ISSN 2042-2695

CENTRE for ECONOMIC PERFORMANCE

CEP Discussion Paper No 1286

July 2014

Because I'm Worth It: A Lab-Field Experiment on the Spillover Effects of Incentives in Health

Paul Dolan
Matteo M. Galizzi

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

ESRC ECONOMIC & SOCIAL RESEARCH COUNCIL

Abstract

We conduct a controlled lab-field experiment to directly test the short-run spillover effects of one-off financial incentives in health. We consider how incentives affect effort in a physical activity task – and then how they spillover to subsequent eating behaviour. Compared to a control group, we find that low incentives increase effort and have little effect on eating behaviour. High incentives also induce more effort but lead to significantly more excess calories consumed. The key behavioural driver appears to be the level of satisfaction associated with the physical activity task, which 'licensed' highly paid subjects to indulge in more energy-dense food.

Key words: Incentives in health, spillover effects, licensing, hidden costs of incentives
JEL: C91; C93; D03; I10

This paper was produced as part of the Centre's Wellbeing Programme. The Centre for Economic Performance is financed by the Economic and Social Research Council.

The project was funded by the Centre for the Study of Incentives in Health (CSIH), from a strategic award from the Wellcome Trust Biomedical Ethics Programme (PI Marteau: 086031/Z/08/Z). We gratefully acknowledge funding from the Centre for the Study of Incentives in Health and the Wellcome Trust. We are very grateful to all the researchers at CSIH for the many discussions and comments. We also thank for comments and suggestions Richard Ashcroft, Saurabh Bhargava, Tim Besley, Luc Bovens, David Bradford, Pablo Brañas-Garza, Antonio Cabrales, Amitav Chakravarti, Robert Cialdini, Simone Ghislandi, Anett Hofmann, Alex Imas, Andrea Isoni, Bernd Irlenbusch, Mireia Jofre-Bonet, Heather Kappes, Praveen Kujal, Kate Laffan, Julian Le Grand, George Loewenstein, Sara Machado, Theresa Marteau, Robert Metcalfe, Raffaele Miniaci, Marisa Miraldo, Natalia Montinari, Daniel Navarro-Martinez, Jeroen Nieboer, Liz Plank, Marianne Promberger, Tommaso Reggiani, Dirk Sliwka, Charitini Stavropoulou, Marcos Vera Hernandez, Daniel Wiesen, and all the participants to: the ESA World Congress at Booth Business School, Chicago; the iHEA Congresses in Toronto and Dublin; the IMEBESS Workshop at Nuffield College, Oxford; the London Health Economics Workshop and the seminars in Behavioural Economics and in Quantitative Social Research at London School of Economics; and seminars at Universities Bocconi, Milan; Brescia; Cologne; Imperial College, London; and at the Wellcome Trust, London. We thank Barbara Fasolo, and Jamie Moss for their support in accessing the LSE Behavioural Research Lab (BRL) and its mailing list. Excellent research assistance by Maria De Francesco, Shira Gall, Sean O'Dea, Enrico Regiroli, and Merata Snedden is gratefully acknowledged.

Paul Dolan is an Associate of the Centre for Economic Performance and a Professor of Behavioural Science in the Department of Social Policy, London School of Economics and Political Science. Matteo Galizzi is a Fellow in the Department of Social Policy and Behavioural Science Research Lab, London School of Economics.

Published by
Centre for Economic Performance
London School of Economics and Political Science
Houghton Street
London WC2A 2AE

I. Introduction

One of the main areas of interest to economists is the analysis of how people react to different types of incentives (Camerer and Hogarth, 1999; Laffont and Martimort, 2002; Gneezy and List, 2013). By the "*basic law of behaviour*" (Gneezy et al., 2011), higher incentives lead to greater effort and performance. There is now good evidence that financial incentives can influence a range of behaviours, such as weight loss (Volpp et al. 2008; John et al. 2011, 2012; Kullgren et al. 2013); smoking cessation (Volpp et al. 2009); gym attendance (Charness and Gneezy, 2009); and children immunization (Banerjee et al., 2010). Financial and non-financial incentives are being considered as a way to bring about changes in a number of risky health behaviours, including heavy drinking, unsafe sex, unhealthy eating, and lack of physical exercise (Loewenstein et al., 2007, 2012; Marteau et al., 2009; Volpp et al. 2011).

