial norms: The case of restaurant tipping during the holidays. *Journal of Economic Behavior & Organization*, 97, 2014.

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.

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.

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.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

Appendix A. Monotone comparative statics and total efforts

Here we show that the set of optimal total efforts is monotone nondecreasing in θ_k. Our starting point is the problem (2.19). 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. Simply using differentiability of $\hat{g}$ produces the sufficient single-crossing condition

$$\frac{\partial^2 U_{\hat{e}}}{\partial \hat{e} \partial \theta_k} = I'' \left(\frac{\partial \hat{g}}{\partial \theta_k} \frac{\partial \hat{g}}{\partial \hat{e}} \right) + I' \frac{\partial}{\partial \theta_k} \left(\frac{\partial \hat{g}}{\partial \hat{e}} \right) \geq 0 \qquad (2.33)$$

Now, note that both $\hat{e}$ and θ_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, of course, 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 λ is the Lagrange multiplier associated with the restriction $e_1 + e_2 - \hat{e} = 0$. If the initial solution to (2.19) had $e_k > 0$, the corresponding necessary Kuhn-Tucker conditions for the PMP imply $\lambda = g'_{e_k}$.[36] Then, $\partial/\partial \theta_k (\partial \hat{g}/\partial \hat{e}) = \partial/\partial \theta_k (g'_{e_k}) = g''_{e_k \theta_k}$, and condition (3.5) is transformed into exactly (2.15).

As for the relationship between g^* and total efforts, we obtain the fol-

[36]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.

lowing from (2.19):

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

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

In summary, if (2.15) holds and $e_k > 0$ initially, optimal total efforts $\arg\max_{\hat{e}} U_{\hat{e}}(\hat{e}; \theta_k)$ are monotone nondecreasing in θ_k as well as in g^*.

Appendix 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 (2.3) 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. □

Proof of Lemma 1

Proof. Suppose $\mathbf{e}' = (e'_1, e'_2)$ solves (2.2) but that $g(\mathbf{e}')$ does not satisfy (2.18) at production level $\hat{e}' = e'_1 + e'_2$. Then there must exist some other vector $\mathbf{e}''$ with $\hat{e}'' = e''_1 + e''_2 = \hat{e}'$ but, by Proposition 1, $g^* \geq g(\mathbf{e}'') > g(\mathbf{e}')$; and then

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

so $\hat{e}'$ cannot solve (2.2).

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 (2.18) for their respective production levels. Now, consider the effort vector $(e'_1 + \hat{e}'' - \hat{e}', e'_2)$. Its sum of efforts is $e'_1 + \hat{e}'' - \hat{e}' + e'_2 = \hat{e}''$, yet if g is strictly increasing,

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

so $\mathbf{e}''$ cannot satisfy (2.18) after all. $\square$

Proof of Proposition 4

Proof. Note that

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

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, the above expression 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 (2.3) that $\lambda \geq 0$, with $\lambda = 0$ if $e_k + e_l < T$ (complementary slackness). Thus, if $e_l^1 = e_k^1 + e_l^1 < T$, ex-ante and ex-post Lagrange multipliers (denoted λ^0, λ^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 $e_l^1 < e_l^0$. □

Proof of Lemma 2

Proof. The solution to system (2.26), if it exists, is

$$\mathbf{e}_\theta^* = -N \left(\mathbf{u}'' + N\mathbf{H}_\mathbf{g}\right)^{-1} \mathbf{g}''_{\mathbf{e}\theta} \tag{2.34}$$

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

$$\frac{dg^*}{d\theta_k} - g'_{\theta_k} = g'_{e_1}\frac{de_1^*}{d\theta_k} + g'_{e_2}\frac{de_2^*}{d\theta_k} = \frac{N^2 g'_{e_1} g''_{e_k\theta_k}\left(g''_{e_1e_2} - g''_{e_le_l}\right)}{|\mathbf{u}'' + N\mathbf{H}_\mathbf{g}|}$$

where the second equality uses the first-order conditions for an interior optimum. This expression is non-negative for $g''_{e_1e_2} \geq g''_{e_le_l}$. □

Proof of Proposition 5

Proof. The solution to system (2.28), if it exists, is

$$\begin{aligned} \mathbf{e}_\theta = -\left(\mathbf{u}'' + (1+I')\mathbf{H}_\mathbf{g} + I''\nabla_\mathbf{e}\mathbf{g}(\nabla_\mathbf{e}\mathbf{g})'\right)^{-1} \\ \left((1+I')\mathbf{g}''_{\mathbf{e}\theta} + I''\nabla_\mathbf{e}\mathbf{g}\left(\nabla_\theta\mathbf{g} - \mathbf{g}_\theta^*\right)'\right) \end{aligned} \tag{2.35}$$

For part A, we use (2.35) to find

$$\frac{de_k}{d\theta_k} = \frac{1+I'}{|\mathbf{u}'' + (1+I')\mathbf{H_g} + I''\nabla_{\mathbf{e}}\mathbf{g}(\nabla_{\mathbf{e}}\mathbf{g})'|} \cdot \Bigg(g''_{e_k\theta_k}\left(-u'' - g''_{e_le_l}(1+I') - \left(g'_{e_1}\right)^2 I''\right) + g'_{e_1}\left(\frac{dg^*}{d\theta_k} - g'_{\theta_k}\right)\left(g''_{e_le_l} - g''_{e_1e_2}\right) I''\Bigg) > 0 \quad (2.36)$$

because $g''_{e_1e_2} \geq g''_{e_le_l}$.

For part B, we use (2.35) to also calculate

$$\frac{de_l}{d\theta_k} = \frac{1+I'}{|\mathbf{u}'' + (1+I')\mathbf{H_g} + I''\nabla_{\mathbf{e}}\mathbf{g}(\nabla_{\mathbf{e}}\mathbf{g})'|} \cdot \Bigg(g''_{e_k\theta_k}\left(u'' + g''_{e_1e_2}(1+I') + \left(g'_{e_1}\right)^2 I''\right) + g'_{e_1}\left(\frac{dg^*}{d\theta_k} - g'_{\theta_k}\right)\left(g''_{e_ke_k} - g''_{e_1e_2}\right) I''\Bigg) \lesseqgtr 0 \quad (2.37)$$

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 (2.36) and (2.37) we find

$$\frac{de_1}{d\theta_k} + \frac{de_2}{d\theta_k} = \frac{1+I'}{|\mathbf{u}'' + (1+I')\mathbf{H_g} + I''\nabla_{\mathbf{e}}\mathbf{g}(\nabla_{\mathbf{e}}\mathbf{g})'|} \cdot \Bigg(g''_{e_k\theta_k}\left(g''_{e_1e_2} - g''_{e_le_l}\right)(1+I') + g'_{e_1}\left(\frac{dg^*}{d\theta_k} - g'_{\theta_k}\right)\left(g''_{e_1e_1} + g''_{e_2e_2} - 2g''_{e_1e_2}\right) I''\Bigg) \geq 0$$

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

Proof of Proposition 6

Proof. For part A, suppose first that production g^k corresponds to some leisure level L^k and that g^l corresponds to L^l. One of the equations defining $g^* = g^k$ is

$$u(m, L^k) + Ng^k > u(m, L^l) + Ng^l$$

and one of the equations specifying g^l as an equilibrium given $g^* = g^k$ is

$$u(m, L^l) + Ng^l - \Delta I_{l-k} > u(m, L^k) + (N-1)g^l + g^k$$

so if everyone else is choosing production level l, then it must be better to choose production l than k. Here $\Delta I_{l-k} > 0$ is the self-image loss that arises when g^l is produced instead of the ideal amount g^k. But combining these two inequalities yield

$$(N-1)(g^k - g^l) > \Delta I_{l-k} > 0$$

Thus, if $g^l > g^k$, we have a contradiction.

For part B, even if the ideal is some $g^* = g^k > g^B$ (i.e. if it is $g^* = g^{AB}$ or $g^* = g^A$), one of the conditions that must be satisfied for $(e_A, e_B) = (0, 1)$ to be a best response is

$$u(m, T-1) + Ng^B - \Delta I_{B-k} > u(m, T-1) + (N-1)g^B + g^A - \Delta I_{A-k}$$

so if everyone else is choosing $(e_A, e_B) = (0, 1)$, then it must be better to choose $(e_A, e_B) = (0, 1)$ than $(e_A, e_B) = (1, 0)$. (Of course, if $g^k = g^A$, $\Delta I_{A-k} = 0$.) Now, since $g^A > g^B$, $\Delta I_{B-k} > \Delta I_{A-k}$. But then the above inequality translates into $g^B - \Delta I_{B-k} > g^A - \Delta I_{A-k}$, which is a contradiction. $\square$

Proof of Proposition 7

Proof. The proof is quite similar to that of Part A of Proposition 6.

Again we associate production g^k with some leisure level L^k and g^l with L^l. Now, $g^* = g^k$ requires for instance

$$u(m-(1-e_A^k)c_A, L^k) + N\hat{g}^k > u(m-(1-e_A^l)c_A, L^l) + N\hat{g}^l$$

and one of the equations specifying g^l as an equilibrium given $g^* = g^k$ is

$$u(m-(1-e_A^l)c_A, L^l)+N\hat{g}^l-\Delta I_{l-k} > u(m-(1-e_A^k)c_A, L^k)+(N-1)\hat{g}^l+\hat{g}^k$$

Combining these two inequalities yields

$$(N-1)\left(\hat{g}^k - \hat{g}^l\right) > \Delta I_{l-k} > 0$$

so if $\hat{g}^l \geq \hat{g}^k$, we have a contradiction. $\square$

Chapter 3

Some Causes are More Equal than Others? The Effect of Similarity on Spillovers in Charitable Giving

3.1 Introduction

People who set out to champion some prosocial cause — be it with respect to their fellow man, their community, the environment, or some abstract ideal — are likely to find that there is more than one way to contribute to that cause. This is perhaps especially apparent in charitable giving, where the number of charities engaging with a given cause is typically very large.[1] Yet even when there are several ways to contribute,

[1]In the United States, for example, charitable organizations for which donations are tax deductible number close to a million. The full list is posted at `http://apps.irs.gov/app/eos/forwardToPub78Download.do` (last accessed on November

interventions to increase contributions may be specific to one prosocial alternative. For example, charities commonly offer matched donations, where a wealthy donor matches incoming contributions dollar-for-dollar or by some other proportion. While there is much research on how such schemes affect donations to the charity being affected (e.g. Karlan and List, 2007; Huck and Rasul, 2011), less well understood is how interventions spill over to contributions made through *other* channels. The present paper contributes to filling that gap by presenting experimental evidence on productivity-driven spillovers across different charities. The main novelty of our experiment derives from an effort to chart their *reach*: that is, do significant spillovers arise even across (very) dissimilar alternatives?

Despite the clear practical relevance of this question, no other study, as far as we are aware, has given it systematic attention. The small number of other experimental studies that analyze this type of spillovers (Reinstein, 2011, 2007; Null, 2011) have focused mainly on identifying the effect itself, typically finding that as the productivity of one charity rises, subjects spend fewer resources on other charities.[2] Such negative spillovers can be viewed as another type of 'crowding-out': one that arises not between public and private contributions, but rather between different modes of private contribution.

The natural suspicion, however, is that spillovers are less pronounced for dissimilar alternatives, with one motivation being the following. Behavioral economists have long argued that much of human decision making is characterized by 'mental accounting' (Tversky and Kahneman, 1981; Read et al., 1999; Thaler, 1999), one aspect of which involves dividing expenditures into budgets or categories (for an interpretation of mental accounts as psychological categories, see Henderson and Peterson, 1992). A defining characteristic of such budgets is that they are not fungible

9^{th}, 2015).

[2]In the simultaneous-decision framework used in these studies (and in the present paper), crowding-out becomes very similar to gross substitution across charities, with matching rates as implicit prices. Another interpretation of these experiments, then, is that they map subjects' substitution behavior within the charity domain.

(Thaler, 1985; Heath and Soll, 1996; Thaler, 1999): people are reluctant to move resources from one account to cover expenses that belong to another account, such as from food purchases to entertainment. For this reason, price shifts may have greater impact within its relevant account than across accounts. In our case, if a person has charities A and B to choose from, and the productivity of A rises, she will be less prone to shift resources from B to A if these charities contribute towards different mental accounts than if they contribute to the same one.[3]

Our experiment, which was conducted in Lund, Sweden, was a real-effort based version of the dictator game where subjects could donate money to two real-world charities by first earning 'points' in a real-effort task. The production-and-allocation decision was repeated over multiple rounds where, to mimic the effect of a charity-specific price shift, the rate of actually donated Swedish crowns (SEK) per point of each charity was systematically varied. A recurring theme in the literature on crowding-out is that the amount of 'good deeds' people are capable of within a given domain is either close to being fixed (so policies simply cause reallocation within that fixed 'budget'), or can increase as a result of a price shift. Our design allowed for both possibilities: negative spillovers follow from standard relative-price effects, but positive spillovers can also arise in principle if, as a result of increasing charity-specific exchange rates, subjects earn more points overall and allocate more to both charities.

Nevertheless, our theoretical prediction, which is derived from a multi-activity generalization (Ek, 2015) of a variant of the warm-glow model for public-goods contributions (Andreoni, 1990; Brekke et al., 2003), is one of partial crowding-out across alternatives: as the exchange rate of

[3]Note that while economists have tended to consider mental accounting a bias, this type of behavior may be entirely consistent with the standard model of consumer choice, with mental accounts simply giving structure to the person's substitution patterns. For example, it may well be rational (though, we suspect, rare) to view the distinct physical processes underlying climate change and eutrophication as represented by different public goods (and hence put them in different accounts) rather than by some monolithic 'environmental good'.

one charity rises, fewer points are spent on the other, but total donations rise. This hypothesis is clearly validated by the experimental data.

Furthermore, to test the idea that the way people categorize alternatives may explain the magnitude of subsequent spillovers, we include four between-subject treatments, isolating two dimensions along which charities may differ: (i) geographical scope, and (ii) the cause involved. In all treatments, one charity was UNICEF, a global organization concerned with the welfare of children. The other charity differed across treatments. In our baseline treatment, UNICEF was paired with another charity addressing children's needs globally. By contrast, our 'dissimilar' treatment paired UNICEF with the local office of a major Swedish environmental NGO. To fully isolate the effect of each dimension (local/global and cause involved), two intermediate treatments were also included. Our results show that while crowding-out occurs significantly in all treatments, the effect is systematically weaker, the more dissimilar are the charity alternatives. In the dissimilar treatment, it is only half as large as in the similar treatment.

It is important to note that these results may be applicable not only to charitable giving but to other kinds of prosocial behavior. Indeed, just as donation matching is charity-specific, public policy to encourage other types of prosocial behavior may be activity-specific. For instance, it is quite common for local governments to promote household recycling in various ways that increase the productivity of individuals' efforts, and this may drive an increase or a decrease in alternative environmental behaviors, such as buying organic products.

In fact, existing field studies on policy-driven 'behavioral spillovers' (Dolan and Galizzi, 2015), while few in number, have also tended to find evidence of crowding-out. Jacobsen et al. (2012) studied a green-electricity program in Memphis, Tennessee, finding that among households that participated at a minimal level (paying the smallest possible increment to fund alternative energy), consumption of (non-green) electricity increased after they had joined the program. Similarly, Tiefenbeck et al. (2013) found that a campaign to conserve water at a housing complex in Massachusetts, while leading to decreased water use, also drove increased electricity consumption. Crowding-out, then, undermines a given policy, and policy impacts will be overestimated by any cost-benefit analysis that ignores spillovers. Our results on similarity suggest that crowding-out will be larger for more similar prosocial activities.

Psychologically, spillovers in these non-charity domains tend to be explained in terms of 'moral licensing' (Cain et al., 2005; Sachdeva et al., 2009; Blanken et al., 2015), which is the tendency of people who have just behaved prosocially to feel justified in subsequently relaxing their moral constraints. The implication is that if an intervention is successful in increasing contributions through the channel being targeted, contributions through other channels may drop as a result.[4] It is therefore worth noting that in our theoretical framework, negative spillovers arise through a process of 'self-image management' which very much captures the flavor of moral-licensing accounts.[5]

The structure of this paper is as follows. Section 2 briefly sketches our

[4]The effect can also run in the opposite direction ('moral cleansing'). For example, Ploner and Regner (2013) present a series of experiments where subjects could donate to charity by participating in a real-effort task which was followed by a dictator game. Those who contributed relatively little in the initial game tended to be especially generous in the dictator game.

Some evidence also exists on an opposing consistency effect suggesting that one prosocial act can spur another. Gneezy et al. (2012) argue that if prosocial behavior is sufficiently costly, people who nevertheless engage in it are led to conclude that they are a prosocial 'type' and will subsequently try to act in accordance. In the same spirit, Brown et al. (2012) found that the more people donated to disaster relief following the 2004 Indian Ocean tsunami, the more likely they were to donate *more*, rather than less, to charity in the future.

[5]Our experiment differs from most, if not all, previous studies on moral licensing in that it involves simultaneous rather than sequential decision making. However, our interpretation of the moral-licensing effect is that it reflects the more basic process of managing one's self-image as a sufficiently moral person. Experimental manipulations (such as the amount initially donated in Ploner and Regner, 2013) work because they cause shifts in the perceived degree of moral self-worth, which in turn affects behavior in other morally charged situations. If our interpretation is correct, the pattern should generalize from sequential to simultaneous decision making, and can inform predictions for either setting.

theoretical framework and presents the resulting predictions. Section 3 describes our experimental design as well as the particularities of each treatment. Section 4 presents our main results, devoting particular effort to arguing that observed treatment effects are in fact due to differences in charity-pair similarity rather than some confounder, such as the subjects' relative preference for each charity. Finally, Section 5 concludes the paper.

3.2 Theoretical model and predictions

This section describes and motivates our experimental predictions. First, we sketch a simple warm-glow model of prosocial behavior which predicts negative spillovers. A similar version of the model analyzed here was considered in Ek (2015).

We consider a population of identical agents choosing the level of 'effort' to exert on two activities in order to contribute to a single public good G. Effort variables x_1, x_2 are measured in time units; labor supply is fixed, so any time not spent contributing is devoted to leisure (L) instead.[6] It is assumed that G is large-scale enough that each agent approximates own impacts on the overall level of the public good by zero. However, agents are warm-glow altruists in the sense of Andreoni (1990), caring about the size of their own contribution. Each agent maximizes the additively separable function

$$U(x_1, x_2) = u(T - x_1 - x_2) + I(g(\theta_1 x_1, \theta_2 x_2)) \tag{3.1}$$

subject to $0 \leq x_1 + x_2 \leq T$, where T is total non-work time. u is the utility of leisure; it is concave and strictly increasing. Each agent's personal public-good production g is strictly increasing in x_1 and x_2. The 'warm-glow' component I is concave and strictly increasing in individual production g. Note that the interpretation of g can be made somewhat

[6]While it is a simple matter to reinterpret this model in monetary terms, the present version corresponds more closely to our real-effort based experiment.

loose by arguing that it represents agents' *perception* of the production function, which may or may not be accurate. Psychologically, warm glow has been interpreted as a favorable sense of social image or self-image as a generous person (Bénabou and Tirole, 2006), the desire to adhere to a moral norm (Brekke et al., 2003), or as the desire to help other contributors feel validated (Rotemberg, 2014). As the model includes no social interaction, we suggest that it is best interpreted in terms of self-image here.

For all identical agents, $\theta_1 > 0$ and $\theta_2 > 0$ regulate the productivity of x_1 and x_2, respectively. They are assumed to reflect activity-specific interventions that affect various costs related to available technology as well as convenience and information, so long as these are relevant for how much is accomplished per unit of effort. For example, if the model is reinterpreted in terms of money, and x_1, x_2 represent donations to different charities, introducing matched donations to one of the charities would cause the corresponding θ to increase, as long as it is credible that the matched donations would not have happened anyway. As another 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 may cause the relevant θ to increase. By contrast, 'nudges' such as simply highlighting average performance in the community will not.

The fundamental trade-off in model (3.1) is standard: agents may either reduce effort and enjoy increased leisure, or increase efforts and, by extension, warm glow. However, the multidimensional production function implies an additional trade-off, between activities. Warm glow is common to both activities and measured in terms of production rather than effort, making this a consequentialist model where agents ask, 'Am I making enough of a difference?' rather than 'Am I trying hard enough?'. In the same vein, note also that while agents maximize utility with respect to effort expended (x_1, x_2), the production function is assumed to take contributions received $(\theta_1 x_1, \theta_2 x_2)$ as its arguments.

In Appendix A, we derive predictions for the effect of an activity-specific intervention. We consider a special case of (3.1), where utility is quasilinear in leisure and each agent's public-good production function is given by the symmetric CES function $g = ((\theta_1 x_1)^\rho + (\theta_2 x_2)^\rho)^{1/\rho}$. There are two main reasons for using this function rather than the approximately linear function that one might expect a single agent to face when producing a large-scale public good. First, although a linear function would imply that activities are perfect substitutes, agents have been found not to perfectly substitute even among closely related alternatives (Null, 2011). Second, this function allows us to represent the effect of similarity on the magnitude of spillovers.[7] The ρ parameter reflects the degree of substitutability between contributions received through activity 1 ($\theta_1 x_1$) and activity 2 ($\theta_2 x_2$), so if activities' perceived substitutability in the production of the public good differs by their similarity, so will the value of ρ.[8]

Focusing on the case where activities are relatively substitutable ($0 \leq \rho < 1$), we then derive the following two results.

Hypothesis 1. When a particular prosocial activity is exposed to a directed positive productivity shock, agents decrease efforts on the other activity.

Hypothesis 2. Total contributions received ($\theta_1 x_1 + \theta_2 x_2$) is increasing in the productivity of the more productive activity.

Finally, we predict a 'cross-domain effect': spillovers become attenuated for more dissimilar activities. As stressed in the introduction, this seems consistent with the tendency of people to engage in mental accounting

[7] Although we depict spillovers across activities as resulting from a nonlinear (perceived) production function, they may just as well be understood as the result of nonlinearity in a multivariate warm-glow function. Our choice here is mainly for simplicity.

[8] A disadvantage of using this function is that, unlike in Ek (2015), interventions are not truly activity-specific. This would require that each parameter has no direct effect on the other activity's productivity, so $g''_{\theta_1 x_2} = g''_{\theta_2 x_1} = 0$; for the CES function, this is true only if $\rho = 1$.

(e.g. Thaler, 1999). The model sketched above is not unequivocal on the issue, though the prediction unambiguously holds for the important benchmark case where initially both activities are equally productive ($\theta_1 = \theta_2$). Even when $\theta_1 \neq \theta_2$, there exists a possibly large interval in ρ where the prediction applies. Loosely speaking, this interval is smaller, the further apart θ_1 and θ_2 are. For the productivity values considered in the present experiment, the prediction holds at least for all $0 \leq \rho \leq 0.63$.

Hypothesis 3. (i) If initially $\theta_1 = \theta_2$, activity-specific positive productivity shocks drive less 'crowding-out', the less similar the activities are. (ii) If $\theta_1 \neq \theta_2$, there exists some $0 < \bar{\rho} < 1$ such that the same conclusion applies for any $0 \leq \rho \leq \bar{\rho}$. (iii) $\bar{\rho}$ is increasing in $a = \theta_2/\theta_1$ if $a < 1$, and decreasing in a if $a > 1$.

There is limited previous evidence on the cross-domain issue. In a similar dictator-game to ours, Reinstein (2007) lets subjects donate to three or more charities and, in shifting the relative productivity of the charities, notes that subsequent spillovers seem larger among more similar charities. Unlike us, however, he does not systematically manipulate the similarity of the charities along several dimensions. Moreover, it is not clear that the pattern he observes is due to perceived similarity per se, as he does not control for ways in which charities may differ other than in terms of similarity, and also does not check whether subjects agreed with his assessment of which charities are similar; hence, competing explanations (e.g. in terms of the relative popularity of the charities) cannot be ruled out.

Results by Blackwell and McKee (2003) also provide some suggestion that if the alternatives are dissimilar in geographical scope, spillovers may be weak. In a variant of the public-good game, the authors let subjects allocate an endowment across three alternatives: a private account, a 'group' public good and a 'global' public good. Returns from the group PG were redistributed only to each participant's particular group of four people, while returns from the global PG were redistributed to all subjects. Increasing the average per capita return on the global public good caused contributions to that good to rise, but did not decrease contributions to the group PG.

Finally, while observing that behavioral spillovers sometimes arise even across dissimilar behaviors (Miller and Effron, 2010), psychologists have hypothesized that the effect may be stronger within a particular domain (such as environmental behavior) than across domains. For example, the recent meta-analysis by Blanken et al. (2015) test this hypothesis, though they find no evidence in favor of it. There are, however, at least three methodological differences between our experiment and the moral-licensing studies that they analyze. First, while our experiment involves simultaneous decision making, psychological studies are explicitly sequential, with an initial (licensing) stage followed by a subsequent decision stage. Second, the initial stage is often hypothetical, e.g. with subjects instructed to describe themselves using positive words. Third, even when the initial stage is choice-based, what is being varied is not the incentives to behave prosocially, but whether a prosocial choice (such as disagreeing with a racist statement) was actually made. Hence, the link between these studies and intervention-driven spillovers is rather indirect.

3.3 Experimental design

The experiment was conducted over nine afternoon sessions in a computer room at the Lund School of Economics and Management (LUSEM). Visually isolated cubicle-like spaces were created by putting up cardboard screens and roll-up banners. The experimental environment could fit 20 subjects, and was fully occupied in six of the sessions. Of the remaining three sessions, one had 19 participants, one had 14, and one had 7, with $N = 160$ in total. Subjects were initially recruited through email from the pool of Swedish-speaking students at LUSEM; for each session, additional participants were recruited on-the-fly on the LUSEM premises. Each session lasted approximately an hour, after which participants were paid a show-up fee of SEK 200 (about €20) in cash.

The experimental setting was a real effort-based variant of the dictator game. Subjects were paid only a show-up fee and were informed that they could generate, with no monetary cost to themselves, donations to two organizations (henceforth, 'charities'): UNICEF Sweden, presented as 'UNICEF' to subjects; and one other (henceforth denoted 'Other'). UNICEF is a global organization concerned with the welfare of children which pools all incoming regular donations internationally. The identity of the 'Other' charity differed across treatments; it will be further described below. Donating required that subjects first earn 'points', and the amount given to charity $k \in \{UNICEF, Other\}$ was $\theta_k x_k$, where $\theta_k > 0$ is here an exchange rate (SEK/point) and $x_k \geq 0$ the number of points allocated towards k. To make donations credible, participants were informed that they would be e-mailed a receipt of the payment, and this was done within days of each session.

Following an initial practice round, the decision to earn and allocate points was repeated over six rounds, with subjects informed that only the choices made in one of those six rounds (to be randomly selected at the end of the session) would translate into actual donations. Five of the rounds were oriented towards identifying spillovers across activities; in each of these θ_{UNICEF} was held constant at 10 SEK/point, while θ_{Other} took on different values (4, 8, 10, 12, and 16 SEK/point). The remaining sixth round had $\theta_{UNICEF} = \theta_{Other} = 16$, and was introduced to examine the effect of a general (as opposed to activity-specific) productivity shock. Being irrelevant for identifying spillover effects, however, this round will not be considered in this paper (with a few clearly marked exceptions). Although all subjects encountered the same exchange rates at some point during a session, the sequence of values of θ_{UNICEF} and θ_{Other} was randomized within and between subjects.

Each round consisted of three computer screens.[9] The first screen simply presented the name and current-round exchange rate of each charity. Subjects were given no further information whatsoever on the charities. The second screen confronted subjects with a slider-adjustment task

[9]The experiment was programmed using the software z-Tree (Fischbacher, 2007).

(Gill and Prowse, 2012). With 48 sliders on the screen and an allotted time of 120 seconds, each slider set to position 50 out of a possible 100, i.e. in the very middle, earned them a single point.[10] After the set time elapsed, subjects progressed to the third screen, where they were shown the number of points they had just earned and were asked to allocate these in integer amounts across the two charities. There was no public or private feedback on this, or any other, decision. The placement of the two charities relative to each other (left/right) was randomized across rounds and subjects and had no significant effect on either the total number of points earned or the allocation across alternatives.

Our four treatments are summarized in Table 3.1. The treatments were between-subject, so each participant faced only one charity pair. The similarity of the charities was systematically varied across treatments by isolating two dimensions along which charities may differ: (i) geographical scope, and (ii) the cause involved. In all treatments, one charity was UNICEF, but the alternative (Other) differed across treatments.

Table 3.1: *Treatment summary*

Treatment	N	No. sessions	Other (non-UNICEF) charity	Gender (% male)	Age (avg.)
GLOBAL-CHILD	40	2	Save the Children Global Action Fund	47.5	23.05
LOCAL-GREEN	40	2	SSNC Lund	52.5	22.95
LOCAL-CHILD	40	3	Save the Children Lund	42.5	23.65
GLOBAL-GREEN	40	2	WWF International	50	22.65

In the GLOBAL-CHILD treatment, UNICEF was paired with the 'Global Action Fund' run by Save the Children in the US. Although the fund's

[10]Thus, referring back to equation (3.1), the experiment was designed to reflect a situation where $T = 48$ and u represents the utility of not exerting effort in the slider-adjustment task.

website[11] states that donations benefit children "in the United States and around the world", subjects were presented only with the name 'Save the Children Global Action Fund', so it is likely that the global dimension was salient. Since both charities address children's needs globally, they are similar along both dimensions, and should be relatively close substitutes.

By contrast, the LOCAL-GREEN treatment had UNICEF paired with the Lund office of the Swedish Society for Nature Conservation (SSNC), which is dissimilar along both dimensions. SSNC is the largest Swedish environmental NGO, and its local branch engages in awareness raising through lectures, workshops, and excursions into the local environment, in addition to lobbying local policy makers.

Finally, to isolate the effect of each dimension (local/global and cause involved), two intermediate treatments were also included: LOCAL-CHILD and GLOBAL-GREEN. In the former, UNICEF was paired with the local Lund office of Save the Children (local/global dimension); in the latter, with WWF International, a major global environmental NGO (cause dimension). Admittedly, the distinctions are not entirely clear-cut; in particular, as there is no need for disaster relief in Sweden, the Lund office of Save the Children focuses on education and community work (especially with respect to child refugees), and so its 'cause' might be viewed as somewhat different from UNICEF. Since this pair is nevertheless likely to be seen as more similar than UNICEF versus the Lund office of SSNC, our hypotheses are little affected.

Figure 3.1 provides a stylized view of the experimental timeline. It also shows that at the end of each session, participants responded to a questionnaire composed of five parts. First, they reported demographic variables: sex, age, and study program. Most subjects were enrolled either in a Bachelor of Science in Business and Economics (67 subjects), Political Science Bachelor (24 subjects), or various Master programs at LUSEM (25 subjects). 16 subjects were not enrolled in any program,

[11] `https://secure.savethechildren.org`

and six were students at the Faculty of Engineering at Lund University (LTH).

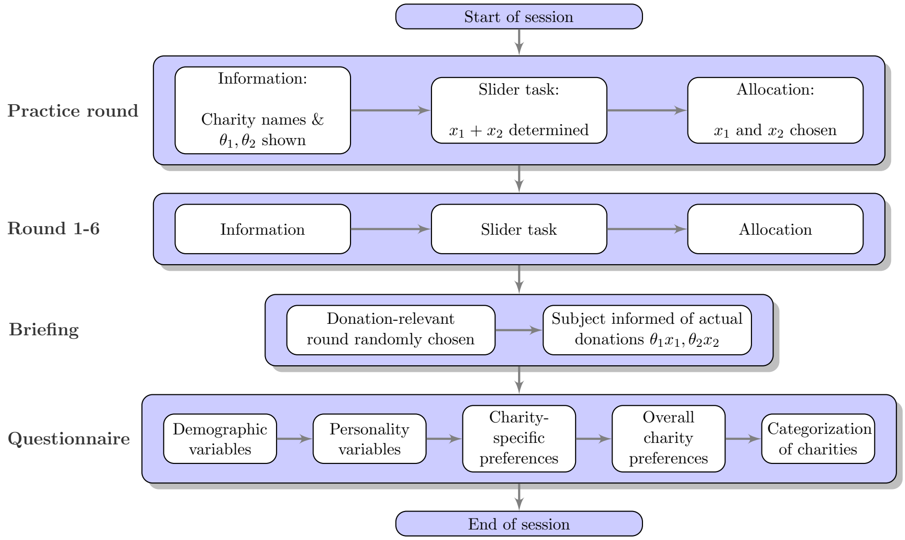


Figure 3.1: *Experimental timeline.*

Second, we elicited various personality variables: a 'satisficing' measure (Turner et al., 2012), the Conscientiousness dimension of Big-Five personality theory (John and Srivastava, 1999); the three 'personal values' dimensions of Benevolence, Universalism, and Achievement (Schwartz, 2003); and cognitive-reflection skills (by the test presented in Frederick, 2005). Except for the cognitive-reflection test, these sections of the questionnaire were in English; the rest of the survey was in Swedish (translated versions of all questions specific to our study are available in Appendix B).[12]

Third, each subject graded (on 5-point scales) his or her knowledge of each charity, along with its perceived favorability and identity/self-

[12] As a robustness check, at a time between 10 days and a month from the experimental session, subjects were invited to re-take the satisficing, conscientiousness and personal values tests through an online tool. We will not use this data set, however, as participation rates were low and we were able to match online results to experimental data for only 78 of the 160 subjects.

image value. We also asked subjects to grade the overall similarity of each charity pair.

Fourth, we asked for overall charity preferences, including the amount donated to any cause per month, the amount donated to each of the charities included in the experiment, and the frequency of giving to people begging on the street.[13]

Fifth and finally, participants were explicitly instructed to sort all five charities that were included in any treatment into any number of categories; they were told to focus on the categories they found "most relevant". We then constructed a dummy variable indicating whether or not each subject thought both charities in his/her particular session belonged to the same category.

Summary statistics on how each charity was perceived on average by subjects are given in Table 3.2. The table presents, from left to right, the number of subjects who reported donating to each charity, average ratings of the favorability and self-image/identity value of each charity as well as subjects' knowledge of it, the average perceived degree of similarity between each charity and UNICEF, and finally the number of subjects who placed each charity in the same arbitrary category as UNICEF. Since subjects rated not only the charities included in their own treatment, we have sorted the results by 'all subjects' as well as only 'within relevant treatment'. The latter includes only answers by those subjects who had the opportunity to give to the charity in question, so e.g. the STC Global row of that part of the table includes only participants in the GLOBAL-CHILD treatment. Since UNICEF was available in all treatments, it is redundant here and not included.

We see that results are similar in both parts of the table. Importantly,

[13]Subjects also reported the perceived effectiveness of each charity, as well as whether the charity was thought to be as effective as UNICEF. These variables were, however, highly correlated with the favorability and overall similarity variables, respectively, and will not be used in our analysis. We also do not use an item on the share of subjects' monthly donations which are done by automatic transfer. The wording of all these questions is reported in Appendix B.

variation in the similarity item, as well as results from explicit categorization by subjects, confirm that our treatments were successful in manipulating the perceived similarity of the charity pairs.

Table 3.2: *Perceptions of charities*

All subjects ($N = 160$)

Charity	Donors (no. subj.)	Favorability (avg.)	Identity (avg.)	Knowledge (avg.)	Similar to UNICEF (avg.)	Same category as UNICEF (no. subj.)
UNICEF	5	4.43 (0.75)	2.71 (1.31)	3.66 (0.96)	-	-
STC (Global)	2	4.48 (0.70)	2.70 (1.33)	3.51 (0.90)	3.80 (1.20)	110
SSNC (Local)	0	3.53 (0.84)	2.06 (1.21)	1.94 (1.02)	1.84 (0.94)	4
STC (Local)	1	3.88 (0.87)	2.41 (1.38)	2.64 (1.25)	2.43 (1.11)	26
WWF (Global)	6	4.34 (0.78)	2.58 (1.41)	3.41 (1.19)	2.86 (1.30)	34

Within relevant treatment ($N = 40$)

Charity	Donors (no. subj.)	Favorability (avg.)	Identity (avg.)	Knowledge (avg.)	Similar to UNICEF (avg.)	Same category as UNICEF (no. subj.)
STC (Global)	0	4.53 (0.64)	2.65 (1.39)	3.58 (0.87)	4.1 (1.10)	28
SSNC (Local)	0	3.55 (0.93)	2.23 (1.10)	1.98 (0.97)	1.98 (1.17)	1
STC (Local)	0	3.83 (0.93)	2.40 (1.43)	2.78 (1.27)	2.50 (1.30)	7
WWF (Global)	3	4.40 (0.63)	2.53 (1.30)	3.50 (1.09)	3.23 (1.40)	6

Standard deviations in parentheses. 'Within relevant treatment' includes subjects in treatments where donating to the relevant charity was possible. For the questions underlying this data set, see Appendix B.1 (number of donors, favorability, identity, knowledge, similarity with UNICEF) and Appendix B.3 (categorization).

3.4 Results

Our first priority in analyzing the experimental dataset is to check whether results are consistent with predictions. Before turning to formal testing, however, we inspect the data visually. Figure 3.2 provides an initial look at the spillover patterns within each treatment. We see, first, that all lines slope downwards, indicating that there was crowding-out across all charity pairs, in support of Hypothesis 1. Indeed, since θ_{UNICEF} was fixed at 10 SEK/point, subjects not only allocated less points to UNICEF as θ_{Other} increased, but donated less money to that charity as well. Also, the spillover effect was much more pronounced in some treatments than in others, in a manner conforming with Hypothesis 3. Our similar treatment, GLOBAL-CHILD, saw the most crowding-out across charities (steepest line); the least similar, LOCAL-GREEN, had the least crowding-out (flattest line); and our intermediate treatments exhibited moderate crowding-out. Because the LOCAL-CHILD and GLOBAL-GREEN lines are very close to each other, the local/global and cause dimensions seem equally important in driving spillovers.

Figure 3.3 reinterprets the data in relative terms, presenting the share of total points allocated to UNICEF, i.e. $x_{UNICEF}/(x_{UNICEF} + x_{Other})$. In terms of relative slope and height, there are hardly any differences between Figures 3.2 and 3.3. We now see clearly that spillovers are asymmetric. When UNICEF is the more productive option in terms of SEK/point, it receives roughly 65-85% of all allocated points, but even when it is less productive, subjects often contribute a large share of points to this charity. Indeed, major differences between treatments only arise when UNICEF is the less productive charity (to the right of the $\theta_{Other} = 10$ mark). This suggests that donating to UNICEF was seen as something approximating a default option by many subjects. While we can only speculate as to why, it seems plausible that it is partly because UNICEF is generally both well known and well liked (Table 3.2). It may also be simply because the productivity of UNICEF was fixed at 10 SEK/point, while θ_{Other} varied.[14]

Table 3.3 categorizes subjects by behavioral pattern and treatment,

[14] This asymmetry is not explained by the theory of Appendix A. As noted there, the cutoff $\bar{\rho}$ below which the similarity prediction holds is smaller at $\theta_{Other} = 16$ than at $\theta_{Other} = 4$, suggesting that the treatment effect should be more, rather than less, apparent when $\theta_{Other} < \theta_{UNICEF}$.

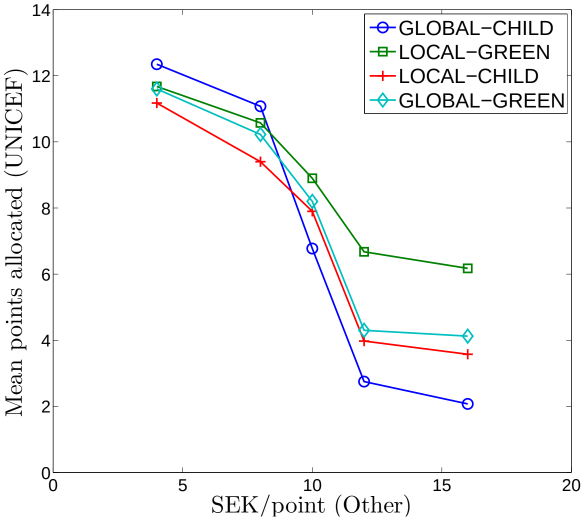


Figure 3.2: *Graphical illustration of spillovers and treatment effects. Allocation in absolute terms (number of points).*

shedding additional light on these patterns. The strong negative spillovers observed in the most similar treatment is seen to have been driven in large part by subjects perfectly substituting across donations received by charities.[15] Also, in line with the asymmetry noted above, fully half of subjects in the dissimilar treatment gave the majority of points to one charity, regardless of relative productivity ('preference for one charity'), and of these 20 subjects, 17 consistently gave most of their points to UNICEF, again suggesting there was a strong preference for this charity. We examine the relative impact on spillovers of similarity and e.g. favorability in more detail in Section 4.1; for now, simply note that these patterns are not inconsistent with an explanation of the treatment effect in terms of similarity, as the low degree of similarity between UNICEF and SSNC may be precisely what makes subjects continue to give most points to UNICEF even when SSNC is more productive.