There is also evidence, however, that highlights the 'hidden costs' of incentives (Fehr and List, 2004), including crowding out of intrinsic motivation (Frey and Oberholzer-Gee, 1997; Deci et al. 1999); changing social norms or individual beliefs about social norms (Gneezy and Rustichini, 2000a,b; Heyman and Ariely, 2004); interacting with reciprocity, reputation, and social comparison concerns (Fehr and Gachter, 1997; Rigdon, 2009; Ariely et al. 2009a; Dur et al. 2010; Gachter and Thoni, 2010; Greiner et al. 2011), and 'choking' due to the anxiety aroused by relating payment to performance (Ariely et al. 2009b). In such cases, incentives may 'backfire', in that they result in the opposite effects to the ones originally envisaged (Fehr and Falk, 2002; Benabou and Tirole, 2003, 2006; Kamenica, 2012). Policy makers obviously need to be alert to such unintended consequences (Dolan et al. 2012; Dolan and Galizzi, 2014).

The economics literature, however, has been largely silent on the *spillover effects* that incentives may have on behaviours *other than the ones directly targeted*. A recent comprehensive review of the experimental evidence in the behavioural sciences suggests that behavioural spillovers are, in fact, pervasive (Dolan and Galizzi, 2013). Consider two subsequent behaviours. On some occasions the second behaviour reinforces the first ('promoting' spillovers). As an example, subjects who wrote down as many arguments they could think of why smoking was bad for their health, then, once outside the lab, waited significantly longer before they lightened up a cigarette (Müller et al., 2009). Households who responded more frequently to health surveys, one year later had higher levels of chlorine in their stored drinking water (Zwane et al. 2011).

In many health contexts, however, the general message seems to be that two subsequent behaviours typically push into opposite directions (‘permitting‘ spillovers). For instance, subjects who took a placebo pill that they believed was a multivitamin supplement, later on indulged in less healthy choices and walked less on their way to return a pedometer than participants told the pill was a placebo (Chiou et al., 2011a,b). Wisdom et al. (2010) made the healthy sandwiches options in *Subways* restaurants more salient by reporting them in the first page of the menus, and found that “*choosing from the healthy menu may have led to a sense of deservingness upon seeing the unhealthy sandwiches that were passed up, leading people to reward themselves with higher-calorie side dishes and drinks*” (p. 171). In Epstein et al. (2010), subjects tended not only to buy more healthy, subsidized, foods, but also to change their purchases of other foods at sale in a way to increase their overall caloric intakes. Similarly, participants who imagined a scenario where they were walking 30 minutes (Werle et al. 2010), or were exposed to exercise commercials (Van Kleef et al. 2011) ate more calories in snacks than control subjects exposed to neutral scenarios.

A key question is whether monetary incentives, which are usually powerful in changing target behaviour, are also affected by spillovers. None of the above studies documenting spillovers, in fact, have explicitly looked at conditional financial incentives. In this paper, we consider the short-term spillover effects of monetary incentives in health, and shed new light on three main questions: i) whether spillover effects exist; ii) whether they depend on the size and the nature of the incentives; and iii) which behavioural mechanisms are most likely to explain the spillover effects.

In our lab-field experiment, subjects were randomly assigned to either a control or one of three treatment groups, and asked to make a real-effort physical task, consisting in stepping as many times they could for two minutes. In the control group, no conditional financial incentive was paid and subjects participated into a lottery after the task. In the high-incentive group subjects were paid 10p for each completed step, whilst in the low-incentive group they were paid 2p for each step. Finally, in the encouragement group, subjects were ‘nudged’ to work hard without any financial incentive. After the lab task all participants were offered an ostensibly unrelated free buffet lunch in a room next to the lab. Unbeknownst to subjects, the number and type of foods and drinks consumed in the lunch by each participant was recorded.

We found that, while subjects in the incentivised arms performed a significantly higher number of steps than in the control group (about 102 and 105 steps, compared to 89), subjects in the high incentive group also exhibited significant forms of spillovers: compared to the control group, they were more likely to have lunch following the task, and, there, they consumed more energy-dense items, in particular side dishes - such as crisps and sweets - and sweetened drinks. As a result, the difference between calories consumed and burned in the high incentive group is significantly higher than the analogous difference in the control (415.9 excess calories compared to 219.7). Although smaller in size, this pattern is similar for the subjects encouraged to work hard in absence of financial incentives, whereas spillovers did not occur in presence of low financial incentives. The key behavioural driver was the level of satisfaction with the stepping performance: although, in fact, they exercised as hard as their peers in the low incentives group, subjects paid high monetary incentives (and, to a lesser extent, subjects 'nudged' to exercise) felt higher levels of satisfaction with the just accomplished task (8.42, and 7.49, compared to 6.39, on a 0-10 scale). Consistent with 'permitting' spillovers and 'licensing' effects, the higher satisfaction with the task fed into feeling that subjects 'deserved a treat' and gave them 'license' to indulge more in tastier, but also more energy-dense, foods and drinks.