[15] Overall, although a clear minority of subjects exhibited perfect substitution, such behavior was more common than in Null (2011), where only about 7% of participants perfectly substituted.

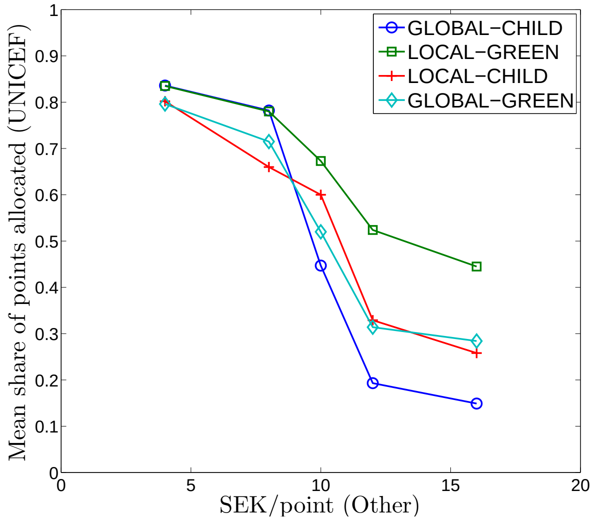


Figure 3.3: *Graphical illustration of spillovers and treatment effects. Allocation in relative terms (share of points).*

Figure 3.3 also implies that points allocated to Other strongly increased with θ_{Other}. Although a positive relationship was predicted in the theory of Appendix A, the fact that it is so pronounced is somewhat notable in light of the elasticities calculated from the field experiment of Huck and Rasul (2011). They derive strictly positive cross-price elasticities of consumption (i.e. of not giving) with respect to the price of giving (donation-matching rates) to the Bavarian State Opera House. Own-price elasticities for donations *received* by the Opera House with respect to the price of giving are negative, however. Translated to our setting, these estimates correspond to $\theta_{Other}x_{Other}$ increasing, but x_{Other} falling as θ_{Other} rises.[16] Clearly, the latter is not true in our data. The former, however, certainly is, and this is consistent with a large number of

[16]The authors argue that since matching schemes presuppose the existence of a lead donor, previous elasticity estimates (Karlan and List, 2007) have confounded the pure price effect with the impact of announcing the lead donor. Since our experiment does not describe the inverse price of giving in terms of matching, the Huck and Rasul (2011) estimates are more likely to be relevant for our study.

Table 3.3: *Substitution patterns in response to shifts in charity-specific productivity*

	GLOBAL-CHILD	LOCAL-GREEN	LOCAL-CHILD	GLOBAL-GREEN	Sum
Always perfectly substituted	19	7	7	11	44
Always weakly substituted	3	1	2	1	7
Mixed pattern (perfect)	2	0	5	1	8
Mixed pattern (weak)	4	1	2	4	11
Mixed pattern (even perfect/weak split)	1	1	1	2	5
Preference for one charity	1	20	8	3	32
Weak preference for one charity	1	2	5	4	12
Allocated inefficiently	2	0	2	4	8
Other pattern	7	8	8	10	33
Sum	40	40	40	40	160

Note: a subject who gave all points to the more productive option in every round where $\theta_{UNICEF} \neq \theta_{Other}$ is taken as having always perfectly substituted. Weak substitution occurred when, in the same situation, the majority of points was given to the more productive option. Some subjects mixed these two patterns; for instance, a subject who, in the four rounds where $\theta_{UNICEF} \neq \theta_{Other}$, gave all points to the most productive charity in one round, and most points to the more productive charity in the other three rounds, is categorized as 'Mixed pattern (weak)'. A subject who in all rounds gave most points to the same charity is considered to have exhibited a preference for that charity; if this was done in all rounds except the one where the non-preferred option had the highest possible relative productivity, the subject was considered to have a 'weak' preference. Finally, a subject who consistently gave most points to the least productive option is categorized as having 'allocated inefficiently'.

studies showing that donations received by a particular recipient depend positively on the productivity or inverse price of giving to that recipient (Andreoni and Miller, 2002; Peloza and Steel, 2005; Karlan and List, 2007).

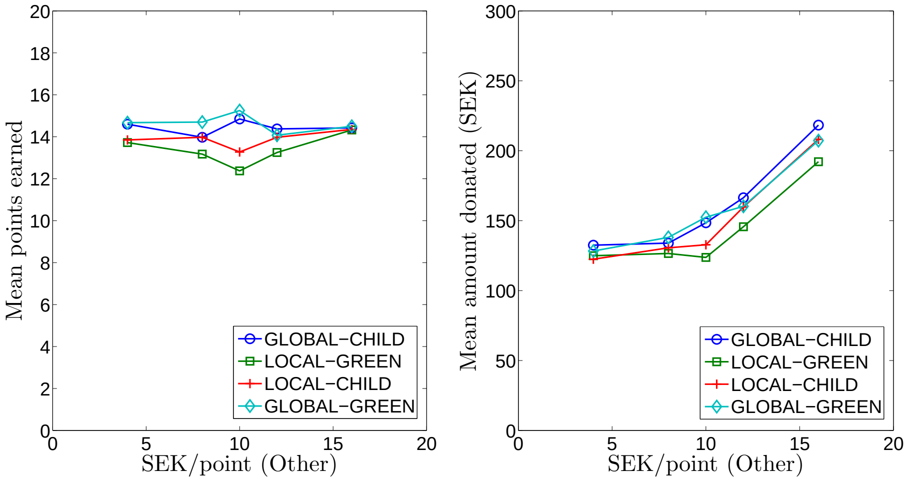


Figure 3.4: *Effect of activity-specific productivity on overall giving. Left: total efforts. Right: total amounts donated.*

The reason why Figures 3.2 and 3.3 are so similar is that, as shown in the left panel of Figure 3.4, the mean number of total points earned by subjects was roughly constant for different values of θ_{Other} and for different treatments. This pattern is consistent with our theoretical model, as well as evidence reported by Imas (2014) that people's prosocial (real) efforts are unresponsive to the magnitude of the resulting benefits. As a result, total amounts donated ($\theta_{UNICEF}x_{UNICEF}+\theta_{Other}x_{Other}$) clearly increased with θ_{Other}, as shown in the right panel of Figure 3.4. Indeed, in confirmation of Hypothesis 2, donations seem to have increased especially strongly when $\theta_{Other} > \theta_{UNICEF} = 10$.

The above figures, while suggestive, only provide information averaged at the treatment level. To analyze the data using within-treatment variation, we perform various formal tests. First, we confirm the asymmetric pattern mentioned above by a series of non-parametric Wilcoxon rank-sum tests. For each value of θ_{Other}, we run three (two-sided) tests comparing the mean number of points allocated to UNICEF in GLOBAL-CHILD versus all other treatments. Table 3.4 gives differences in means and significance levels for all tests, confirming that differences between treatments with respect to contributions to UNICEF tend to emerge only when $\theta_{Other} \geq \theta_{UNICEF}$.[17]

Table 3.4: *Results of non-parametric Wilcoxon rank-sum tests: points allocated to UNICEF*

	LOCAL-GREEN	LOCAL-CHILD	GLOBAL-GREEN
$\theta_{Other} = 4$	-0.675	-1.175	-0.75
$\theta_{Other} = 8$	-0.5	-1.675	-0.85
$\theta_{Other} = 10$	2.125*	1.125	1.425
$\theta_{Other} = 12$	3.925***	1.225	1.55
$\theta_{Other} = 16$	4.1***	1.5	2.05**

Note: table presents differences in means for x_{UNICEF} between GLOBAL-CHILD and each other treatment, conditional on θ_{Other}. Also given are (two-tailed) significance levels (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$) when testing for equality of means in x_{UNICEF}.

Table 3.4, in analyzing contribution levels, provides an indirect formal test for treatment effects. However, a more direct test looks for differences in the *slope* of contributions, as in Figure 3.2. This corresponds to running regressions of the form[18]

$$x_{UNICEF,i} = \beta_0 + \beta_1 \theta_{Other,i} + \boldsymbol{\beta_2'} \boldsymbol{X_i} + \boldsymbol{\beta_3'} \theta_{Other,i} \boldsymbol{X_i} + \boldsymbol{\beta_4} \boldsymbol{T_i} + \boldsymbol{\beta_5} \theta_{Other,i} \boldsymbol{T_i} + \epsilon_i \quad (3.2)$$

where $\boldsymbol{X_i}$ is a vector of control variables, and $\boldsymbol{T_i}$ is a vector of treatment dummies. In our regressions, we treat GLOBAL-CHILD as the baseline, so $\boldsymbol{T_i}$ includes dummy variables for LOCAL-GREEN, LOCAL-CHILD,

[17] We also tried using LOCAL-GREEN as our benchmark, comparing the mean x_{UNICEF} in this treatment with the means of each other treatment. This also confirms the asymmetry observed in Figures 3.2 and 3.3, and unlike GLOBAL-CHILD, LOCAL-GREEN is significantly different from LOCAL-CHILD for $\theta_{Other} > 10$.

[18] Robust standard errors are clustered at the subject level in all regressions reported in this paper.

and GLOBAL-GREEN. Note that treatment effects are given by interaction terms $\boldsymbol{\beta_5}\theta_{Other,i}\boldsymbol{T_i}$ in (4.1).

Results are given in Table 3.5.[19] Column 1 presents outcomes from the simplest version of equation (4.1), with no control variables; hence, it closely corresponds to Figure 3.2. Note that although regression equation (4.1) includes non-interacted terms for relevant treatment and (if applicable) control variables, the table reports results only on θ_{Other} itself and all interaction terms. We see that in the similar GLOBAL-CHILD treatment (the non-interacted θ_{Other} coefficient), there is clear evidence of crowding-out, confirming Hypothesis 1. Moreover, consistent with Hypothesis 3, this effect is less pronounced in the other treatments, though significantly so only for LOCAL-GREEN, the least similar charity pair, where the effect's magnitude is roughly halved. Column 2 shows that the significance of these interaction coeffients is robust to adding demographic controls (sex, age, study program) as well as each observation's round number.[20]

[19]We ran a number of variants of the OLS regressions summarized in Table 3.5 and beyond. First, we tried using LOCAL-GREEN rather than GLOBAL-CHILD as the baseline (Tables 3.5 and 3.6 only). Second, we included subject fixed effects. Third, we replaced x_{Other} as dependent variable with the share-of-points variable $x_{UNICEF}/(x_{UNICEF} + x_{Other})$. For all of these extensions, results were qualitatively (and indeed, often quantitatively) very similar to the OLS results presented in this paper.

[20]In an exploratory analysis, we tested for individual heterogeneity by running OLS and Tobit regressions of x_{UNICEF} on θ_{Other} and interacted questionnaire items. Few patterns emerge. Master students do substitute more, consistent with e.g. an ‘indoctrination effect’ (Frank et al., 1993; Bauman and Rose, 2011) due to advanced students being more familiar with economic ideas such as the law of demand. Note, however, that subjects did not specify which Master program they were enrolled in, so these results may not be driven by economics majors. Two psychological measures are significant: self-reported ‘satisficers’ and subjects with good cognitive-reflection (CR) skills both exhibit greater crowding-out. While these two types may seem like opposites, high-CR subjects were also more likely to choose boundary solutions in the allocation decision. In one interpretation, while giving *most* points to the charity with the highest exchange rate seems an obvious heuristic, the stronger rule of allocating *all* points to this charity (and hence, none to the other) is more restrictive, may appear aversive at first glance, and is therefore more likely to be followed by high-CR subjects.

Table 3.5: *Treatment effects*

VARIABLES	Dependent variable in all regressions: x_{UNICEF}			
	OLS (1)	OLS (2)	Tobit (3)	Tobit (4)
θ_{Other}	-0.979***	-0.518	-1.329***	-0.724
	(0.144)	(0.496)	(0.213)	(0.696)
$\theta_{Other}\times$GLOBAL-CHILD		Reference category		
$\theta_{Other}\times$LOCAL-GREEN	0.469***	0.457***	0.728***	0.715***
	(0.177)	(0.173)	(0.244)	(0.233)
$\theta_{Other}\times$LOCAL-CHILD	0.273	0.243	0.416*	0.380
	(0.177)	(0.169)	(0.249)	(0.236)
$\theta_{Other}\times$GLOBAL-GREEN	0.270	0.252	0.433*	0.406*
	(0.183)	(0.176)	(0.251)	(0.239)
$\theta_{Other}\times$Round		-0.046*		-0.062*
		(0.027)		(0.035)
Demographic controls	NO	YES	NO	YES
Observations	800	800	800	800
R-squared	0.244	0.278	0.050 (pseudo)	0.056 (pseudo)

Robust standard errors clustered at the subject level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Only coefficients on interaction terms reported. All interacted variables were included separately in each regression. Demographic controls included: sex, age, and study program.

Out of all 800 observations, 185 have the boundary solution $x_2 = 0$, and we may suspect that some subjects, if able, would have been willing to set x_2 to a negative number in order to further increase donations to their preferred option. It may therefore be of interest to compare OLS to a model which assumes that the distribution of x_2 values is censored at zero. Hence, columns 3 and 4 of Table 3.5 repeat the analysis using a Tobit regression model. Although unsurprisingly the magnitude of the treatment effect is somewhat greater in the Tobit regressions (causing all treatment differences to be at least marginally significant), on the whole differences between the models are minor.

Next, we pool data across all treatments and regress total efforts only on θ_{Other}. Results confirm that there is no significant relationship between charity-specific productivity and total efforts.[21] There is also little indication that the number of points earned reacted to changes in the maximum of θ_{UNICEF} and θ_{Other} ($p = 0.251$).[22] Furthermore, in a substantial departure from contribution patterns typically observed in endowment-based dictator games, the number of points earned was roughly normally distributed, and the number of points earned was below 5 in only 17 observations (out of 960), two of which had $x_{UNICEF} + x_{Other} = 0$. It thus appears that most subjects found the slider-adjustment task relatively easy to perform.

Finally, we confirm Hypothesis 2, namely that total amounts donated ($\theta_{UNICEF} x_{UNICEF} + \theta_{Other} x_{Other}$) increases strongly with θ_{Other} when $\theta_{Other} \geq 10$ ($\beta = 11.316$, $p = 0.000$), as seen in the right panel of Figure 3.4. Indeed, regressing total donations on θ_{Other} using only observations where $\theta_{Other} < 10$ still yields a marginally significantly positive slope coefficient ($p = 0.089$). On the whole, therefore, our data indicates that total donations received are increasing in the productivity of both the less productive and the more productive charity. This is turn implies

[21] Overall effort was not constant over time, as regressing total points on the 'round' variable produces a highly significant (p=0.000) coefficient of 0.674.

[22] If the general-productivity shock round (where $\theta_{UNICEF} = \theta_{Other} = 16$) is included in the sample, the coefficient is positive and marginally significant ($p = 0.106$).

that spillovers can be seen as partial, attenuating but not reversing the effect of a charity-specific productivity shift on total donations received.

3.4.1 Robustness tests

All results presented so far are consistent with predictions. We will now argue that treatment effects are in fact caused by the differing similarity of charity pairs and not by some confounding factor, such as differences in how well known and well regarded each 'Other' charity is compared to UNICEF. Returning to Table 3.2, although patterns are not clear-cut, we see e.g. that SSNC Lund was both the least well known and the least popular charity, so subjects may have given less to SSNC than to the Save the Children Global Action Fund because of the latter's relative popularity rather than because of their varying similarity with UNICEF.

To settle the issue, we examine individual variation within each treatment. Figure 3.5 provides a look at the raw correlation between substitution patterns and perceived similarity, as well as some possible confounders. All panels are based on a simple pooled interaction regression of the form

$$x_{UNICEF,i} = \alpha + \beta\theta_{Other,i} + \gamma z_i + \delta\theta_{Other,i} z_i + \epsilon_i$$

for some interacted variable z_i. In the upper-left panel, z_i is the perceived similarity between UNICEF and Other, and the figure shows the marginal effect of θ_{Other} on x_{UNICEF} (i.e. spillovers) for various values of this variable. Hence, the slope of the line represents δ, the difference that z_i makes for the magnitude of spillovers. Clearly, greater perceived similarity is correlated with stronger crowding-out. This is confirmed by the underlying regression, where the estimated $\delta < 0$ has $p = 0.000$.

The other three panels represent similar correlations for the *difference* in perceived favorability, identity/self-image value, and knowledge between UNICEF and Other (calculated as $z_i = z_{UNICEF,i} - z_{Other,i}$). Although all suggest a positive slope — for instance, the more favorable UNICEF

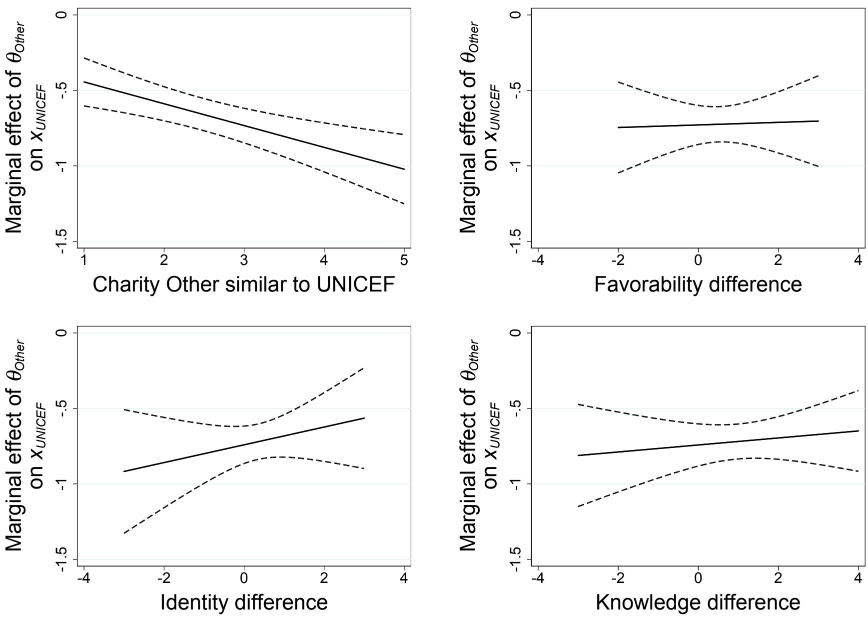


Figure 3.5: *The impact of charity-specific preference and belief variables on spillovers. Dashed lines give 95% confidence intervals.*

is perceived compared to Other, the less crowding-out — none are significantly increasing. In addition, of the four lines, the one associated with the similarity variable is the steepest: a one-point change in similarity has a greater effect on spillovers than a one-point change in any of the three difference variables. This suggests that of the four variables, similarity is the most important for explaining spillover patterns.

A more careful test involves adding these four variables as controls in regression (4.1). These results are presented in Table 3.6. Column 1 adds differenced variables for perceived favorability, identity/self-image value, and knowledge; clearly, this has only a minor impact on treatment effects, suggesting they are not driven by such differences among charities. Indeed, if anything, including these variables makes the treatment effect more prominent compared to Table 3.5. By contrast, adding also the similarity variable (column 2) does in fact render the treatment effect insignificant, while the similarity variable itself is highly significant and has the expected sign.[23] As in Table 3.5, when we repeat the analysis using the Tobit model (columns 3 and 4), results are broadly similar. This confirms that variation in substitution patterns across treatments does seem to reflect the perceived similarity of the charities, rather than e.g. their relative popularity among subjects.

Finally, we address a different concern, namely that the sequence of θ_{Other} values, while randomly generated, may make a difference for subject behavior. That is, subjects may 'anchor' their beliefs about reasonable contribution levels based on first-period exchange rates. To examine this possibility, we generate dummy variables for each possible value of first-period θ_{Other}. As shown in Table 3.7, there were no differences in either the level of effort (column 1) or the magnitude of spillovers (column 2).[24] A Tobit regression corresponding to column 2 (not reported) yields qualitatively identical results.

[23] In a variant of these regressions, we included the favorability, identity/self-image, and knowledge variables for Other instead of first taking differences. The pattern with respect to treatment effects and the similarity variable was very similar, with one exception, namely that identity/self-image was significant at the 5 % level both

Table 3.6: *Robustness checks*

VARIABLES	Dependent variable in all regressions: x_{UNICEF}			
	OLS (1)	OLS (2)	Tobit (3)	Tobit (4)
θ_{Other}	-0.388	0.261	-0.535	0.291
	(0.511)	(0.586)	(0.710)	(0.797)
θ_{Other}×GLOBAL-CHILD		Reference category		
θ_{Other}×LOCAL-GREEN	0.620***	0.324	0.882***	0.502*
	(0.195)	(0.207)	(0.255)	(0.265)
θ_{Other}×LOCAL-CHILD	0.331*	0.097	0.471**	0.174
	(0.170)	(0.184)	(0.235)	(0.252)
θ_{Other}×GLOBAL-GREEN	0.266	0.159	0.396*	0.267
	(0.177)	(0.179)	(0.237)	(0.237)
θ_{Other}×Round	-0.050*	-0.045*	-0.068*	-0.062*
	(0.027)	(0.027)	(0.035)	(0.034)
θ_{Other}×Favorability diff.	-0.086	-0.083	-0.093	-0.094
	(0.066)	(0.062)	(0.084)	(0.079)
θ_{Other}×Identity diff.	0.080	0.023	0.145	0.066
	(0.075)	(0.081)	(0.101)	(0.108)
θ_{Other}×Knowledge diff.	-0.069	-0.049	-0.099	-0.071
	(0.055)	(0.053)	(0.069)	(0.066)
θ_{Other}×UNICEF/Other similarity		-0.139***		-0.182***
		(0.050)		(0.066)
Demographic controls	YES	YES	YES	YES
Observations	800	800	800	800
R-squared	0.343	0.354	0.072 (pseudo)	0.075 (pseudo)

Robust standard errors clustered at subject level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Only coefficients on interaction terms reported. All interacted variables were included separately in each regression. Demographic controls included: sex, age, and study program.

Table 3.7: *Anchoring effects*

VARIABLES	Dep. variable: $x_{UNICEF}+x_{Other}$	Dep. variable: x_{UNICEF}
	(1)	(2)
θ_{Other}		-0.668***
		(0.173)
$I(\theta_{Other} = 4 \text{ in round } 1)$	Reference category	
$I(\theta_{Other} = 8 \text{ in round } 1)$	-0.430	0.297
	(0.996)	(2.408)
$I(\theta_{Other} = 10 \text{ in round } 1)$	1.413	3.680
	(1.519)	(2.787)
$I(\theta_{Other} = 12 \text{ in round } 1)$	0.874	0.042
	(1.264)	(2.846)
$I(\theta_{Other} = 16 \text{ in round } 1)$	-0.320	0.068
	(0.991)	(2.403)
$\theta_{Other} \times I(\theta_1 = 4 \text{ in round } 1)$	Reference category	
$\theta_{Other} \times I(\theta_1 = 8 \text{ in round } 1)$		-0.060
		(0.207)
$\theta_{Other} \times I(\theta_1 = 10 \text{ in round } 1)$		-0.310
		(0.229)
$\theta_{Other} \times I(\theta_1 = 12 \text{ in round } 1)$		0.069
		(0.229)
$\theta_{Other} \times I(\theta_1 = 16 \text{ in round } 1)$		0.100
		(0.201)
Constant	13.64***	13.85***
	(0.858)	(2.063)
Observations	695	695
R-squared	0.018	0.219

Robust standard errors clustered at subject level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. $I(\cdot)$ is an indicator function which is equal to one if the condition within parentheses holds, and equal to zero otherwise.

3.5 Conclusion

Does an intervention targeting a particular prosocial activity spill over onto other activities? This paper has argued, on the basis of experimental data, that it may. In particular, we have presented evidence of a partial crowding-out effect within the charitable-giving domain: spillovers attenuate, though do not completely reverse, the intervention's impact. This raises important questions about the efficiency properties of interventions such as donation-matching. According to our results, such schemes can be expected to only partly raise additional donations, while distorting the initial pattern of donations.

Furthermore, we have presented data indicating that spillovers decrease in magnitude as activities grow more distant in terms of similarity. Hence, for example, a policy affecting an environmental activity may spill over onto other environmental acts but, perhaps, not onto e.g. charitable giving to the poor. Whether this is actually the case is, of course, an empirical question which merits further investigation. In this context, it is noteworthy that the similarity item included in our questionnaire was highly predictive of negative spillovers both within and across treatments. There may be scope for using questions of this type in ex-ante policy evaluation. Ideally, of course, spillover effects should be estimated by well-identified empirical analysis rather than by survey. Nevertheless, once more empirical evidence on spillovers has accumulated, it is worth checking whether similarity ratings are highly correlated with results; if they are, survey items may eventually prove useful as a cheap shortcut for policy analysts.

---

when similarity was included and when it was not.

[24] Note that in Table 3.7, we have dropped entirely those subjects who had $\theta_{UNICEF} = \theta_{Other} = 16$ in the first round; including them does not make any difference. Neither does interacting the maximum of exchange rates in the first round, or the minimum, regardless of whether such subjects are included or not (there is a single exception: when these subjects are not included, those who faced a maximum exchange rate of 16 in the first round exhibit less crowding-out, with $p = 0.071$, than those who faced a maximum exchange rate of 10; given the sheer number of regressions involved, we find this less than persuasive).

A final note on the external validity of our results. We have in this paper focused on the direct effect of activity-specific productivity. This is not to argue that relative productivity is the only factor underlying choices on how to contribute; our assumption of no social interaction, in particular, has been made for simplicity rather than realism. In addition, interventions may impact behavior indirectly through means other than productivity, including by signaling to agents in the economy that the public good targeted is more important than previously thought, or that a particular charity is of high quality. The main contribution of this paper has been to provide evidence on similarity as one of the relevant factors at work in determining the magnitude of productivity-led spillovers.

Acknowledgments

Financial support from the Jan Wallander and Tom Hedelius Foundation is gratefully acknowledged. Erik Wengström and Ola Andersson provided valuable comments which greatly improved this paper. Helpful comments were also provided by Tommy Andersson, Jens Gudmundsson, Håkan J. Holm, Karl McShane, Amrish Patel, Marco Piovesan, Eskil Rydhe, Margaret Samahita, and Anna Welander, as well as participants at seminars and conferences in Lund, Copenhagen, Aarhus, Heidelberg and Tampere.

References

J. Andreoni. Impure Altruism and Donations to Public Goods: A Theory of Warm-glow Giving. *The Economic Journal*, 100(401), 1990.

J. Andreoni and J. Miller. Giving according to garp: An experimental test of the consistency of preferences for altruism. *Econometrica*, 70 (2), 2002.

Y. Bauman and E. Rose. Selection or indoctrination: Why do economics students donate less than the rest? *Journal of Economic Behavior & Organization*, 79(3), 2011.

R. Bénabou and J. Tirole. Incentives and Prosocial Behavior. *American Economic Review*, 95(5), 2006.

C. Blackwell and M. McKee. Only for my own neighborhood? Preferences and voluntary provision of local and global public goods. *Journal of Economic Behavior & Organization*, 52(1), 2003.

I. Blanken, N. van de Ven, and M. Zeelenberg. A Meta-Analytic Review of Moral Licensing. *Personality and Social Psychology Bulletin*, 41(4), 2015.

K.A. Brekke, S. Kverndokk, and K. Nyborg. An economic model of moral motivation. *Journal of Public Economics*, 87(9), 2003.

S. Brown, M.N. Harris, and K. Taylor. Modelling charitable donations to an unexpected natural disaster: Evidence from the U.S. Panel Study of Income Dynamics. *Journal of Economic Behavior & Organization*, 84(1), 2012.

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.

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. A Multi-Activity Approach to Public-Good Contributions, 2015. Working Paper 2015:26, Department of Economics, Lund University.

U. Fischbacher. z-tree: Zurich Toolbox for Ready-Made Economic Experiments. *Experimental Economics*, 10(2), 2007.

R.H. Frank, T. Gilovich, and D.T. Regan. Does Studying Economics Inhibit Cooperation? *Journal of Economic Perspectives*, 7(2), 1993.

S. Frederick. Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*, 19(4), 2005.

D. Gill and V. Prowse. A Structural Analysis of Disappointment Aversion in a Real Effort Competition. 2012.

A. Gneezy, A. Imas, A. Brown, L.D. Nelson, and M.I. Norton. Paying to Be Nice: Consistency and Costly Prosocial Behavior. *Management Science*, 58(1), 2012.

C. Heath and J.B. Soll. Mental Budgeting and Consumer Decisions. *Journal of Consumer Research*, 23(1), 1996.

P.W. Henderson and R.A. Peterson. Mental Accounting and Categorization. *Organizational Behavior and Human Decision Processes*, 51 (1), 1992.

S. Huck and I. Rasul. Matched fundraising: Evidence from a natural field experiment. *Journal of Public Economics*, 95(5-6), 2011.

A. Imas. Working for the 'warm glow': On the benefits and limits of prosocial incentives. *Journal of Public Economics*, 114, 2014.

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.P. John and S. Srivastava. The Big Five Trait Taxonomy: History, Measurement, and Theoretical Perspectives. In L.A. Pervin and O.P. John, editors, *Handbook of personality: Theory and research*. New York: Guilford Press, 2nd edition, 1999.

D. Karlan and J.A. List. Does Price Matter in Charitable Giving? evidence from a Large-Scale Natural Field Experiment. *American Economic Review*, 97(5), 2007.

D.T. Miller and D.A. Effron. Psychological License: When it is Needed and How it Functions. 43, 2010.

C. Null. Warm glow, information, and inefficient charitable giving. *Journal of Public Economics*, 95(5-6), 2011.

J. Peloza and P. Steel. The Price Elasticities of Charitable Contributions: A Meta-Analysis. *Journal of Public Policy & Marketing*, 24(2), 2005.

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. Substitution Between (and Motivations for) Charitable Contributions: An Experimental Study. 2007. Discussion Paper No. 648, Department of Economics, University of Essex.

D. Reinstein. Does One Charitable Contribution Come at the Expense of Another? *The B.E. Journal of Economic Analysis & Policy*, 11(1), 2011.

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.

S.H. Schwartz. A proposal for measuring value orientations across nations, 2003. Chapter 7 in *Questionnaire development report of the European Social Survey.* Retrieved from `http://www.europeansocialsurvey.org/docs/methodology/core_ess_questionnaire/ESS_core_questionnaire_human_values.pdf`.

R. Thaler. Mental Accounting and Consumer Choice. *Marketing Science*, 4(3), 1985.

R.H. Thaler. Mental Accounting Matters. *Journal of Behavioral Decision Making*, 12(3), 1999.

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.

B.M. Turner, H.B. Rim, N.E. Betz, and T.E. Nygren. The Maximization Inventory. *Judgment and Decision Making*, 7(1), 2012.

A. Tversky and D. Kahneman. The Framing of Decisions and the Psychology of Choice. *Science*, 211(4481), 1981.

Appendices

Appendix A. Theoretical predictions

Consider the quasilinear utility function

$$U(x_1, x_2) = T - x_1 - x_2 + \beta \ln \left[((\theta_1 x_1)^\rho + (\theta_2 x_2)^\rho)^{1/\rho} \right]$$

where production is represented by the symmetric CES function $g = ((\theta_1 x_1)^\rho + (\theta_2 x_2)^\rho)^{1/\rho}$ nested inside the warm-glow term. The parameter $\beta > 0$ reflects how much the agent cares about warm glow and hence measures her degree of (impure) altruism. ρ reflects the degree of substitutability between contributions received through activity 1 ($\theta_1 x_1$) and activity 2 ($\theta_2 x_2$): if $\rho = 1$, activities are perfect substitutes; if $\rho \to -\infty$, activities are perfect complements; and production is Cobb-Douglas for $\rho = 0$.

We will focus on the case where activities are relatively substitutable (so $0 \leq \rho < 1$) and examine comparative statics for θ_1, results for a shift in θ_2 being completely analogous. At an interior solution, optimal efforts on activity $k = 1, 2$ are given by

$$x_k^* = \frac{\beta \theta_l^{1-c}}{\theta_1^{1-c} + \theta_2^{1-c}} \tag{3.3}$$

for $k \neq l$ and $c \equiv 1/(1-\rho)$;[25] note that $0 \leq \rho < 1$ corresponds to $c \geq 1$. It is then straightforward to show that

$$\frac{dx_2^*}{d\theta_1} = \frac{\beta(1-c)\theta_2^{1-c}}{\theta_1^c \left(\theta_1^{1-c} + \theta_2^{1-c}\right)^2} \leq 0 \tag{3.4}$$

[25] The solution obtained when the 'outer' utility function is of the Cobb-Douglas type, i.e.

$$U(x_1, x_2) = \alpha \ln (T - x_1 - x_2) + (1 - \alpha) \ln \left[((\theta_1 x_1)^\rho + (\theta_2 x_2)^\rho)^{1/\rho} \right]$$

is identical to (3.3) if $\beta = (1 - \alpha)T$.

for $c \geq 1$. Hence, x_2 is decreasing in θ_1 (negative spillovers):

Hypothesis 1. When a particular prosocial activity is exposed to a directed positive productivity shock, agents decrease efforts on the other activity.

It also follows immediately from (3.3) that total efforts $x_1^* + x_2^* = \beta$ and hence $dx_1^*/d\theta_1 = -dx_2^*/d\theta_2 \geq 0$.[26] To evaluate the effect of θ_1 on total contributions received $(\theta_1 x_1^* + \theta_2 x_2^*)$, it is illuminating to first rewrite this quantity as

$$\begin{aligned}\theta_1 x_1^* + \theta_2 x_2^* + \theta_2 x_1^* - \theta_2 x_1^* &= \theta_2(x_1^* + x_2^*) + (\theta_1 - \theta_2)x_1^* \\ &= \beta\theta_2 + (\theta_1 - \theta_2)x_1^*\end{aligned}$$

Now differentiate with respect to θ_1 to yield

$$x_1^* + (\theta_1 - \theta_2)\frac{dx_1^*}{d\theta_1}$$

As $dx_1^*/d\theta_1 \geq 0$, this expression is positive when $\theta_1 > \theta_2$, but indeterminate otherwise. This is because even if the productivity of the *less* productive activity increases, agents will substitute towards it. In that case, the net change in total contributions received will be given by the preexisting stock of x_1, which is made more productive, and the magnitude of the substitution towards x_1.[27] In summary, we have the

[26]While it is not true for more general g that total efforts $x_1^* + x_2^*$ are constant, the effect of activity-specific productivity on total efforts is typically ambiguous (for instance, this is true in the perfect-substitutes case where $\rho = 1$). Intuitively, the ambiguity arises because when productivity increases it becomes possible to do more with less effort, and this may be desirable since I measures self-ideal discrepancy only in terms of production.

[27]Note, however, that agents' (perceived) public-good production g is still increasing in θ_1. The change in production is given by

$$\frac{dg(x_1^*, x_2^*)}{d\theta_1} = \frac{\partial g}{\partial \theta_1} + \frac{\partial g}{\partial x_1} \cdot \frac{dx_1^*}{d\theta_1} + \frac{\partial g}{\partial x_2} \cdot \frac{dx_2^*}{d\theta_1}$$

but at an interior optimum, we have $\partial g/\partial x_1 = \partial g/\partial x_2$; and we have seen also that following:[28]

Hypothesis 2. Total contributions received $(\theta_1 x_1 + \theta_2 x_2)$ is increasing in the productivity of the more productive activity.

Finally, we examine the effect of similarity on the magnitude of $dx_2^*/d\theta_1$. We assume that activities that are more similar are perceived by agents as having a larger value of ρ (i.e. closer to one). As c is increasing in ρ, the hypothesis is that $dx_2^*/d\theta_1$ is decreasing in c. Differentiating (3.4) with respect to c, we find

$$\frac{d}{dc}\left(\frac{dx_2^*}{d\theta_1}\right) = -\frac{\beta\theta_2^{1-c}}{\theta_1^c\left(\theta_1^{1-c}+\theta_2^{1-c}\right)^3}\Big[\theta_1^{1-c}+\theta_2^{1-c} + (1-c)\left(\theta_2^{1-c}-\theta_1^{1-c}\right)\left(\ln\theta_1 - \ln\theta_2\right)\Big]$$

Clearly, the sign of this derivative will be determined by the expression within brackets, the sign of which is itself ambiguous. At first glance, the bracketed factor as a whole seems to be decreasing in the spread of the productivity parameters; this intuition can be confirmed, and further analysis greatly simplified, by redefining θ_1 and θ_2 in terms of their relative size $a = \theta_2/\theta_1$ and their geometric mean $\hat{\theta} = \sqrt{\theta_1\theta_2} = \sqrt{a}\theta_1$. We then have $\theta_1 = \hat{\theta}/\sqrt{a}$ and $\theta_2 = \sqrt{a}\hat{\theta}$, and the bracketed expression can be rewritten as

$$\frac{\hat{\theta}^{1-c}}{a^{(1-c)/2}}\left(1 + a^{1-c} - \left(a^{1-c}-1\right)(1-c)\ln a\right)$$

$dx_1^*/d\theta_1 = -dx_2^*/d\theta_2$. Hence, $dg(x_1^*, x_2^*)/d\theta_1$ is given simply by $\partial g/\partial\theta_1$, evaluated at the optimum. But

$$\frac{\partial g}{\partial\theta_1} = \theta_1^{\rho-1}x_1^{\rho}\left((\theta_1 x_1)^{\rho} + (\theta_2 x_2)^{\rho}\right)^{\frac{1-\rho}{\rho}}$$

which is nonnegative for any x_1 and x_2.

[28]It can be shown that hypotheses 1 and 2 also hold when allowing for underlying efficiency differences, i.e. for $g = \left(\alpha_1(\theta_1 x_1)^{\rho} + \alpha_2(\theta_2 x_2)^{\rho}\right)^{1/\rho}$, with $\alpha_1, \alpha_2 > 0$.

We conclude that the sign of the similarity effect will depend on whether or not

$$1 + a^{1-c} - \left(a^{1-c} - 1\right)(1-c)\ln a \geq 0 \tag{3.5}$$

Note that the prediction on similarity is independent of the magnitude of the productivity parameters, as expressed by their geometric mean. This is clearly desirable, as the choice of unit for public-goods production is arbitrary.

Although (3.5) cannot be explicitly solved for a or c, a number of observations can be made. The first is that in the benchmark case when $a = 1$ ($\theta_1 = \theta_2$), the left-hand side of (3.5) is equal to 2, so the similarity prediction then holds for any c. Moreover, even for $a \neq 1$, it is equal to 2 when $c = 1$ ($\rho = 0$). On the other hand, it then also tends towards $-\infty$ as $c \to \infty$ ($\rho \to 1$). To see why, suppose first that $a < 1$. Then, rewriting (3.5) as

$$a^{1-c}\left(\frac{1}{a^{1-c}} + 1 - \left(1 - \frac{1}{a^{1-c}}\right)(1-c)\ln a\right)$$

we see that as the contents of the outer parenthesis approaches $-\infty$ when c becomes very large, the entire expression must as well. Now suppose $a > 1$. Then, for large c, the dominant term in (3.5) will be $-c\ln a$, which of course tends to $-\infty$ as $c \to \infty$. We conclude that for any $a \neq 1$, there must exist some cutoff $\bar{c} > 1$ such that the similarity prediction holds for any $1 \leq c < \bar{c}$.[29]

Finally, we show that $\bar{c}$ is increasing in a for $a < 1$, and decreasing in a for $a > 1$; loosely speaking, the larger the spread of the productivity parameters, the smaller is the interval for which the similarity prediction holds. This we demonstrate by simply differentiating (3.5) with respect to a, showing that for any $c > 1$ the entire left-hand side of (3.5) is

[29] This statement is silent on whether the left-hand side of (3.5) is monotonic in c. If not, there could exist ranges of c larger than $\bar{c}$ where (3.5) again holds. It is simple to show (by differentiation) that there is monotonicity when $a > 1$ but not necessarily when $a < 1$. We have been unable to produce an example of nonmonotonicity beyond $\bar{c}$, however.

increasing in a for $a < 1$, and decreasing in a for $a > 1$. The cutoff must then move in an analogous manner. The derivative in question is

$$-\frac{(1-c)^2 \ln a}{a^c}$$

Clearly, the sign of this expression is equal to that of $-\ln a$.

In our experiment, $a \in [5/8, 5/2]$. $\bar{c}$ is lower at the upper bound of this interval: there, $\bar{c} \approx 2.68$, implying that negative spillovers are larger for more similar alternatives approximately at all $0 \leq \rho \leq 0.63$. Thus, a sufficient condition for the prediction to be valid is that all charity pairs included in the experiment have $0 \leq \rho \leq 0.63$.

In summary, we have the following statement.