The rest of the paper is structured as follows. Section 2 describes the methods and the experiment. Section 3 reports the main results. Section 4 discusses the results and concludes.

II. Methods

The experiment was designed to test for short-term spillover effects in health behaviours that are relevant for both research and policy purposes. We therefore looked at one-off incentives for physical exercise – 'calories out' – and explored their possible spillovers for eating behaviour – 'calories in'. Physical exercise is of increasing interest among behavioural and health economists (Della Vigna and Malmendier, 2006; Acland and Levy, 2013; Cawley et al. 2013) and incentives for physical exercise have already been proved to be effective (Charness and Gneezy, 2009). We opted for *stepping for two minutes* because it is a familiar, effortful but not overly exhausting task, and because it results in only around 15 calories being burned (Keytel et al., 2005). As in Werle et al. (2010), Wisdom et al. (2010), and Van Kleef et al. (2011) we considered a task related to healthy eating as the non-targeted health behaviour where spillovers might occur.

The experimental protocol was approved by the LSE Research Ethics Committee and by the Board of Directors of the Centre for the Study of Incentives in Health (CSIH), which funded the experiment. All experimental sessions were run at the LSE Behavioural Research Lab (BRL) between June and September 2012. Subjects were recruited from the volunteers in the BRL mailing list (about 5,000 subjects, mostly current and former students of the University of London). There was no other eligibility or exclusion criterion to select participants. In the email invitation, subjects were not informed about the exact nature of the experiment that would be conducted, and were only told that: the experiment would last about an hour; they would receive £10 for their participation; they would have the chance to get an extra payment related to their tasks. Subjects could sign up to any of five one-hour sessions starting every hour between 11 am and 3 pm at every working day in the week.

A total of 156 subjects participated into 24 experimental sessions. At their arrival into the BRL, subjects were identified anonymously using an ID code assigned by the online recruitment system (SONA), and asked to read an informed consent form, and to sign the latter if they agreed on carry on with the experiment.[1] Next, by asking them to draw a number for their cubicle in the lab, subjects were randomly assigned to either a control group (C) or one of three treatment groups (H, L, and E, to be explained below). Randomization resulted in 40 subjects assigned to treatment H, 39 each to groups L and E, and 38 to the control group C. Treatments were equally spread across the various sessions of the day (Table 1): a Pearson chi-squared test cannot reject the null hypothesis that the experimental groups and the timeslots were independent ($p=0.887$).

[1] All subjects who turned to the BRL lab gave their consent and decided to take part to the experiment. When, before the exercise task, they were asked to confirm their consent, all subjects agreed again and regularly concluded the experiment.

The experiment started with all subjects entering the main 'laboratory' room at BRL, sitting at their assigned desk, and reading a set of written instructions.[2] In the instructions, subjects were informed that the experiment consisted in three subsequent tasks: tasks 1 and 3 were questionnaires, while task 2 was a very simple physical task. Next subjects took part in task 1, which was designed to be a 'filler' questionnaire task and asked about basic background characteristics. While they were answering the questionnaire at their cubicles in the main lab room, the subjects were individually approached and asked if they could follow the experimenter to take part in task 2.

The experimenter accompanied the subject in one of the five small rooms that can be accessed from the main lab room. In the room there was a scale, a vertical meter, a 6-inches high stepper, and a professional heart-rate meter. The experimenter then measured and weighted each subject, and immediately before the physical exercise task, also took and recorded the heart rate measure using a professional heart rate meter (in beats per minute, *bpm*). Also before the physical exercise task, subjects rated how much 'full of energy' they felt, as well as three other happiness dimensions, using 0-10 Likert scales and tests directly taken from the wellbeing literature (Dolan et al., 2011).

[2] Sample instructions for the experiment can be requested from the authors.

In all groups, subjects were then asked to do as many times as they could do in two minutes the simple task they were then showed by the experimenter, which consisted in stepping up and down on the stepper with both feet. Subjects were then asked to confirm that they wished to continue with the experiment and, if so, were invited to start.

The difference between the four groups was that, while, at the end of their task, subjects in the control group (C) participated in a lottery to win an amount of £20, subjects in treatment groups H and L were given conditional financial incentives to exercise. Subjects in the high-incentive (H) group were paid 10p for each time they were able to step, whereas subjects in the low-incentive (L) group were paid 2p for each performed step. In the last treatment (E), participants did not receive any financial incentives and were not exposed to money in any form, but rather they were instead 'nudged' to exercise hard using verbal encouragement. Every 20 seconds the experimenter said encouragement sentences like "*well done, keep going, you're doing really well, only another 40 seconds to go...*"

Immediately after the physical exercise task, the experimenter measured again the heart rate with the professional heart rate meter. Also immediately following the task, subjects were asked to rate how satisfied they were with their performance using a 0-10 Likert scale. The subjects also rated again (always using 0-10 Likert scales) how much ‘full of energy’ they felt, as well as the other happiness dimensions. The experimenter then told subjects the number of steps they had just completed, asked subjects in C to draw a number for the lottery, and announced the amount of the extra-payment for subjects in treatments H and L. Before leaving the small room with the stepper, the experiment invited the subjects to take a rest for few minutes in another small room nearby, before going back to their desk in the main lab room in order to complete the task 1 and task 3 questionnaires.