Hypothesis 3. (i) If initially $\theta_1 = \theta_2$, activity-specific positive productivity shocks drive less 'crowding-out', the less similar the activities are. (ii) If $\theta_1 \neq \theta_2$, there exists some $0 < \bar{\rho} < 1$ such that the same conclusion applies for any $0 \leq \rho \leq \bar{\rho}$. (iii) $\bar{\rho}$ is increasing in $a = \theta_2/\theta_1$ if $a < 1$, and decreasing in a if $a > 1$.

Appendix B. Survey questions

The following questions were posed in Swedish and have been translated.

Appendix B.1. Charity-specific preferences and beliefs

1. *Do you regularly give money to X?* (regular giving, Charity X)

To what extent do you agree with the following statements?

2. *X does valuable work* (favorability of Charity X)

3. *Money given to X has a large effect* (effect of Charity X)

4. *I would be willing to wear a symbol for X on my jacket* (identity/self-image value of Charity X)

5. *I know roughly what work is done by X* (knowledge of Charity X)[30]

6. *When it comes to donations to charity, I think that gifts to UNICEF is roughly the same thing as gifts to X* (similarity of Charity X to UNICEF, overall)

7. *A Swedish crown given to UNICEF has as large an effect as a Swedish crown given to X* (marginal product of Charity X relative to UNICEF)

Appendix B.2. Overall giving preferences

1. *Roughly how much (in SEK) do you donate to various charitable organizations each month?* (monthly giving)

2. *How large a share (in percent) of the above donations are made by automatic transfer?* (share of automatic transfers)

[30]For the Save the Children Global Action Fund, the wording was slightly modified as '*...what work is done based on X*'.

3. *How often (in number of occasions) do you give money to beggars in a month?* (monthly giving to beggars)

4. *Do you show to the world that you support a charity? This might be for example by sticking a symbol on your jacket/bag or by sometimes wearing a T-shirt with a logo.* (overall giving identity)

Appendix B.3. Categorization
Consider the following charities:

Save the Children Global Action Fund
The local office of SSNC in Lund
WWF International
The local office of Save the Children in Lund
BRIS in Lund
UNICEF

You are now to sort these organizations/alternatives into categories. Use the categories that you find most relevant.

Chapter 4

Behavioral Spillovers from Food-Waste Collection in Swedish Municipalities

with Jurate Miliute-Plepiene

4.1 Introduction

Ideally, environmental policy evaluation should include all impacts of the intervention under scrutiny. Foremost among these are immediate and future costs and benefits directly associated with the policy: does it have the desired effect on the environmental variable of interest? What are associated direct costs? Other relevant effects can be viewed as *spillover* effects. Some of these, being mediated by financial incentives, can be readily analyzed using the toolset of economic theory: for example, a subsidy on the recycling of aluminum waste is likely to have indirect effects through its impact on the price of scrap metal (Palmer et al., 1997; Kaffine, 2014). However, other spillovers may occur across (at least partly) voluntary environmental behaviors, and in response to information campaigns or other non-economic policy instruments. This type of spillover has received significant attention only recently and is rarely, if ever, considered in policy evaluations, such as in cost-benefit analyses. We contribute to filling that gap by estimating policy-driven spillovers across household environmental activities. As far as we are aware, the present study is the first to do so by applying standard econometric methods to large-scale data from a natural experiment.

The intervention that we examine is food-waste collection from Swedish households. Collection has currently been implemented in roughly two-thirds of the 290 Swedish municipalities, with introduction occurring gradually from 1990 onwards. We examine whether these staggered policy changes have affected households' efforts to sort packaging waste. Our data set merges policy-introduction dates with a yearly panel on the packaging amounts collected in each municipality across 2006-2014, and allows estimation of the causal spillover effect from food-waste collection within that time interval.

The effect we examine is an example of a 'behavioral spillover'. Broadly defined, this is any path dependency in behavior — where an agent is more/less likely to perform action y having just performed action x — that is not mediated through monetary incentives, but rather through changes in self-image or in the salience of particular motives for action (Dolan and Galizzi, 2015), or through the formation of general (as opposed to specific) habits. In principle, such spillovers may have either sign. The prime example of a self-image-mediated spillover effect is 'moral licensing' (Merritt et al., 2010). This is a behavioral pattern observed in lab experiments (Blanken et al., 2015), where people who have just behaved prosocially subsequently relax their moral constraints. If indeed households that begin separating food waste at source feel licensed to reduce their recycling of other waste fractions at the same time, we should observe a negative-sign spillover effect from the introduction of food-waste collection. On the other hand, if households value consistency in environmental and other behaviors, as in the psychological accounts of Festinger (1957) and Bem (1967), they might exhibit positive spillovers, with one behavior reinforcing the other.[1]

We consider the variables of food and packaging waste particularly suitable for spillover estimation. Given that both arise from waste-related environmental activities, they are likely to be seen by individuals as rather similar in scope and purpose, which may be important in light of evidence that spillovers are larger in magnitude for similar activities than for dissimilar ones (Ek, 2015; Reinstein, 2007). Also, Swedish recycling rates for packaging are high but not extremely high: in 2012, according to the Swedish Environmental Protection Agency, it stood at 69%; [2] thus, there is plenty of room for positive-sign as well as negative-sign spillovers.

With respect to the sign of spillovers, we do not set out to test a particular hypothesis, as prior evidence from several strands of literature is mixed. Most results from lab studies on have pointed in favor of negative spillovers in prosocial and environmental behavior. Gneezy et al. (2012) argue that spillovers are more likely to be positive when the initial prosocial act is costly, and stress that in most experimental studies the 'initial action' (to be followed by some possibly prosocial action) is essentially hypothetical, e.g. with subjects writing essays about themselves using positive words (Sachdeva et al., 2009).[3] For this reason,

[1]Dolan and Galizzi (2015) further disaggregate spillovers by the sign of the initial shift in behavior (due to the intervention). For instance, negative spillovers in prosocial behavior can be either 'permitting' or 'purging', based on whether what came first was a 'good' or a 'bad' deed; moral licensing, clearly, is an example of the former (for examples of the latter, see Ploner and Regner, 2013; Blanken et al., 2014). In the present setting, we expect that food-waste collection will cause households to source separate a larger, rather than smaller, share of food waste, and so will simply use the terms positive and negative spillovers.

[2]This figure was calculated from data in Swedish Environmental Protection Agency (2014), as a weighted average of recycling rates for each packaging type considered in this paper: glass, paper packaging, plastic packaging, and metal.

[3]Mazar and Zhong (2010) found that subjects who had simply been exposed to 'green products' — by specifying which ones they would hypothetically like to buy — subsequently behaved *more* prosocially than control subjects, while those that actually did buy green products displayed the opposite pattern. The word 'buy' is

external validity may be low.

However, even in the field, where the behaviors in question are generally costly, both positive and negative spillovers have been demonstrated. Jacobsen et al. (2012) study a green-electricity program in Memphis, Tennessee, and find that although the program did promote the consumption of clean electricity, it also increased the overall use of electricity, though only among a subset of participants. In the same vein, Tiefenbeck et al. (2013) study an information campaign to conserve water at a housing complex in Massachusetts, and find that while participants did decrease water use, they increased electricity consumption at the same time. By contrast, Carlsson et al. (2016) find positive spillovers on electricity use from a similar Colombian water-use information campaign.

Finally, Miliute-Plepiene and Plepys (2015) analyze the same behavioral spillover effect as we do here — from food-waste collection to packaging waste — and note that waste-management practitioners often claim to observe a positive-sign effect. They perform a simple before-and-after comparison within a single Swedish municipality, confirming that greater household efforts to recycle and prevent waste are observed subsequent to treatment; they argue that a set of demographic and socio-economic variables are insufficient to explain the increase. The particular municipality that they study is a socioeconomic outlier, limiting external validity; and in any case, the method used implies that results (while suggestive) cannot readily be given a causal interpretation.[4]

The present paper, by contrast, attempts to produce causal estimates, by two methods. First, we regress our packaging-waste variable on the share of households that participate in the collection of food waste, within a municipal fixed-effects framework. Second, we perform a difference-in-difference analysis by replacing the continuous treatment variable with an institutional dummy indicating whether a given municipality has in-

somewhat misleading here, however, as subjects did not pay for the products out of their own pocket.

[4]The authors also present a survey in which 47% of households state that they sort more packaging waste due to the introduction of food-waste collection.

troduced food-waste collection. Results in the former case are qualitatively and quantitatively similar to the intention-to-treat estimates of the latter case. This suggests that post-adoption increases in packaging waste are mediated by households' engagement with these waste systems, and supports the idea that behavioral spillovers are at work.

Our analysis indicates that spillovers are positive: food-waste collection causes more packaging waste to be collected. The effect is growing over time, likely due in part to slow implementation in many areas. Spillovers are on the order of 10%, but are somewhat reduced (and sometimes insignificant) when we control for certain changes in the monetary and non-monetary incentives for recycling facing households. Information campaigns and curbside collection of packaging waste from multi-family housing remain uncontrolled for in our analysis; if these were systematically adopted concurrently with food-waste collection, estimates are likely to be biased upward. There are institutional reasons to doubt that this is the case, however, and we perform robustness tests that provide no compelling evidence of remaining bias.

The remainder of this paper is organized as follows. Section 2 provides technical and institutional background on the collection of food and packaging waste. Section 3 describes our empirical strategy and the data sets that we use. Section 4 then presents results, including various robustness tests. Section 5 concludes.

4.2 Background

4.2.1 Household food waste[5]

In 2003, the Swedish government introduced a national target stating that no later than 2010, 35% or more of the food waste generated by households, restaurants, catering facilities, and grocery stores should

[5]Much of this section is based on a Swedish-language report by Swedish Waste Management (2015).

undergo biological treatment. The purpose of this target was to reduce waste incineration. There are two main types of biological treatment. First, food waste can undergo anaerobic digestion to produce biogas and digestate, with the latter used as soil conditioner. Second, it can be composted, with no biogas production; the resulting compost can be also be used as soil conditioner or fertilizer. Hence, unlike composting, digestion reclaims both the matter and the energy content of food waste.

While the target, which was the first of its kind in Sweden, was not met, an updated and somewhat refined target was adopted in 2012. This mandates a 50% biological treatment share by 2018, with 40% treated with energy recovery (i.e. by digestion).

Implementation of policies for biological treatment occurs at the municipal level, as these institutions run waste management either directly or through a (public or private) contractor. As of 2015, approximately two-thirds of Swedish municipalities have introduced systems for the source separation of food waste. A few collect food waste only from e.g. restaurants and schools. According to the latest available data (from 2013), 21% of the total (household and non-household) food waste produced in Sweden is digested, and 10% is composted. These shares have been rising for a number of years. In many municipalities introduction has been gradual, with only a subset of residential areas affected initially; it is not uncommon for introduction to begin as a pilot study in a single neighborhood, lasting for several years.

For households, two different methods for separating food from residual waste are in use. In one system, the two fractions are deposited in different containers. In the other, only one garbage can is used but food waste and residual waste is nevertheless kept separate, either in different compartments (multi-compartment bins) or in different-colored bags (e.g. green for food waste) which subsequently undergo automatic optical sorting at a specialized plant. Any food waste that is not separated at source is combusted as residual waste.

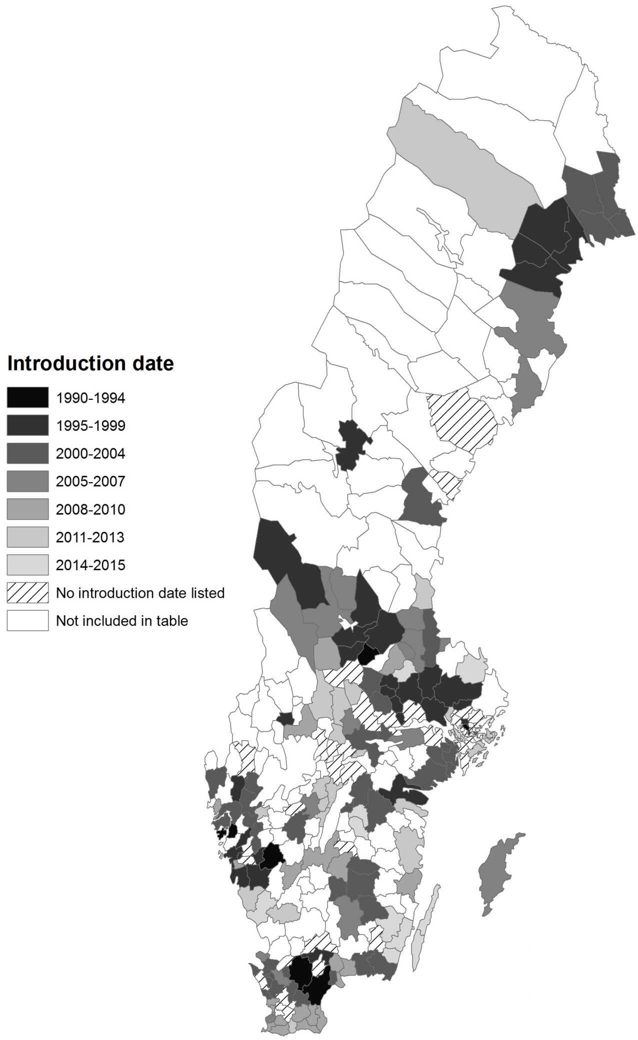
In around half of the municipalities that now collect food waste, participation is mandatory for households. Where sorting is voluntary, economic incentives such as lower waste rates for participating homes, are commonly used to encourage household efforts. In addition, municipalities often monitor sorting efforts, though household-level monitoring is typically not possible for multi-family housing, where waste is deposited anonymously. Various means are used to increase compliance: for instance, single-family households may be informed by telephone or mail that sorting efforts are unsatisfactory, and (under voluntary sorting) may be offered the higher rate as a last resort (Britta Moutakis, Swedish Waste Management, personal communication).

Figure 4.1 shows that municipalities where food waste is collected are spread across most parts of Sweden. The majority of areas where food is not (yet) source separated are rural, low population density areas, especially in the northern half of Sweden. To some extent, the same pattern holds within municipalities as well, with collection of food waste often expanding gradually outward from initial trials within population centres. This is partly because treatment facilities involve large fixed costs and, to break even, tend to require larger waste inputs than those generated in small rural towns.

4.2.2 Household packaging waste

In 1994, the Swedish government issued an ordinance (SFS 1994:1205, now 2014:1074) establishing the principle of extended producer responsibility (EPR) with respect to packaging and newspaper waste. EPR shifts (financial or physical) responsibility for waste management from municipalities to producers and retailers of the products that end up as waste (Lindhqvist, 2000). Under this system, producers are required to organize collection and management of waste products in accordance with certain regulations. They are also required to meet predefined targets (set by government) for packaging-waste collection as well as material and energy recovery for specific materials.

Liable companies can choose to implement the EPR regulations individ-



Note: this figure was constructed based on information in the Swedish Waste Management table on municipal food-waste collection, described in Section 3. Hence, 'Introduction date' refers to the beginning of implementation. 'Not included in table' is best interpreted as 'no food-waste collection'.

Figure 4.1: Introduction of food-waste collection across Swedish municipalities

ually or collectively. Overall, the response of producers and/or retailers of packaging has been to organize themselves by packaging type into five different member companies, which are themselves owners of the Packaging and Newspaper Collection Service, FTI (in Swedish, “Förpacknings- och tidningsinsamlingen”), which runs the collection and recycling of packaging and newsprint waste. Revenue from selling recycled materials is insufficient to cover costs, so FTI charges membership fees that are differentiated by the packaging type produced by each company. Although membership is voluntary, more than 90% of Swedish packaging producers and retailers have joined FTI (Miliute and Plepys, 2009).

FTI maintains roughly 6000 recycling sites where households can drop off their waste. It is also possible to drop off waste at recycling centers run by municipalities themselves, though these are typically fewer in number and further from residential areas than the FTI sites, being also designed for the collection of bulky waste. While household participation in the EPR system is mandatory in principle, it is rarely enforced. Nevertheless, most Swedish households actively sort and drop off packaging waste, so that current national targets for separate collection and recycling (set in a 2005 addition to the FTI ordinance) have been fulfilled. In 2014, almost 36% of all household waste was recycled (Swedish Waste Management, 2015).

In addition to FTI and municipal recycling sites, many municipalities have introduced systems for the curbside collection of packaging waste. Most collect packaging only from apartments, where individual property owners (municipal housing companies, housing cooperatives, etc.) typically decide independently whether or not to order curbside collection from municipal waste companies or private contractors. In the municipalities where curbside collection from apartments is common, such systems usually became widespread several years before the introduction, if applicable, of food-waste collection (Jon Nilsson-Djerf, Swedish Waste Management, personal communication).

A minority of municipalities also offer curbside collection of packaging waste from single-family homes.[6] The most common system is for households to place a pair of four-compartment bins on their property.[7] Other systems (e.g. sorting into different bags) are also in use, and some municipalities collect only a subset of packaging materials. A few municipalities introduced curbside collection concurrently with food-waste collection, though most did not. As with food-waste collection, introduction was often gradual, with initial trials involving a small number of households but lasting several years. Participation rates vary widely, from less than 1% of single-family households to all of them.

As of 2015, households that do not recycle their packaging waste face nonzero marginal cost, in the form of weight-based residual waste fees ('pay-as-you-throw'), in only 32 municipalities. The limited uptake of weight-based fees may be due to concerns that such fees provide incentives for opportunistic behavior such as illegal waste dumping, putting one's waste in a neighbor's waste container, etc. (Fullerton and Kinnaman, 1995; Bucciol et al., 2015). Indeed, two municipalities have recently discontinued their use of weight-based fees.

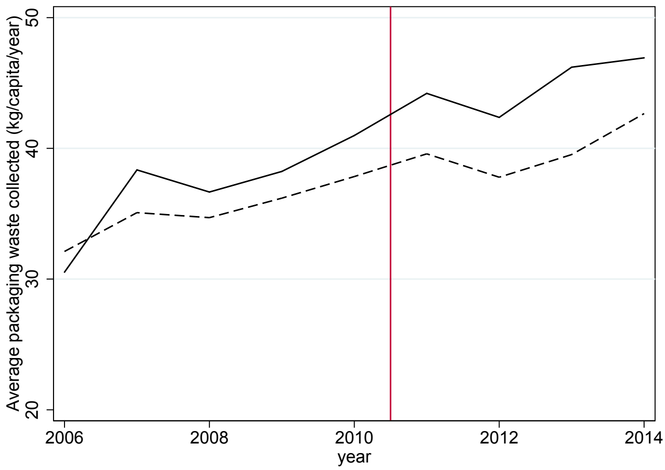
4.3 Empirical strategy and data

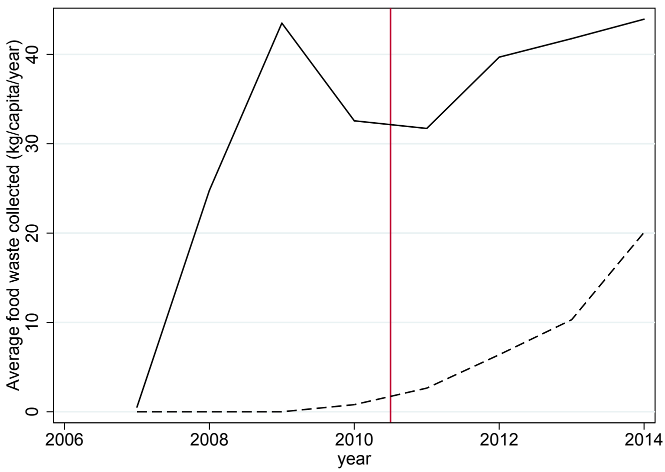
4.3.1 Empirical strategy

The upper panel of Figure 4.2 plots average collected amounts of packaging waste for two groups which differ with respect to treatment timing: municipalities where introduction occurred between 2006 and 2010 (the solid line), and in 2011 or later (dashed line), respectively. If food-waste

[6] This reflects the dissatisfaction, in some areas, with the perceived inconvenience facing households under the EPR system. Demands to increase the density of collection points have been resisted by FTI due to the costs involved. Producers have agreed, however, to increase recycling rates by 2017 and 2020, as well as to improve collection standards.

[7] The eight fractions, typically, are: colored glass, uncolored glass, metal containers, plastic packaging, paper packaging, waste paper, food waste, and residual waste.





Note: This figure depicts yearly-average amounts of packaging waste (upper panel) and food waste (lower panel) collected, divided by early adopters (solid lines) and late adopters (dashed lines). Early adopters introduced food-waste collection in 2007-2010 (left of vertical line), while late adopters introduced such systems in 2011 or later (right of vertical line). To separate the groups, the vertical line is placed midway between 2010 and 2011.

Figure 4.2: Food-waste collection by timing of introduction.

introduction has significant spillovers, we might expect early adopters (introduction to the left of the vertical line) to experience a rise in packaging-waste collection before late adopters (introduction right of vertical line) do. There is some indication of positive spillovers: in particular, packaging amounts appear to increase relatively more rapidly among early adopters in the first half of the period. This may be benchmarked against the lower panel of Figure 4.2, which focuses on direct effects. Here, using as outcome variable the average food-waste amounts collected, we find a clear pattern that food-waste amounts begin to increase earlier for early adopters than for late adopters.

Of course, these patterns should not be taken as direct evidence of causality. We estimate the causal effect of introducing food-waste collection on sorting of packaging waste by two main econometric approaches. Both exploit the staggered implementation across municipalities, eliminating all sources of omitted-variable bias that are time-invariant at the municipal level and controlling for year-specific shocks. First, we use a municipality fixed-effects framework where we regress recycled amounts of packaging on either (i) the share of households formally participating in food-waste sorting, and (ii) an indicator variable which is equal to one if a given municipality has introduced food-waste collection, and zero otherwise.[8] Second, to check for pretreatment[9] effects and to estimate the evolution of spillovers over time, we also use another regression difference-in-difference design that includes lags and leads of food-waste introduction dates.

In both cases, estimation relies on regression equations of the type

$$y_{mt} = \alpha_m + \lambda_t + \delta T_{mt} + \boldsymbol{\beta'} \boldsymbol{X}_{mt} + \epsilon_{mt} \tag{4.1}$$

where m indexes municipalities and t indexes time. The dependent vari-

[8]In the latter case, no information on exactly when households were affected is used, so those results should be interpreted as intent-to-treat estimates at the municipal level.

[9]Note that throughout the rest of this paper, the term ‘treatment’ will refer to the adoption of food-waste collection systems, and should not be confused with literal waste treatment (e.g. combustion, digestion).

able y_{mt} is given by the amount of packaging collected per capita in each municipality and year. α_m are municipality fixed effects and λ_t is a set of year dummies. Even with fixed effects, unobserved factors probably produce error terms that are autocorrelated within municipalities (Bertrand et al., 2004); thus, as is standard in this type of approach, we cluster standard errors at the municipal level. T is the treatment variable, while $\boldsymbol{X}_{mt}$ is a vector of time-varying controls. In certain specifications, we also include lags and leads of the treatment variable.

All control variables are municipality-specific and include demographic (population, population density, mean age, share of multi-family housing), socioeconomic (mean earned income, number of university graduates) and political variables (vote shares in national elections, dummies for left-wing, right-wing, or broad coalition majorities in local government). Finally, we partially control for monetary and non-monetary incentives for recycling by including data on weight-based fees, curbside collection of packaging waste from single-family homes, and the number of drop-off sites (FTI sites and municipal recycling centers) per capita. We represent weight-based fees by a dummy variable indicating whether a given municipality has introduced such incentives. For curbside collection, we use participation rates among single-family homes; due to lack of data on such rates, however, participation is typically treated as constant after introduction.

The crucial assumptions required for our difference-in-difference strategy to yield results with a causal interpretation are that (i) the treatment status of one municipality does not causally influence outcomes in other municipalities, and (ii) at each time of treatment, there should be no year-specific shocks that shift packaging recycling in the treated group relative to the non-treated group. While testing these assumptions directly is not possible, we consider (i) unlikely to pose problems, as there are typically multiple drop-off sites for packaging waste within any given municipality and no obvious incentive to deposit waste elsewhere; though admittedly we cannot rule out more subtle cross-municipality effects due to e.g. diffusion of social norms.[10]

Assumption (ii) implies that there can be no omitted-variable bias due to time-varying factors. It is clear from Figure 4.1 that introduction of food-waste collection has expanded in clusters of municipalities and thus has not been random. That non-randomness may (but need not) be due to time-varying factors related to recycling of packaging. One obvious candidate is environmental awareness within municipalities, which (if increasing) are likely to drive both the introduction of food-waste collection and increased recycling overall, leading to spillover estimates that are biased upward. We attempt to provide a proxy for environmental awareness by including the share of university graduates, as well as the summed municipal vote share received by the Green Party and the Centre Party in national elections. These parties occupy different positions on the political spectrum and are well known for being pro-environment.[11] A different approach is to argue that our outcome is a behavioral variable, while the treatment decision is institutional. Hence, given the inertia likely present in local political processes, we should probably observe an increase in packaging recycling *before* the start of food-waste collection. We will check for exactly this pattern by including leads (and lags) to regression (4.1).[12]

[10]Bucciol et al. (2015) present evidence that adoption of weight-based waste fees may cause 'waste tourism', i.e. households traveling to adjacent municipalities, where no monetary incentive for source separation exists, to drop off their waste. While only one municipality (Falkenberg) adopted weight-based fees and food-waste collection at the same time, in separate regressions (results available on request) we try including a dummy for whether a given municipality is adjacent to (at least) one other where municipality weight-based fees are used. Our central results change very little.

[11]Note that if the spillover account is to be believed, any proxy based on acts expressing engagement with environmental issues may be subject to the 'bad-control' problem (Angrist and Pischke, 2009), as it may itself form part of spillovers from food-waste introduction.

[12]A different point is that treatment-effect size may depend on the time-*in*variant factors included in municipality fixed effects. For example, suppose we hold environmental awareness constant within, but not between, municipalities. Areas where awareness is high may then be more likely to implement food-waste collection as well as be more responsive to treatment, exhibiting larger spillovers. A similar pattern may arise if local authority take expected spillovers directly into account when de-

An absence of pre-treatment effects is not enough to conclude that estimates are causal. In particular, concerns remain that adoption coincided with overall (mainly upward) shifts in the monetary and non-monetary incentives for recycling, as is clearly the case if food-waste collection is introduced at the same time as curbside collection of packaging waste. Our data allows us to partially eliminate such upward bias by controlling for weight-based fees, curbside collection from single-family homes, and the number of drop-off sites, but not from multi-family housing curbside collection or information campaigns that coincide with the introduction of food-waste collection. We will attempt to address the latter variables indirectly in our analysis.

4.3.2 The data

Our main dependent variable — collected amounts of packaging waste per capita and year — is drawn from two sources: a data set provided by FTI, and Avfall Web, which is a database maintained by Swedish Waste Management (Avfall Sverige), a Swedish stakeholder and trade asssociation representing mostly municipal waste organizations. Both data sets are municipality-year panels. The FTI data covers 2006-2014 and all 290 Swedish municipalities, but excludes packaging materials that are not collected in collaboration with an FTI member company. Avfall Web typically includes packaging collected through any channel, but does not cover 2006, has a large proportion (45%) of missing values, and reports disaggregated data only for 266 municipalities. For example, seven municipalities in the county of Västergötland co-manage waste collection through a single company, and report only aggregate data to Avfall Web. The number of nonempty cells per year is generally rising over time, from 37 (2007) to 175 (2014), with a maximum at 223 (2011).

We construct a balanced panel by using the Avfall Web data whenever possible, and the FTI data for all municipality-year pairs where Avfall

ciding whether to begin collecting food waste. Under these circumstances, estimates should be interpreted as average treatment effects on the treated.

Web lacks information.[13] The resulting data set covers 2006-2014 and includes 230 municipalities. Separate items for glass, paper packaging, plastic packaging, and metal are summed to obtain a single measure of the packaging waste collected per capita and year in each municipality-year pair.

Avfall Web also provides information on collected amounts of food waste for each municipality and year. The data reports the amounts that are composted, digested, and treated at sewage plants, respectively; we sum these into a single measure. Again, there are many missing values, and as FTI does not collect food waste, no more information is available.

Table 4.1 displays summary statistics. Per-capita amounts collected are quite comparable across the FTI and Avfall Web data sets, though slightly larger in Avfall Web, as expected. Waste amounts are right-skewed, with mean amounts above the median. This is confirmed in Figure 4.3, which highlights the cross-section dimension of our merged panel, showing that for most municipality-year pairs, less than 50 kilograms of packaging was collected per capita and year. Figure 4.4 displays the time-series dimension of the data by plotting pooled yearly averages of collected packaging amounts against time. The figure reflects a clear upward trend in recycling over time ($p = 0.000$), of about 1 kilogram per year.[14]

As our independent variable, we use either a continuous measure (share of households formally sorting food waste) or binary one (whether a given municipality has introduced food-waste collection). We construct

[13] To check the robustness of this approach, we also ran regressions using only the Avfall Web or FTI data (results available upon request). Both data sets produce estimates that are, if anything, larger and more significant than the main results reported in the present paper.

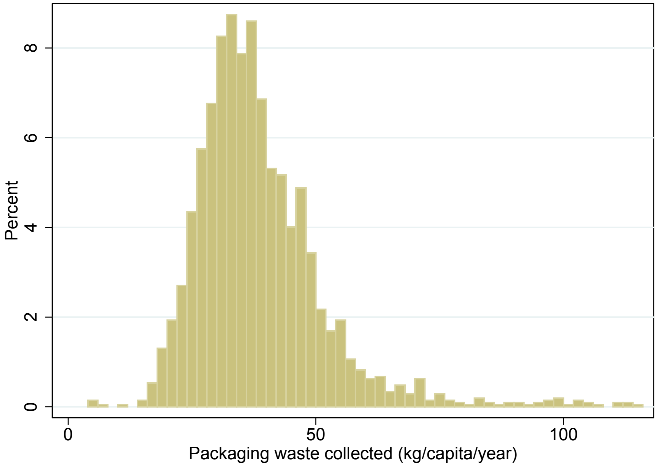
[14] The corresponding figures for food waste are not very illuminating, as more than half of all observations are then empty, and 38% of nonempty observations are equal to zero. The cross-section pooled distribution of nonzero, nonempty observations is double-peaked, with one maximum close to zero and another at roughly 50 kilograms per capita and year. Six observations have more than 100 kilograms per person and year. Yearly averages are 20-30 kg/capita/year apart from a sudden spike to around 40 kg in 2008 and 2009. There is no significant time trend in these averages.

Table 4.1: *Summary statistics: waste data*

Variable, by data source	Mean	Median	Standard deviation
FTI data			
Packaging waste	38.00	35.84	12.41
Glass	18.45	17.33	7.15
Paper packaging	12.85	12.11	5.62
Plastic packaging	4.77	3.95	3.34
Metal	1.93	1.75	0.94
Avfall Web			
Packaging waste	40.32	37.91	14.32
Glass	18.86	17.71	7.97
Paper packaging	13.67	12.54	6.96
Plastic packaging	5.65	4.88	3.61
Metal	2.06	1.76	1.60
Food waste	24.92	11.96	27.97
Share of households recycling food waste (%)	34.15	2	42.73
Merged FTI-Avfall Web data			
Packaging waste	38.64	36.32	13.48
Glass	18.50	17.32	7.69
Paper packaging	13.26	12.30	6.47
Plastic packaging	4.87	4.04	3.48
Metal	2.01	1.80	1.30

Note: values refer only to the subset of 230 municipalities included in the merged data set. All figures in kg/capita/year except 'share of households recycling food waste', which is given in percent.

the latter based on food-waste introduction dates provided by Swedish Waste Management. That information, along with data used for control variables, is further described in Appendix A; all introduction years relevant for our study are listed in Table 4.8 in Appendix B. Although by necessity we drop all municipalities for which introduction dates are known but which are not in our packaging-waste panel, we do retain as an additional control group all municipalities that are included in the



Note: $N = 2070$, but 2 outlier observations with Packaging waste > 120 lie outside the range of the figure.

Figure 4.3: The distribution of packaging waste collected (in kg/capita/year) across all observations (municipality-year pairs).

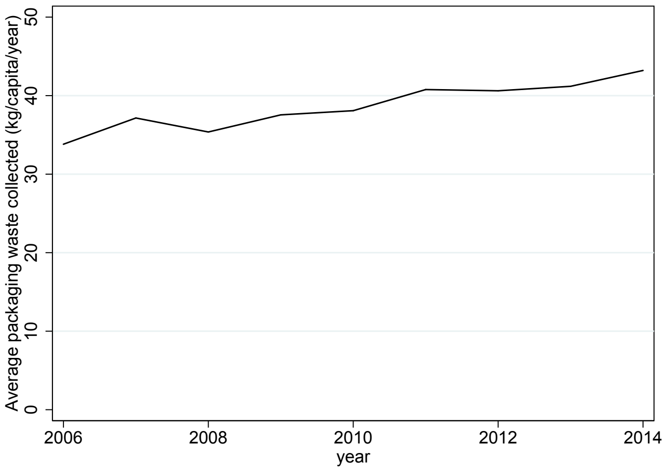


Figure 4.4: Pooled yearly averages of recycled packaging amounts recycled.

panel but lack known introdution dates. These municipalities are very likely to have never implemented food-waste collection.[15]

4.4 Results

In this section, we estimate the effect of introducing food-waste collection on household sorting of packaging. Our main regression results are presented in Table 4.2. Columns 1-3 utilizes the continuous treatment variable, while columns 4-6 repeats the same trio of regressions using the binary treatment indicator. In column 1 and 4, we regress per-capita packaging amounts collected only on treatment and time dummies. In Column 2 and 5, we include all control variables except those directly influencing incentives to recycle (weight-based fees, curbside recycling, number of drop-off sites), which are instead added in column 3 and 6. Across all regressions, spillover estimates are consistently positive and change little from inclusion of non-waste related municipal controls. However, controlling for shifts in recycling incentives causes the estimated treatment effect to deflate by up to half its initial size, while also rendering it insignificant. Clearly, adding these factors removes a degree of upward bias, probably because adoption of curbside collection of packaging from single-family homes has tended to coincide with the introduction of food-waste collection.[16]

Though insignificant, point estimates are substantial even with all controls included, corresponding to up to 9% of the data average. As conditional-on-positive participation rates for food-waste collection average 67%, estimates are broadly comparable across the two treatment variables. This is notable, as it suggests that the intention-to-treat estimates obtained in columns 4-6 of Table 4.2 are being mediated by

[15] Of the relevant 93 municipalities, we directly contacted 81, none of which stated that they had such systems in place during 2006-2014.

[16] Of the three variables added in columns 3 and 6 of Table 4.2, the curbside-collection dummy is the only one which, if dropped, single-handedly causes treatment estimates to become significant.

Table 4.2: *Spillover effects of food-waste collection on household sorting of packaging waste*

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment (continuous)	0.028**	0.035***	0.020			
	(0.014)	(0.013)	(0.014)			
Treatment (binary)				3.271***	3.524***	1.325
				(1.182)	(1.142)	(1.117)
Participation, curbside coll.			12.236**			12.488**
			(5.848)			(5.680)
Weight fee			4.455***			2.883
			(1.645)			(1.904)
Drop-off sites/capita			5,045**			4,683**
			(2,363)			(2,242)
Municipality fixed effects	YES	YES	YES	YES	YES	YES
Year dummies	YES	YES	YES	YES	YES	YES
Additional controls	NO	YES	YES	NO	YES	YES
Observations	1,343	1,343	1,283	2,070	2,070	1,355
R-squared (within)	0.111	0.125	0.144	0.156	0.181	0.141
Number of municipalities	216	216	213	230	230	217

Robust standard errors clustered at the municipal level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Additional controls are: population, population density, mean age, share of multi-family housing, mean earned income, number of university graduates, municipal government majority (left/right/broad coalition), and summed vote share for Green/Centre party. The variable 'Treatment (continuous)' gives the share of households that are (formally) participating in food-waste collection. 'Treatment (binary)' is equal to one for all municipality-year pairs where food-waste collection systems are in use.

household participation: the spillover appears to be, in essence, behavioral.

The insignificant results in columns 3 and 6 of Table 4.2 may mask substantial treatment dynamics due to the often slow rate of implementation. Thus, in Table 4.3, we subdivide (binary) treatment into distinct periods by adding lags and leads, with these differential effects presented in chronological order. Each coefficient then captures the effect, if any, at that distance to (the start of) implementation of food-waste collection. As discussed previously, this specification also checks for certain types of endogeneity: if significant effects are found prior to introduction, treatment estimates may be confounded e.g. by environmental awareness, for which we include only a proxy.[17] Starting from the baseline specification with all controls included, we add indicator variables corresponding to two years before treatment, through the year of adoption, to four years after treatment, and finally a dummy for all observations five years of more after treatment.

Spillovers in Table 4.3 are indeed seen to grow over time, consistent with slow implementation. In no case do we observe significant pre-treatment effects, suggesting that shifts in packaging waste are at least driven by factors which systematically coincide with food-waste collection. Also, even when observable incentive shifts are accounted for, third-year spillover effects are significant and estimated at 12-16% of the data average. There is also a notable drop in sorting five years or more subsequent to treatment, for which we have no immediate explanation. Allcott and Rogers (2014) do find a pattern of 'action and 'backsliding' in household efforts, where large initial behavioral changes are made in response to policy, but behavior gradually reverts to pre-intervention levels in the long term. While it is plausible that this also applies to spillovers, we will see below that the pattern does not emerge with respect to the targeted behavior itself (column 6 of Table 4.4, below).[18]

[17] There may likewise be reverse causality, with increased household efforts to sort packaging causing municipalities to adopt collection of food waste, rather than vice versa.

[18] A different possibility is that estimates are distorted by early adopters for which

Table 4.3: *Treatment-effect dynamics*

	(1)	(2)	(3)
Treatment_{t+2}	0.586	0.279	0.955
	(0.833)	(0.776)	(0.879)
Treatment_{t+1}	-0.455	-0.707	0.173
	(1.686)	(1.572)	(1.075)
Treatment_{t0}	0.708	0.685	-1.138
	(1.420)	(1.283)	(1.436)
Treatment_{t-1}	3.436**	3.667***	2.104
	(1.495)	(1.378)	(1.672)
Treatment_{t-2}	4.454**	4.657***	2.988
	(1.757)	(1.713)	(1.937)
Treatment_{t-3}	5.875***	6.222***	4.493**
	(1.722)	(1.626)	(1.855)
Treatment_{t-4}	5.223***	5.579***	4.205**
	(1.976)	(1.847)	(2.121)
$\text{Treatment}_{t-5\ \text{forward}}$	4.810**	5.338***	1.011
	(2.020)	(1.875)	(2.509)
Participation, curbside coll.			12.724**
			(5.799)
Weight fee			2.585
			(1.706)
Drop-off sites/capita			4,561**
			(2,259)
Municipality fixed effects	YES	YES	YES
Year dummies	YES	YES	YES
Additional controls	NO	YES	YES
Observations	2,070	2,070	1,355
R-squared (within)	0.162	0.189	0.154
Number of municipalities	230	230	217

Robust standard errors clustered at the municipal level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Additional controls are: population, population density, mean age, share of multi-family housing, mean earned income, number of university graduates, municipal government majority (left/right/broad coalition), and summed vote share for Green/Centre party. 'Treatment (binary)' is equal to one for all municipality-year pairs where food-waste collection systems are in use. Dummy variables Treatment_{t+2} through Treatment_{t-4} are associated with the binary treatment variable and equal to one in only the corresponding year, while $\text{Treatment}_{t-5\ \text{forward}}$ is equal to one for all observations at least five years subsequent to introduction.

Figure 4.5 provides graphical representation of these results. The upper panel corresponds to column 2 of Table 4.3, and hence does not include waste-related controls, while the lower panel is based on column 3 and includes all controls. The figure plots point estimates of leads and lags with associated 95% confidence intervals, thus depicting the evolution of collected packaging prior to, during, and after treatment, while controlling for municipality fixed effects, year effects, and municipal variables.