A research assistant then accompanied the subjects to one of two rooms that were prepared, ostensibly for the subjects taking a rest. The adjacent rooms had some chairs, and a table where some foods and drinks were placed. Subjects were told by the research assistant to help themselves with whatever food or drink they liked. On the table, different types and varieties of food were available: in particular a variety of high-calories sandwiches (bacon, lettuce and tomato; chicken and bacon; ham and cheese; tuna and sweet corn), low-fat and -calories sandwiches (roast chicken salad; bacon and chicken; tuna and cucumber), and vegetarian sandwiches (cheese and tomato; cheese and celery) was available. Moreover, crisps, fruits (apples), and different types of sweets (chocolate chips muffins, chocolate bars) were also available. Finally, different individual drinks (coke, orange juice, and water) were also available. All foods and drinks were bought from a leading UK supermarket chain whose policy is to report on each item well visible labels with nutritional information: in particular, each item reports on the front package a summary label where 'pies' with green, amber and red 'signposts' briefly report how well each item contributes to the recommended daily amounts (RDA) of calories, sugars, fats, saturated fats, carbohydrates, and salt. On the back of the package, each item also reports a more detailed table with full nutritional information according to the guidelines on daily allowances (GDA) recommendations. The low-fat sandwiches also report a further label saying "*Be good to yourself: less than 3% fat*".[3]

All foods and items were presented on the table within their own packages, in the same position and order to each subject. In fact, soon after that a subject had eventually consumed something and decided to leave the room to go back to the main lab room, by looking at the table (and the trashing bin) in the room, the research assistants recorded the number and type of foods and drinks

[3] A table with the main nutritional intakes in each food and drink item used in the experiment can be found in the Appendix (Table 11).

the subject had consumed. Then, the assistants also replaced the same food and drink items in exactly the same position for the next participant. This allowed us to readily calculate how many calories nutritional intakes were consumed in the lunch by each participant after the physical task. The overall set-up was designed in a way to give the impression that the buffet was merely an act of kindness for having participated to the experiment, as commonly used at LSE for taking part in surveys and research events. The design is similar to the one in Wisdom et al. (2010) and is consistent with the more general idea of finding ways to minimize the experimenter demand effect by attempting to 'obfuscate' the ultimate objective of the experiment (Zizzo, 2010). The envisaged objective to make the whole situation sounding very natural to subjects is confirmed by the fact that all participants promptly agreed they were keen on taking a rest after their stepping task (especially after having seen how much they heart rate increased) and most subjects indeed had something to eat, or at least to drink.

Subjects who had completed the physical task and the following lunch, thus went back to the main lab room where they finished the questionnaire for task 1. Next, they also answered a questionnaire for task 3, which basically served to collect control questions related to the main experimental tasks. Among other things, subjects were, in fact, asked about: i) when was the last time they had eaten something before coming to the lab; ii) what they ate in that occasion; iii) at what time, and what, they had for breakfast that day, and for breakfast, lunch and dinner the day before; iv) how much time they spent in moderate and vigorous physical activities in the last 24 hours; v) how many calories subjects thought to have burnt in the physical exercise task; vi) how many calories they thought to have eaten in the lunch buffet; and vii) their usual activities, habits, and life style (e.g. portions of different types of foods in the last week; drinking; exercising; smoking). After completing task 3, subjects in the H and L groups received payment based on their performance in task 2 (10p or 2p for each counted step), whereas subjects in C who had then drawn a number corresponding to the number of their desk were also paid £20. All subjects then read and signed a debriefing form, were thanked, and left the lab.

The experimental design allows us to directly test our three above hypotheses on the spillover effects of incentives in health. If the incentives to perform the physical activity also have spillover effects on how subjects behave in the subsequent ostensibly unrelated lunch (question i), we should observe choices that are significantly different across the H, L, and E treatments and the control group, mainly in terms of caloric intakes, but also of the overall nutritional balance of the foods and drinks consumed. The head-to-head comparison between the three treatments groups allows us to assess whether spillover effects are conditional on the size and type of incentives offered (question ii). Statistically significant differences between these treatments point against the null hypothesis that spillover effects are independent on the size of the financial incentives. Similarly, treatments H, L can be compared to treatment E, to test the null hypothesis that spillover effects do not depend on whether incentives are in the form of financial incentives or non-financial nudges.

To gain insight into the mechanisms that underpin the spillovers (question iii), we can have a closer look at the sign and nature of the short-term spillovers. If incentives lead to *less* calories consumed, then the (financial or non-financial) incentives to exercise result in 'promoting' spillovers (Dolan and Galizzi, 2013). These, among others, may be in line with '*preference for consistency*' motives similar to the ones documented by the *cognitive dissonance* or *foot-in-the-door* literatures (Festinger, 1957; Freedman and Fraser, 1966; Cialdini, Trost, and Newsom, 1995). If, on the other hand, incentives lead to *more* calories consumed, then incentives (or nudges) result in 'permitting' spillovers. Several behavioural mechanisms can explain such 'permitting' spillovers. For instance, consistently with *licensing* effects (Monin and Miller, 2001; Cain et al., 2005), the reward and 'deservingness' from being paid for effort provides a moral licence to later indulge more in food (Werle et al. 2010; Wisdom et al., 2010; Chiou et al., 2011a,b; Van Kleef et al., 2011).

Alternatively, consistent with *ego-depletion* effects, subjects who exercised hard under the strain of incentives, can then feel 'depleted' in their physical or mental energy after the task, and thus exert less self-control in restraining their intakes in the lunch (or, attempt to replete the pool of energy resources by consuming more glucose, which is converted from food into neurotransmitters: Baumeister et al., 1998; Gailliot et al., 2007).[4] By explicitly asking subjects about their satisfaction with task performance as well as about their 'energy depletion' feeling, we are able to conduct explicit regression analysis to isolate the most likely mechanism beyond the spillovers.

III. Results

Table 2 provides a description and the main descriptive statistics for all the variables in our experiment.[5] We have directly checked whether each variable in the analysis is normally distributed using a standard skewness/kurtosis test for normality. The distributions for most variables are non-normal, and, therefore, in

[4] Of course, although not previously documented by the behavioural literature, other forms of 'permitting' spillovers are possible in principle. For instance, high financial incentives can trigger higher cognitive attention to the fact of receiving money for the task. This, in turn, can induce subjects to pay more attention to the 'reward' aspect related to their participation in the experiment, and lead them to see their 'overall reward' for participating in the experiment as the sum of two parts: the financial rewards conditional to their physical task, plus the lunch offered afterwards. If so, one can speculate that, driven by the idea to maximize their '*overall reward*', subjects can be eager to '*take the most*' out of the experiment when walking into the room with the refreshments, and can thus eat more.

[5] All the tables with descriptive statistics and estimates' results are reported in the Appendix.

what follows we report the results of the Mann-Whitney (two-sample Wilcoxon rank-sum) non-parametric tests. Following random assignment of subjects to treatments, there were no significant differences in background characteristics across treatments, including age (*Age*), gender (*Female*), weight (*Weight*), self-assessed health (*SAH*), baseline levels of heart rate per minute (*HRBefore*), happiness (*Happy*), and hunger (*Hungry*), or weekly expense for accommodation (*RentExpense*, a proxy for income) (Table 3).

An overview of the main results in the overall experiment, and across each experimental group, is provided in Tables 4 and 5.

The first result concerns the effect of financial incentives for physical exercising. Paying subjects any conditional amount of money made subjects working harder in the stepping task. As can be seen in Table 4, subjects in both treatment groups with financial incentives performed a higher number of steps: the average number of steps (*Steps*) was about 102.55 in H, 105.54 in L, 92.48 in E, and 89.42 in C. The difference between the number of steps in the arms with financial incentives compared to the control group is both noticeable in size (12-15 steps more) and statistically significant (*p=0.0004* in H and *p=0.0002* in L). The difference in the level of performed steps is generally reflected in the calculated individual difference between the two measures of the heart rate immediately before and after the stepping exercise (*DiffHR*): subjects in treatments H and L showed differences in heart rates of 62.4 and 68.67 pulses, compared to 63.38 pulses in treatment E, and 51.65 pulses in C. The differences between the financial incentives treatments and the control group are statistically significant at 0.05 level (*p=0.0332* for H, *p=0.0156* in L, whereas *p=0.0778* for E).