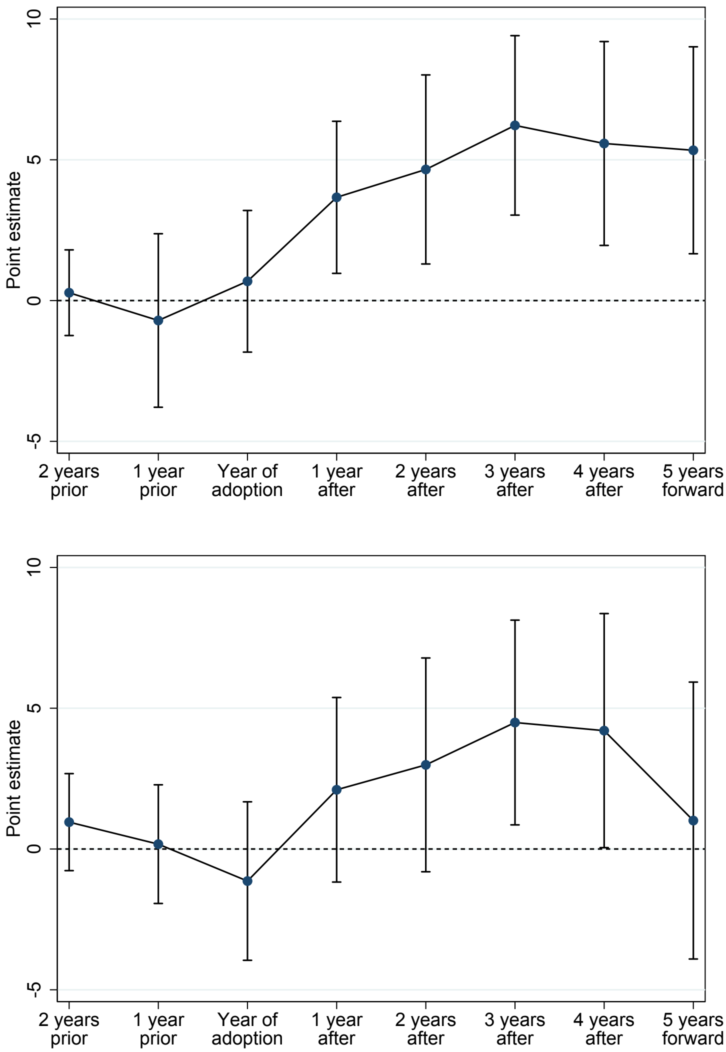
As a benchmark, we also consider the direct effect of the intervention on the source separation of food waste itself (Table 4.4). Columns 1 and 2 report results for the continuous treatment variable, with and without controls, while columns 3 and 4 repeat the analysis using the treatment dummy.[19] In all cases, we find a positive and significant treatment effect indicating, in line with Figure 4.2, that introducing food-waste collection in a given municipality causes collected amounts of food waste to increase. Adjusting for participation rates, estimates are again roughly comparable across columns, implying a sizable effect of at least 10 kg/capita/year that is little changed when control variables are added (columns 2 and 4). In columns 5 and 6, we again subdivide (binary) treatment into distinct periods by adding lags and leads. Crucially, significant effects appear only after introduction, and rise in a monotonic fashion over time.[20]

Finally, notice that the coefficient on curbside collection is everywhere positive and significant, suggesting that such systems increase food-waste recycling over and above the direct effect from introducing food-

the entire progression of lags is not identified. For example, excluding the set of municipalities treated in 2002-2006 (161 observations) produces a fifth-and-forward lag which, at 4.2 kg/capita/year ($p = 0.061$), is only a slight drop from the estimated fourth-year lag of 5.7 kg/capita/year. Picking slightly different thresholds for exclusion produces similar results.

[19] In practice, of the 1116 observations for which the treatment dummy is equal to zero, 61 report nonzero collected amounts of food waste.

[20] When also municipality-specific linear time trends are included, results become slightly negative and insignificant. Our interpretation is that the time period under study is too short to properly identify such trends, which is why we do not include them in our main analysis.



Note: The figure shows the estimated effect (controlling for municipal variables) of food-waste collection on collected packaging waste, for years before, during, and after introduction. Vertical lines denote 95% confidence intervals, i.e. 1.96 times the standard error of estimates. The raw data for the figures are taken from column 3 (upper panel) and 4 (lower panel) of Table 4.3.

Figure 4.5: Estimated treatment effect over time

Table 4.4: *The direct effect of food-waste collection on the sorting of food waste*

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment (continuous)	0.209***	0.194***				
	(0.045)	(0.045)				
Treatment (binary)			10.822***	10.811**		
			(4.059)	(4.194)		
Participation, curbside coll.		11.515***		13.617***		13.197***
		(2.169)		(3.312)		(3.228)
Weight-based fee		-5.264**		-2.473		-1.361
		(2.514)		(1.646)		(2.028)
Drop-off sites/capita		-137.1		-1,814		-1,449
		(5,721)		(5,801)		(5,751)
Treatment_{t+2}					0.272	0.373
					(1.848)	(1.634)
Treatment_{t+1}					1.619	1.541
					(2.174)	(1.978)
Treatment_{t0}					10.124*	10.109*
					(5.989)	(6.000)
Treatment_{t-1}					12.994**	12.809**
					(5.016)	(4.941)
Treatment_{t-2}					17.959***	17.928***
					(5.116)	(5.080)
Treatment_{t-3}					21.218***	20.972***
					(5.578)	(5.532)
Treatment_{t-4}					22.972***	22.652***
					(5.950)	(5.865)
$\text{Treatment}_{t-5\text{ forward}}$					29.468***	26.280***
					(5.754)	(5.522)
Municipality fixed effects	YES	YES	YES	YES	YES	YES
Year dummies	YES	YES	YES	YES	YES	YES
Additional controls	NO	YES	NO	YES	NO	YES
Observations	1,044	1,022	963	905	963	905
R-squared (within)	0.175	0.180	0.073	0.103	0.120	0.137
Number of municipalities	245	241	220	211	220	211

Robust standard errors clustered at the municipal level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Additional controls are: population, population density, mean age, share of multi-family housing, mean earned income, number of university graduates, municipal government majority (left/right/broad coalition), and summed vote share for Green/Centre party. The variable 'Treatment (continuous)' gives the share of households that are (formally) participating in food-waste collection. 'Treatment (binary)' is equal to one for all municipality-year pairs where food-waste collection systems are in use. Dummy variables Treatment_{t+2} through Treatment_{t-4} are associated with the binary treatment variable and equal to one in only the corresponding year, while $\text{Treatment}_{t-5\text{ forward}}$ is equal to one for all observations at least five years subsequent to introduction.

waste collection. This is evidence of a spillover running in the opposite direction from that examined in this paper (i.e. from packaging waste to food waste). Although potentially relevant in its own right, we do not pursue the issue further here.

4.4.1 Checking for remaining bias

As discussed in Section 3.1, some upward bias may remain in the spillover estimates reported above, though it is not clear how much. In particular, we have identified two factors that, provided they were systematically introduced at the same time as food-waste collection, may cause bias: (i) curbside collection of packaging waste from multi-family homes, and (ii) recycling information campaigns.

These omitted variables may be less problematic than those that we do include. First, as housing companies cannot 'go it alone' with respect to food waste, bias from packaging collection in multi-family housing will arise only if these agents time adoption of curbside systems to coincide with the introduction of food-waste collection at the municipal level. In reality, as we have noted, curbside collection is much more widespread among multi-family homes than single-family ones, and crucially most major housing companies that operate such systems have done so since several years prior to food-waste collection. Hence, the risk of confounding food-waste collection with curbside collection is, relatively speaking, small. Second, we expect information campaigns concurrent with treatment to target both food and packaging waste only if both types of waste are collected by the same company, which is typically not the case.[21]

So as not to proceed on faith alone, in this section we check for signs that the above concerns are valid. Regarding curbside collection from multi-family homes, if this is a cause of bias, the remaining estimated treatment effect should be smaller when estimated from a sample con-

[21] We cannot exclude the possibility that information campaigns specifically directed towards food waste (and coincided with treatment) had spillover effects on efforts to sort packaging waste.

sisting mostly of single-family homes (where we do control for curbside collection) than when estimated from a sample with mostly multi-family homes (where we do not). Moreover, if no causal spillover exists (and there are no other omitted variables of note), treatment effects in a sample of mostly single-family homes should be small and insignificant.

To see whether this is the case, in Table 4.5 we divide our data into two groups of municipalities separated by the median share of multi-family homes (approximately 33%). We then run separate regressions for the two subsamples. We consider the continuous (columns 1-2) as well as binary (columns 3-4) treatment variable, including under the assumption that introduction dates are lagged a single year (columns 5-6), to account for the apparent pattern that treatment has little impact already in the year of introduction. We find no immediate cause for concern: although spillovers in the subsample with few multi-family homes are insignificant, and relatively small more often than not, none of the cross-model Wald tests reject the null hypothesis that treatment effects are equal across subsamples.

A similar test may be performed to check for the presence of general-purpose information campaigns that target packaging waste as well as food waste, and are timed to coincide with treatment. In Section 3.1 we argued that such campaigns are more likely to occur in municipalities where both types of waste are collected by the same company, as is the case under curbside collection of packaging from single-family homes. We therefore look for the effect of information campaigns by checking whether the presence of curbside collection — prior to food-waste collection — makes a difference for subsequent spillovers. As only 17 municipalities had curbside collection in place at the time of treatment, we do not run separate regressions, but simply drop these areas from the sample.

Results are given in Table 4.6 where, for ease of comparison, we also reproduce the relevant baseline estimates from Table 4.2. Note that all regressions control for the presence, as such, of curbside collection from single-family homes. There is no indication that prior adoption

Table 4.5: *Treatment effects by municipal share of multi-family housing*

	(1) Few multi-family	(2) Many multi-family	(3) Few multi-family	(4) Many multi-family	(5) Few multi-family	(6) Many multi-family
Treatment (continuous)	0.021	0.007				
	(0.0215)	(0.015)				
Treatment (binary)			-0.320	1.349		
			(2.138)	(1.081)		
Treatment (binary), lagged					2.844	3.052***
					(2.795)	(0.998)
Participation, curbside coll.	21.004**	1.124	22.047**	1.033	20.430**	0.612
	(9.843)	(1.816)	(9.854)	(1.982)	(9.560)	(2.225)
Weight fee	7.013***	3.845*	2.152	3.084	2.020	3.040
	(1.841)	(2.209)	(2.199)	(2.226)	(2.204)	(1.987)
Drop-off sites/capita	7,889***	-757.0	8,382***	-174.8	8,283***	-234.5
	(2,388)	(2,896)	(2,331)	(2,564)	(2,294)	(2,659)
Municipality fixed effects	YES	YES	YES	YES	YES	YES
Year dummies	YES	YES	YES	YES	YES	YES
Additional controls	YES	YES	YES	YES	YES	YES
Observations	581	702	609	746	609	746
R-squared (within)	0.193	0.158	0.192	0.160	0.195	0.169
Number of municipalities	103	110	106	111	106	111

(Two-sided) cross-model Wald tests for equality of treatment-effect coefficients, p-values:

Model (1) vs. (2)	0.574		
Model (3) vs. (4)		0.391	
Model (5) vs. (6)			0.920

Robust standard errors clustered at the municipal level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Additional controls are: population, population density, mean age, share of multi-family housing, mean earned income, number of university graduates, municipal government majority (left/right/broad coalition), and summed vote share for Green/Centre party. The variable 'Treatment (continuous)' gives the share of households that are (formally) participating in food-waste collection. 'Treatment (binary)' is equal to one for all municipality-year pairs where food-waste collection systems are in use. 'Treatment (binary), lagged' is equal to one whenever food-waste collection was introduced at least one year previously. Whether a given municipality has 'few' or 'many' multi-family homes is determined by the median (across municipality means for 2006-2014) share of multi-family homes, which was 0.334.

Table 4.6: *Treatment effects, adjusting for prior presence of curbside collection of packaging from single-family homes*

	(1) Baseline	(2) No prior curb-side	(3) Baseline	(4) No prior curb-side	(5) Baseline	(6) No prior curb-side
Treatment (continuous)	0.020	0.019				
	(0.014)	(0.012)				
Treatment (binary)			1.325	2.135**		
			(1.117)	(0.931)		
Treatment (binary), lagged					3.474***	3.703***
					(1.246)	(0.940)
Participation, curbside coll.	12.236**	18.543	12.488**	19.642	11.707**	19.875
	(5.848)	(13.121)	(5.680)	(12.694)	(5.466)	(12.639)
Weight fee	4.455***	4.495**	2.883	2.697	2.790*	2.574
	(1.645)	(1.796)	(1.904)	(2.024)	(1.645)	(1.775)
Drop-off sites/capita	5,045**	5,328**	4,683**	5,015**	4,606**	4,889**
	(2,363)	(2,373)	(2,242)	(2,240)	(2,215)	(2,246)
Observations	1,283	1,179	1,355	1,247	1,355	1,247
R-squared (within)	0.144	0.127	0.141	0.127	0.148	0.133
Number of municipalities	213	197	217	201	217	201

(Two-sided) cross-model Wald tests for equality of treatment-effect coefficients, p-values:

Model (1) vs. (2)	0.810	
Model (3) vs. (4)	0.188	
Model (5) vs. (6)	0.723	

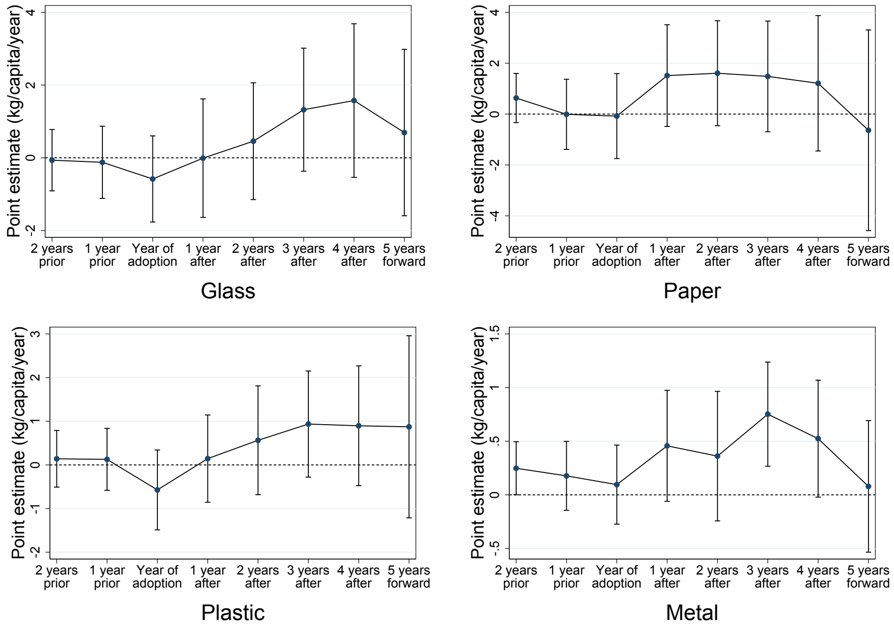
Robust standard errors clustered at the municipal level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Additional controls are: population, population density, mean age, share of multi-family housing, mean earned income, number of university graduates, municipal government majority (left/right/broad coalition), and summed vote share for Green/Centre party. The variable 'Treatment (continuous)' gives the share of households that are (formally) participating in food-waste collection. 'Treatment (binary)' is equal to one for all municipality-year pairs where food-waste collection systems are in use. 'Treatment (binary), lagged' is equal to one whenever food-waste collection was introduced at least one year previously. 'Prior curbside' applies to all municipalities which had curbside-collection systems for packaging waste in place at the time of treatment; hence, it includes municipalities that adopted both systems at the same time.

of curbside-collection systems modify the spillovers, as cross-model test statistics are all insignificant. Indeed, if anything, point estimates are larger for the restricted sample, which is the opposite of what one would expect if information campaigns are a cause of bias.

4.4.2 Extension: individual waste fractions

So far, we have concerned ourselves with packaging waste as a whole. Yet packaging is itself composed of several waste fractions: glass, paper packaging, plastic packaging, and metal. Data on each are available in our merged panel, and it is therefore possible to check for spillovers on individual types of packaging. This exercise can be framed as a robustness test: since households face very similar conditions in recycling each packaging type (whether through FTI drop-off sites, municipal recycling centers, or curbside collection), there is little reason to expect large differences across fractions. Indeed, the spillover account would arguably be undermined if e.g. treatment estimates are significantly positive only for some fractions, while negative for others.

To check whether that is the case, we repeat some of the regressions of the previous section for each packaging type. These fractions are substantially more variable than packaging as a whole, as seen in Table 4.1. Our results, given in Figure 4.6 and Table 4.7, are also less precise, with point estimates that are mainly insignificant (all control variables are everywhere included), though overwhelmingly positive for our continuous and binary treatment variables. Moreover, for all fractions, coefficients on leads and lags indicate that spillovers grow subsequent to treatment in a way which is broadly in line with previous results. Thus, there is no indication that food-waste collection has affected different packaging types in qualitatively different ways. Effect sizes three or four years after adoption range from 8% (glass) to as much as 30-40% (metal).



Note: The figure shows the estimated effect (controlling for municipal variables) of food-waste collection on various packaging-waste fractions, for years before, during, and after introduction. Vertical lines denote 95% confidence intervals, i.e. 1.96 times the standard error of estimates. The raw data for the figures are taken from column 3 (glass), 6 (paper), 9 (plastic), and 12 (metal) of Table 4.3.

Figure 4.6: Estimated treatment effect over time, individual fractions

4.5 Concluding remarks

This paper has attempted to identify behavioral spillover effects from an environmental policy intervention, namely the introduction of systems for collecting food waste in Swedish municipalities. We have done so by regressing, within fixed-effects and difference-in-difference frameworks, collected amounts of packaging waste on variables representing food-waste collection. Most of our regression specifications display a positive

Table 4.7: *Main treatment effect, disaggregated waste fractions*

	(1) Glass	(2) Glass	(3) Glass	(4) Paper pack.	(5) Paper pack.	(6) Paper pack.
Treatment (continuous)	0.008			0.005		
	(0.007)			(0.008)		
Treatment (binary)		0.160			0.955	
		(0.483)			(0.640)	
Treatment_{t+2}			-0.065			0.631
			(0.430)			(0.493)
Treatment_{t+1}			-0.123			-0.009
			(0.506)			(0.703)
Treatment_{t0}			-0.581			-0.079
			(0.604)			(0.852)
Treatment_{t-1}			-0.009			1.513
			(0.831)			(1.020)
Treatment_{t-2}			0.457			1.606
			(0.819)			(1.052)
Treatment_{t-3}			1.323			1.481
			(0.863)			(1.109)
Treatment_{t-4}			1.575			1.210
			(1.077)			(1.358)
$\text{Treatment}_{t-5\text{ forward}}$			0.695			-0.636
			(1.167)			(2.012)
Participation, curbside coll.	2.910**	3.310***	3.063***	4.663	4.287	4.890*
	(1.209)	(1.102)	(1.118)	(2.979)	(2.892)	(2.951)
Weight fee	-0.122	-0.377	-0.374	2.900**	1.695*	1.430
	(0.913)	(0.989)	(0.941)	(1.126)	(0.965)	(1.104)
Drop-off sites/capita	1,220*	1,029	975.8	3,046	2,895	2,875
	(622.2)	(754.1)	(743.8)	(2,241)	(2,018)	(2,043)
Observations	1,283	1,355	1,355	1,283	1,355	1,355
R-squared (within)	0.037	0.039	0.043	0.093	0.091	0.100
Number of municipalities	213	217	217	213	217	217

Table 4.7: *Main treatment effect, disaggregated waste fractions, cont'd*

	(7) Plastic pack.	(8) Plastic pack.	(9) Plastic pack.	(10) Metal	(11) Metal	(12) Metal
Treatment (continuous)	0.00413			0.00305*		
	(0.00534)			(0.00164)		
Treatment (binary)		-0.0340			0.243	
		(0.343)			(0.166)	
$\text{Treatment}_{t\,+\,2}$			0.140			0.248*
			(0.331)			(0.126)
$\text{Treatment}_{t\,+\,1}$			0.127			0.177
			(0.362)			(0.164)
Treatment_{t0}			-0.573			0.0954
			(0.467)			(0.188)
$\text{Treatment}_{t\,-\,1}$			0.144			0.457*
			(0.510)			(0.264)
$\text{Treatment}_{t\,-\,2}$			0.564			0.361
			(0.636)			(0.308)
$\text{Treatment}_{t\,-\,3}$			0.936			0.752***
			(0.620)			(0.248)
$\text{Treatment}_{t\,-\,4}$			0.896			0.524*
			(0.700)			(0.278)
$\text{Treatment}_{t\,-\,5\text{ forward}}$			0.873			0.0788
			(1.064)			(0.313)
Participation, curbside coll.	3.687**	3.896**	3.667**	0.976	0.994*	1.104*
	(1.657)	(1.591)	(1.691)	(0.595)	(0.589)	(0.595)
Weight fee	1.992	1.918	1.944	-0.316	-0.353	-0.415
	(1.309)	(1.402)	(1.290)	(0.307)	(0.268)	(0.313)
Drop-off sites/capita	678.8	762.2	718.0	100.1	-2.722	-7.804
	(476.6)	(550.4)	(566.9)	(301.8)	(278.0)	(278.1)
Observations	1,283	1,355	1,355	1,283	1,355	1,355
R-squared (within)	0.440	0.448	0.457	0.035	0.035	0.043
Number of municipalities	213	217	217	213	217	217

Robust standard errors clustered at the municipal level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. All regressions include municipality fixed effects, time dummies, and additional controls (population, population density, mean age, share of multi-family housing, mean earned income, number of university graduates, municipal government majority, and summed vote share for Green/Centre party). The variable 'Treatment (continuous)' gives the share of households that are (formally) participating in food-waste collection. 'Treatment (binary)' is equal to one for all municipality-year pairs where food-waste collection systems are in use. Dummy variables $\text{Treatment}_{t\,+\,2}$ through $\text{Treatment}_{t\,-\,4}$ are associated with the binary treatment variable and equal to one in only the corresponding year, while $\text{Treatment}_{t\,-\,5\text{ forward}}$ is equal to one for all observations at least five years subsequent to introduction.

treatment effect which is not present prior to introduction of food-waste collection and which increases gradually over a number of years thereafter, up to roughly 10% of the population-average amount. The same pattern emerges both for packaging waste as a whole and for each separate type of packaging, although results are typically not significant in the latter case.