We then estimated the individual number of calories burned in the physical exercise based on subjects' gender, age, weight, and heart rate immediately after the physical task. The calculation followed the models fitted by Keytel et al. (2005) for the cases where direct measures are not available for maximal oxygen consumption (so-called VO_2max, which, in principle, can be measured by breathing into a mask during exercising: e.g. Canning et al. 2014). Although other models can be found in the literature, Keytel et al. (2005) gender-specific equations have been widely adopted and validated and, as such, have also been used by numerous popular software, apps, and online calories calculators for physical exercises (e.g. www.shapesense.com). We thus estimated the calories burned in the physical exercise (*KcalOut*) for each subject in our sample based on their gender and age, and their directly measured weight and heart rate (after the task). The average estimated calories burned across treatments were Kcal=16.95 in H; Kcal=16.61 in L; Kcal=15.40 in E; and Kcal=13.26 in C. The differences between the financial incentives treatments and the control group are statistically significant at 0.05 level ($p=0.0293$ for H, $p=0.0376$ in L, while the difference is not statistically significant for E, with $p=0.2176$).

Interestingly, the difference in the objective indicators of physical effort across the treatments does not merely reflect an analogous pattern for the subjective perception of the effort (*DiffEnergy*), which was measured as the individual difference between the values (on a 0-10 scale) of the responses to the question on 'how full of energy' the participants felt immediately before and after the stepping task scale. While the subjects in treatments H and E showed comparable levels of decay in their perceived feeling 'full of energy' (0.225 and 0.243, respectively), a much more pronounced decrease in energy was perceived by subjects in treatment L: 1.128 out of 10. Subjects in treatment C reported to feel even fuller of energy after the physical task, with responses characterised by an *increase* of 0.131 out of 10. Only in treatment L, however, is the perceived energy decay significantly different from the control group ($p=0.4330$ for H, $p=0.0182$ for L, and $p=0.3759$ for E).

The patterns for the objectively measured and subjectively perceived effort are also substantially different from what emerges from the self-assessed level of immediate satisfaction (on a 0-10 scale) related to the just accomplished task (*ImSatisf*). Here, again, the subjects exposed to high or low financial incentives, despite performing substantially the same number of steps, showed very distinct patterns of satisfaction for the completed task. Such a divergence, however, does not mirror the one for the perceived effort: while subjects in group H rated their own performance with an outstanding 8.42, subjects in the L group rated it as 7.06, with a difference from the control group (6.39) which is statistically significant only for treatment H ($p<0.0001$ for H, $p=0.0921$ for L). Moreover, the value of satisfaction for the stepping among subjects in treatment E (7.49) is also statistically higher than the one reported in the control group ($p=0.0012$).

Thus, although changes in the level of the financial incentives did not turn out to affect objectively measured performance, it substantially altered subjects' perception of the performance. The latter, in turn, can represent a behavioural mechanism driving potential spillover effects between the incentivised task and other, subsequent, tasks. When, at the very end of the experiment, subjects were asked again how much satisfied they were with the performed task (*EndSatisf*), although all subjects reported lower levels of satisfaction than reported immediately after the task, the main differences across treatments substantially persisted over the course of the experiment: in group C subjects reported on average a satisfaction level of 6.04, which was significantly lower than in treatments H (7.9, $p<0.0001$), E (7.13, $p=0.0035$), and L (6.81, $p=0.0280$).

In the lunch after the stepping task, subjects in treatment H consumed significantly higher calories (*KcalIn*) than in the control group (432.9 versus 233.09 Kcal, $p=0.0017$). The differences in the calories consumed in treatments L and E are only marginally significant (319.23 Kcal, $p=0.0673$ and 350.12 Kcal, $p=0.0538$, respectively). Looking in greater detail to the caloric consumption for each item (Table 6), compared to the control group, subjects in treatment H consumed a significantly higher amount of calories in crisps (79.2 versus 13.89 Kcal, $p<0.0001$), drinks (86.45 versus 37 Kcal, $p=0.0001$), and sweets (97.72 versus 38.31 Kcal, $p=0.0643$). As a result, subjects in H also consumed significantly higher intakes of fats (*p=0.0004*), saturated fats (*p=0.0024*), sugars (*p<0.0001*), and salt (*p=0.0305*) than in C. The differences in the calories consumed in each category of food are not statistically significant for subjects in treatments L and E.

The number of total calories consumed, minus the above calculated number of calories burned, was then used to define a new variable (*ExcessKcal*, Table 4) capturing the net balance between 'calories in' and 'calories out' and the extent to which, during the subsequent meal, subjects in our experiment replenished the calories spent in the physical task. As it can be seen, this difference is positive and conspicuous in all treatments: subjects consumed more than what they burned for 415.94 Kcal in group H; 302.62 Kcal in L; 334.72 Kcal in E; and 219.77 Kcal in C. The excess balance, however, is statistically higher than in the control group only in treatment H (*p=0.0020*) while is only marginally higher for group E (*p=0.0682*).

A look at subjects' answers in the control questions in task 3 helps assessing the extent to which subjects were able to correctly predict the amount of calories burned (*EstKcalOut*) and consumed (*EstKcalIn*). On average subjects, thought to have burned about 80.34 Kcal in the physical task, which is clearly an over-estimate of the real number of calories burned (between 13 and 17 Kcal). The estimated number of calories burned in the two treatments with financial incentives (and higher number of steps) is significantly higher than in the control group (69.39 in H and 112.12 in L compared with 58.95 in C, with $p=0.0385$ and $p=0.0165$, respectively). A partly different picture emerges concerning the question on the estimated amount of calories consumed in the lunch. The average predicted figure of 196.76 Kcal is clearly an under-estimation of the actual caloric intakes (335.10 Kcal). The estimated number of calories consumed, however, is not statistically significantly different across the groups (200.61 Kcal in H; 170.83 in L; 246.43 in E; and 169.03 in C).

The results are fully confirmed by formal regression analysis, which allows to account and control for, broadly defined, individual heterogeneity. A second look at the data shows that the significant differences in the average number of excess calories across treatments are due to two, conceptually distinct, factors. First, while in the control group 71.05 percent of the subjects chose to have lunch, 90, 84.62, and 76.92 percent of the subjects did it in treatments H, L, and E, respectively, with the proportion being significantly higher for group H ($p=0.0169$, while only marginally higher in L, with $p=0.0756$). Second, among the ones who indeed chose to have lunch at first place, how much they decided to eat or drink also varied across treatments: the average calories consumption for the subjects who had lunch - that is, *E(KcalIn | KcalIn*>0*)* - in fact, was of 481 Kcal (SD 273.91) in group H, 377.28 Kcal (SD 269.92) in group L, 455 Kcal (SD 240.95) in group E, and 327.98 Kcal (SD 318.03) in the control group, whose average consumption is significantly different from treatments H ($p=0.0191$) and E ($p=0.0212$).

An appropriate and flexible approach to analysis thus uses a two-part (or hurdle) model. We model, accordingly, the two decisions of i) whether or not to consume anything at lunch, and then ii) how many excess calories to eat or drink, as generated by potentially different underlying mechanisms. In particular, the first part of the hurdle model consists of a binary choice whether to have lunch (*KcalIn*>0) or not (*KcalIn=0*) modelled by a standard *probit* taking values 1 or 0, respectively.[6] We then model the second part using a linear regression model in logs to describe how many calories subjects decided to consume in excess of the calories burned, given the option to have lunch, that is, *E(Ln ExcessKcal* |*ExcessKcal*>0*)*. To enhance the comparability of results in the two separate parts, we use analogous sets and subsets of regressors in the two estimations.

[6] Given the limited amount of calories that could be burned in the physical exercise, and our specific choice of food items, whose minimum caloric intake was of 47 Kcal (i.e. an apple), the fact that a subject consumed any entire food or drink item (*KcalIn > 0*) automatically implies that also their excess calories balances were in fact strictly positive (*ExcessKcal > 0*).

In Tables 7-8 in the Appendix, we thus present a set of probit estimations for i) whether the subjects decided to consume anything at the lunch after the task. The dependent variable is the observed choice to have lunch (*KcalIn* > 0) or not (*KcalIn =0*), while the explanatory and control variables are the treatment dummies (*TreatH, TreatL, TreatE*), and a set of individual characteristics including: subjects' gender (*Female*); the baseline level of hunger before the task (*Hunger*); the number of performed steps (*Steps*); the differences in heart rates (*DiffHR*); the difference in feeling full of energy (*DiffEnergy*); the subjective estimates of calories burned (*EstKcalOut*, while the estimate for calories consumed *EstKcalIn* is clearly omitted because of endogeneity); the feelings of satisfaction with the task (*ImSatisf*).

In Table 7, a simple probit model (*Model 1*) where the only independent variables are the dummies for the treatments, shows that subjects in the high incentive group were significantly more likely to have lunch after the stepping (at 0.05 level). This effect is fully robust to the introduction in the model of variables controlling for subjects' gender, level of hunger, and time passed since their last meal (*Models 2-4*), where it also emerges that hungrier subjects were, intuitively, more likely to have lunch (always significant at least at 0.05 level). The effect of being assigned to the high-incentive group appears robust to alternative specifications including variables that control for the objective or perceived level of effort, such as the number of completed steps, the feeling of being energy depleted, or the objective change in the heart rate, none of which have significant effects (*Models 5-7*).