Given the assumption that these results can be accepted as *prima facie* evidence of spillovers, they indicate that environmental policy can have substantial spillovers that are typically not accounted for in cost-benefit assessments. Furthermore, our data paints a somewhat rosier picture than much of the literature on behavioral spillovers in moral contexts, where spillovers are most often found to be negative. However, lab studies typically omit some factors which may be crucial for driving synergies across prosocial activities. These include the signaling effect that policy may have, revealing to individuals that not only is the targeted activity more important than previously recognized, but (we might assume) so are e.g. all environmental or waste-related activities.

With respect to studies on water and electricity use (Tiefenbeck et al., 2013; Carlsson et al., 2016), we note that waste sorting may also be special in that targeted and non-targeted activities (source separation of food and packaging waste, respectively) are performed in very close temporal and spatial proximity and also have similar motives. As a result, if and when a household begins source separating food waste, a broad review of recycling behavior may be triggered at the same time. This is in line with the psychological 'habit discontinuity hypothesis', in which an initial habitual behavior will be disrupted when the context within which it occurs changes (Wood et al., 2005). The idea also bears some similarity to practice theory in sociology. This theory adopts a holistic perspective, taking as its object of study not individual activities *per se* but the entire set of images, skills, and material elements involved in the enactment of everyday 'social practices' (Shove and Pantzar, 2005), and arguing that changing a single practice is generally contingent on shifting an entire set of wider social practices (Hargreaves, 2011; Kurz et al., 2015). Hence, we should perhaps not be surprised if household sorting of different waste fractions changes at the same time and in the same direction.

Importantly however, certain concerns remain that our estimates are a product of upward omitted-variable bias. In particular, the data does not permit us to fully control for all shifts in the recycling-related (monetary and non-monetary) incentives facing households, although we point to institutional factors to argue that the remaining bias from such sources is probably small. We also test for some plausible specific causes of endogeneity without finding clear cause for concern. A different (though obvious) difficulty relates not to the validity of our results, but to their policy relevance. Specifically, because we do not control for household waste generation itself (due to lack of reliable data), the spillovers we find need not represent an environmental gain. In practice, increases in collected packaging waste may be driven by consumption of packaging, household efforts to sort waste, or any combination of the two.

While not all policies may lend themselves to rigorous empirical analysis as readily as food-waste collection, we believe that identification of spillovers is in some ways relatively straightforward: the very distance between targeted and non-targeted activities may mitigate some of the endogeneity issues inherent in traditional policy analysis. For example, the idea that policy decisions are more likely to consider direct effects than spillovers may strike empirical researchers as quite convenient. Caveats notwithstanding, then, it is hoped that at the very least this study demonstrates the feasibility of using data from natural experiments to analyze the effect of environmental policy on variables other than the one targeted.

Acknowledgments

This study was supported by the Swedish Institute as a post-doctoral scholarship for Dr. J. Miliute-Plepiene. Financial support from the Jan

Wallander and Tom Hedelius Foundation is gratefully acknowledged. The authors are thankful to Avfall Sverige for kind support with data and comments. Ola Andersson, Alessandro Martinello, Karl McShane, Natalia Montinari, Amrish Patel, and Erik Wengström provided valuable comments which greatly improved this paper.

References

H. Allcott and T. Rogers. The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation. *American Economic Review*, 104(10), 2014.

J.D. Angrist and J. Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press, Princeton, NJ, USA, 2009.

D.J. Bem. Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological Review*, 74(3), 1967.

M. Bertrand, E. Duflo, and S. Mullainathan. How Much Should We Trust Differences-in-Differences Estimates? *The Quarterly Journal of Economics*, 119(1), 2004.

I. Blanken, N. van de Ven, M. Zeelenberg, and M.H.C. Meijers. Three Attempts to Replicate the Moral Licensing Effect. *Social Psychology*, 45(3), 2014.

I. Blanken, N. van de Ven, and M. Zeelenberg. A Meta-Analytic Review of Moral Licensing. *Personality and Social Psychology Bulletin*, 41(4), 2015.

A. Bucciol, N. Montinari, and M. Piovesan. Do Not Trash the Incentive! Monetary Incentives and Waste Sorting. *Scandinavian Journal of Economics*, 117(4), 2015.

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.

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? Behavioral Spillovers in Charitable Giving, 2015. Working Paper 2015:29, Department of Economics, Lund University.

L. Festinger. *A theory of cognitive dissonance*. Stanford University Press, 1957.

D. Fullerton and T.C. Kinnaman. Garbage, Recycling, and Illicit Burning or Dumping. *Journal of Environmental Economics and Management*, 29(1), 1995.

A. Gneezy, A. Imas, A. Brown, L.D. Nelson, and M.I. Norton. Paying to Be Nice: Consistency and Costly Prosocial Behavior. *Management Science*, 58(1), 2012.

T. Hargreaves. Practice-ing behaviour change: Applying social practice theory to pro-environmental behaviour change. *Journal of Consumer Culture*, 11(1), 2011.

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.

D.T. Kaffine. Scrap Prices, Waste, and Recycling Policy. *Land Economics*, 90(1), 2014.

T. Kurz, B. Gardner, B. Verplanken, and C. Abraham. Habitual behaviors or patterns of practice? explaining and changing repetitive climate-relevant actions. *Wiley Interdisciplinary Reviews: Climate Change*, 6(1), 2015.

T. Lindhqvist. Extended Producer Responsibility in Cleaner Production, 2000.

N. Mazar and C. Zhong. Do Green Products Make Us Better People? *Psychological Science*, 21(4), 2010.

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.

J. Miliute and A. Plepys. Driving Forces for High Household Waste Recycling. lessons from Sweden. *Environmental Research, Engineering & Management*, 1(47), 2009.

J. Miliute-Plepiene and A. Plepys. Does food sorting prevent and improve sorting of household waste? a case in sweden. *Journal of Cleaner Production*, 101, 2015.

K. Palmer, H. Sigman, and M. Walls. The Cost of Reducing Municipal Solid Waste. *Journal of Environmental Economics and Management*, 33(2), 1997.

M. Ploner and T. Regner. Self-image and moral balancing: An experimental analysis. *Journal of Economic Behavior & Organization*, 93, 2013.

D. Reinstein. Substitution Between (and Motivations for) Charitable Contributions: An Experimental Study. 2007. Discussion Paper No. 648, Department of Economics, University of Essex.

S. Sachdeva, R. Iliev, and D.L. Medin. Sinning Saints and Saintly Sinners. *Psychological Science*, 20(4), 2009.

E. Shove and M. Pantzar. Consumers, Producers and Practices: Understanding the invention and reinvention of Nordic walking. *Journal of Consumer Culture*, 5(1), 2005.

Swedish Environmental Protection Agency. Avfall i sverige 2012, 2014. Report 6619.

Swedish Waste Management. Avfallsavgifter 2010: Insamling och behandling av hushållsavfall — former och utförande, 2012. Report U2012:01.

Swedish Waste Management. Miljöstyrande taxa? En vägledning till viktbaserad avfallstaxa inför beslut, vid införande och drift, 2014. Report U2014:05.

Swedish Waste Management. Hushållsavfall i siffror — kommun- och länsstatistik 2014, 2015. Report 2015:25.

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.

W. Wood, L. Tam, and M. Guerrero Witt. Changing Circumstances, Disrupting Habits. *Journal of Personality and Social Psychology*, 88 (6), 2005.

Appendices

Appendix A. Data

Treatment variables

Data on participation rates in food-waste systems is available in Avfall Web. We construct the binary treatment variable from a table, also provided by Swedish Waste Management, which lists municipal introduction dates (years) for household food-waste collection up until and including 2015. This table is freely available at `http://www.avfallsverige.se/fileadmin/uploads/Rapporter/Biologisk/Matavfallsinsamling_nov_2015.pdf`. Its first iteration, constructed in 2005, was based on an email survey where the relevant item was phrased 'When was a collection system for food waste introduced?' As introduction has often been gradual, local officials may have found this question difficult to answer. However, after 2005, municipalities not included in the initial table were emailed annually and asked whether collection had been introduced in the past year. Given the incremental nature of table updates, post-2005 introduction years are likely to reflect the beginning, rather than completion, of implementation.

The current iteration of the food-waste table includes 197 Swedish municipalities. We drop three (Halmstad, Katrineholm, Laholm) which state "No collection" for both single-family and multi-family housing. The likely reason for their inclusion in the table is that they do collect food waste from sources not relevant for our study, e.g. restaurants and schools. Though unlikely, stated introduction dates could in principle refer to collection from such institutions, with collection from residential areas starting only at a later, and unknown, date (Britta Moutakis, Swedish Waste Management, personal communication). Of the remaining municipalities, 39 are not associated with an introduction year, or state more than one introduction year; these are likewise excluded. We do retain 10 municipalities that report an interval of two or more years, rather than a single introduction year. In these cases we have used as introduction date the first year in the stated interval.

There is partial overlap between our packaging-waste panel and the introduction-date list, with none forming a strict geographical subset of the other. Merging the two, 134 out of the remaining 230 municipalities have known food-waste introduction dates in the period 1990-2015. 47 municipalities exhibit variation in the treatment variable such that adoption occurred in 2007-2014; as of 2014, these areas contained approximately 18 % of the national population, while the panel as a whole covers roughly 70 % of the population.

Control variables

Our control variables are taken from multiple sources. From Statistics Sweden, we obtain municipal data on population, population density, mean earned income, mean age, number of university graduates, shares of multi-family housing, and municipal vote shares for the Green Party and the Centre Party in national elections. Data on political majority was taken from the Swedish Association of Local Authorities and Regions. The number of drop-off sites per capita in each municipality is obtained from Avfall Web; as this variable has 35% missing observations, we include it at some cost to statistical power. Data on weight-based fees and curbside collection of packaging waste from single-family housing is drawn from Avfall Web, various Swedish-language reports (Swedish Waste Management, 2012, 2014), and in many cases by direct contact with the relevant municipal authorities. The resulting data set is presented in Table 4.9 and 4.10 in Appendix B.

Appendix B. Introduction dates

Table 4.8: *Introduction of food-waste collection from households*

Introduction year	Municipalities
1990	Stenungsund
1991	Borås
1993	Borlänge
1994	Hässleholm, Kristianstad, Sollentuna, Tjörn
1995	Bollebygd, Falun, Gagnef, Mark
1996	Ale, Leksand, Rättvik, Sala, Solna, Upplands Väsby, Uppsala, Älvdalen
1997	Enköping, Göteborg, Hallstahammar, Knivsta, Norberg, Osby, Surahammar, Östersund
1998	Boden, Fagersta, Heby, Kil, Luleå, Norrköping, Partille, Piteå, Älvsbyn
1999	Färgelanda, Kungsbacka
2000	Arboga*, Bjuv, Falköping*, Orust, Skinnskatteberg, Tanum, Trollhättan, Uddevalla, Vänersborg, Åstorp, Örebro
2001	Alingsås, Gnesta, Karlskrona, Nykvarn, Smedjebacken, Södertälje, Trosa
2002	Eksjö, Hörby, Höör, Mellerud, Nyköping, Oxelösund, Ronneby, Sundsvall, Uppvidinge, Vetlanda, Överkalix
2003	Haparanda, Kalix, Lidingö, Motala*, Pertorp, Sävsjö, Vadstena*, Övertorneå

Table 4.8: *Introduction of food-waste collection from households, cont'd*

Introduction year	**Municipalities**
2004	Bromölla, Gävle*, Höganäs, Landskrona*, Linköping, Svalöv*, Ängelholm
2005	Båstad, Klippan, Ockelbo*, Skellefteå, Växjö
2006	Eskilstuna, Malung-Sälen, Mora, Orsa, Skövde*, Älvkarleby*
2007	Gotland, Hofors*, Malmö, Nora, Sandviken*, Umeå
2008	Hedemora, Jönköping, Mölndal
2009	Burlöv, Karlstad, Kävlinge, Oskarshamn*, Sundbyberg, Trelleborg, Ystad
2010	Karlshamn, Lomma, Lysekil, Nässjö, Olofström, Simrishamn, Sjöbo, Svedala, Sölvesborg, Tomelilla, Tranemo, Vansbro, Vellinge
2011	Hällefors, Jokkmokk, Kumla, Ljusnarsberg, Tyresö
2012	Haninge, Skurup, Söderhamn, Värmdö, Västervik
2013	Filipstad, Grästorp, Hjo*, Hylte, Håbo, Karlsborg*, Karlskoga, Söderköping, Tibro*, Töreboda*, Varberg
2014	Boxholm, Falkenberg, Kalmar*, Mjölby, Mörbylånga*, Nybro*, Torsås*, Öckerö
2015	Avesta, Borgholm, Ekerö, Östhammar

Table 4.8: *Introduction of food-waste collection from households, cont'd*

Introduction year	Municipalities
Not included in Swedish Waste Management table	Alvesta, Arjeplog, Arvidsjaur, Arvika, Berg, Bjurholm, Bollnäs, Bräcke, Dals-Ed, Dorotea, Eda, Essunga, Finspång, Flen, Forshaga, Gislaved, Gnosjö, Grums, Gullspång, Gällivare, Götene, Habo, Hagfors, Hallsberg, Hammarö, Herrljunga, Hudiksvall, Hultsfred, Härjedalen, Högsby, Kinda, Kiruna, Kramfors, Krokom, Kungälv, Lessebo, Lidköping, Lilla Edet, Ljungby, Ljusdal, Lycksele, Malå, Mirestad, Markaryd, Mullsjö, Munkedal, Munkfors, Mönsterås, Nordanstig, Nordmaling, Norrtälje, Norsjö, Ovanåker, Pajala, Ragunda, Robertsfors, Sollefteå, Sorsele, Sotenäs, Staffanstorp, Storfors, Storuman, Strömstad, Strömsund, Sunne, Svenljunga, Säffle, Säter, Tidaholm, Tierp, Timrå, Tingsryd, Torsby, Tranås, Ulricehamn, Vaggeryd, Valdemarsvik, Vara, Vilhelmina, Vimmerby, Vindeln, Vingåker, Vårgårda, Vännas, Värnamo, Ydre, Åmål, Ånge, Åre, Årjäng, Åsele, Åtvidaberg, Ödeshög

* Municipality included in Avfall Web only as part of larger organization

Table reports introduction dates for food-waste collection from households, and is based on information provided by Swedish Waste Management (see Appendix A). Only municipalities with an introduction date later than 2006 and before 2015 exhibit variation with respect to treatment.

Table 4.9: *Introduction of weight-based fees*

Introduction year	**Municipalities**
1995	Varberg
1996	Eda, Emmaboda, Linköping, Ulricehamn, Vaggeryd
1997	Danderyd, Haparanda, Ånge
1998	Borgholm, Härryda, Lerum, Mönsterås, Nordmaling, Partille, Robertsfors, Skurup, Sollentuna, Umeå
2000	Bjuv (discontinued in 2007)
2003	Kalix, Sundsvall
2005	Kramfors, Storuman
2008	Gotland
2009	Vilhelmina
2010	Göteborg, Katrineholm, Örnsköldsvik
2012	Stockholm
2014	Bjurholm, Falkenberg, Vännäs

Table reports introduction dates for weight-based waste fees, and is partly based on Swedish Waste Management (2014). Only municipalities who had such systems in place at some point in 2006-2014 are included.

Table 4.10: *Introduction of curbside collection from single-family homes*

Municipality	Introduction year	Fractions collected	System type	Participation rate
Alingsås	2012	All	2 four-compartment	30%
Bjuv	2000	All	Bags (2000-06), 2 four-compartment (2007-)	55-60% (2000-06), 87% (2007-)
Botkyrka	2012	All	2 four-compartment	5-7%
Båstad	2005	All	2 four-compartment	40%
Eskilstuna	2010	All except glass	Bags and optical sorting	99%
Eslöv	2010 (trial), 2013 (whole municip.)	All	2 four-compartment	3-7% (trial), 75% (whole municip.)
Haninge	2012	All	2 four-compartment	5%
Helsingborg	2004	All	2 four-compartment	65%
Herrljunga	2002	All	Bags and crates	10%
Huddinge	2012	All	2 four-compartment	10-15%
Håbo	2004 (discontinued in 2014)	All	2 four-compartment	1-2%
Härnösand	2014	All	2 four-compartment	65%
Hässleholm	2005	All	Bags	100%
Höganäs	2004	All	Containers, bags and crates	60-70%
Hörby	2010 (trial), 2013 (whole municip.)	All	2 four-compartment	6% (trial), 75% (whole municip.)
Höör	2010 (trial), 2013 (whole municip.)	All	2 four-compartment	6% (trial), 75% (whole municip.)
Jönköping	2008 (trial), 2013 (whole municip.)	All	2 four-compartment	4% (trial), 65%
Karlshamn	2010	All	2 four-compartment	75%
Karlstad	2006 or earlier	Specified by household (all are possible)	Bags	1%
Klippan	2005	All	2 four-compartment	90%

Table 4.10: *Introduction of curbside collection from single-family homes*

Municipality	Introduction year	Fractions collected	System type	Participation rate
Kungsbacka	Before 2006	Glass, waste paper	Bags	10-15%
Laholm	2012	All	Bags	1%
Lomma	2010	All	2 four-compartment	75-80%
Lund	2001	All	2 four-compartment	75%
Lysekil	2013	All	2 four-compartment	10-15%
Mölndal	2012 (trial, discontinued in 2014)	All	2 four-compartment	4%
Norrköping	2014	All	2 four-compartment	12%
Norsjö	2005	All	Local drop-off sites (not run by FTI)	50%
Nynäshamn	2012	All	2 four-compartment	2%
Olofström	2010	All	2 four-compartment	40%
Osby	Before 2006	All	Bags, buckets, containers	100%
Perstorp	2003	All	2 four-compartment	80-90%
Salem	2012	All	2 four-compartment	5%
Sjöbo	2010	All	2 four-compartment	80-85%
Säffle	2007 (discontinued in 2014)	Glass, metal, waste paper	Containers	80-85%
Söderköping	1990s	All	Bags and crates	25%
Sölvesborg	2010	All	2 four-compartment	95%
Tjörn	2013	All	Bags	2%
Torsås*	2000 (discontinued in 2008)	All	Bags	100%
Trelleborg	2014	All	2 four-compartment	35%
Vara	2002	All	Bags and crates, or 1 four-compartment	10%
Vårgårda	2002	All	Bags and crates	10%

Table 4.10: *Introduction of curbside collection from single-family homes*

Municipality	Introduction year	Fractions collected	System type	Participation rate
Västervik	2003 (trial), 2007 (whole municip.)	Specified by household (all are possible)	1-3 two-compartment	65% (whole municip.)
Åstorp	2000	All	Bags (2000-06), 2 four-compartment (2007-)	70% (2000-06), 79% (2007-)
Älmhult	2014	All	Local drop-off sites (not run by FTI)	4%
Örkelljunga	2006	All	2 four-compartment	95%
Östra Göinge	Before 2006	All	Bags, buckets, containers	100%

* Municipality included in Avfall Web only as part of larger organization

Table reports introduction dates for curbside collection of packaging waste from single-family homes. Only municipalities who had such systems in place at some point in 2006-2014, and which collect more than one fraction, are included; a few (Danderyd, Falun, Lidingö, Täby) have curbside collection of paper waste but no other recyclable fraction. Reported participation rates are among single-family homes, and are approximations for 2006-2014; they are likely to have changed over time. As rollout to all parts of a municipality may have taken several years, 'introduction year' should be understood as the year when introduction began.