When, in Table 8, also the subjective estimate of the calories burned in the stepping task is introduced as an explanatory variable, its effect is significant (at 0.05 level), with a corresponding reduced level of significance in the treatment H dummy (*Model 9*). Substantially the same occurs when the variable capturing the level of satisfaction with the performance is directly introduced in the analysis (*Model 10*). A specification testing the joint effect of the latter two variables suggests that the key significant driver beyond the decision to have lunch appears to be the perception of the burned calories (*Model 11*). Together, these findings signal that the higher likelihood that subjects in group H decided to have lunch is partly mediated by their satisfaction with the task, and, especially, by their perception of having burned more calories in the stepping exercise.[7]

In Tables 9-11 in the Appendix, we model the second part for ii) how many excess calories the subjects consumed in the lunch, conditional to having

[7] An explicit analysis (available on request) excludes the hypothesis of significant interaction effects between the treatment groups, on the one hand, and either the subjective perception of the calories burned, or the level of satisfaction, on the other, suggesting that the effects of the latter variables are substantially symmetric across groups.

lunch. We thus present a set of OLS regressions with heteroskedatisc-robust standard errors where the dependent variable is the log of the, individually calculated, balance between 'calories in' and 'calories-out' (*Ln ExcessKcal | ExcessKcal>0*) and the explanatory and control variables are the ones above. As can be seen in Table 9, in the simplest model where the only independent variables are the dummies for the treatments (*Model 12*), subjects in the high incentive and the 'nudge' groups consumed significantly more excess calories in the lunch (about 60-62 percent more calories than in the baseline, significant at a 0.05 level). The two treatment effects are fully robust and remain consistently significant at 0.05 level also when more variables are introduced to control for subjects' gender (with females tending to consume less calories), level of hunger, and time passed since their last meal (*Models 13-16*).

From Table 10, the effect of being assigned to either the 10p incentive or the encouragement treatment appear robust to alternative specifications controlling also for the objective or perceived level of effort: none of variables for the number of completed steps, the perceived level of energy depletion, or the objective change in the heart rate have significant effects (*Models 18-23*). The significance of the two treatment effects remains unaltered also when the analysis controls for the subjective estimate of the calories burned in the stepping task (*Model 25*, Table 8). The higher calories consumption of subjects in group H, and to a lesser extent E, however, is mostly mediated by their level of satisfaction with the just accomplished task: when it is also included in the analysis, in fact, the level of satisfaction of the subjects is significant at 0.05 level while the significance level of the treatment dummies disappears or declines (*Model 26*). The explicit inclusion of interaction terms (*Model 27*) shows that the effect of the most satisfied subjects in groups H and E is significant (at least at 0.05 level): in both treatments, an increase of one point of satisfaction within the 0-10 scale accounts for an increase of about 30 percent in the level of calories consumption compared to the baseline group (about 70 Kcal more).

Overall, the results from the two-part model estimates suggest that the likelihood to whether or not to have lunch was mainly explained by the level of hunger of the subjects, and by their subjective perception of how many calories they had burned during the stepping task. These two variables, however, were not the key drivers of how many calories subjects consumed during the lunch. The caloric intakes were instead mainly explained by participants' gender and by the high level of satisfaction with the accomplished stepping task, induced by the high monetary rewards or by the verbal encouragement. All these findings reinstate the importance of explicitly elicit and control for subjects' satisfaction with the experimental task. In absence of such a direct indicator, one could be led to conclude that, to drive the different 'permitting' spillovers following the incentivised task was merely the presence of different levels of financial incentives. In light of the direct evidence on the subjective feelings for the task, we can instead infer that the spillovers were not just triggered by the presence of different monetary incentives *per se*, but, rather, it was mediated by the impact that these had on the subjective level of satisfaction.

We also conducted a broad set of robustness checks and variants of the empirical models (all available on request) mainly by: i) using the dependent variable *ExcessKcal* in levels rather than logs; ii) focusing the analysis on a sub-sample where few 'outliers' values were 'trimmed' off (e.g. three subjects consumed more than 1,000 calories at lunch, one in each of the groups H, L, and C); iii) using *KcalIn* as an alternative dependent variable (in both levels and logs); iv) using as dependent variables either the number of calories consumed in each food category (e.g. sandwiches, sweets, drinks), or more refined classes of nutritional intakes (e.g. sugars, fats, saturated fats, sodium); v) using discrete choice models for the likelihood that a subject consumed 'healthy', or 'energy dense' items within each food category (e.g. sandwiches, side dishes, drinks). Typically, the set of robustness estimates provide substantially identical results: the general message is that subjects in treatment H tended to consume more, absolute or excess, calories; higher intakes of fats, saturated fats, sugars, and salt, and more energy-dense food and drink items.

Overall, the evidence from the main findings and the regression analysis documents the existence of spillover effects in presence of high financial incentives, and seems to bring forward particular support to an interpretation of such spillovers in light of the 'licensing' behavioural account: although they, in fact, exercised as hard as subjects who received just 2p per step, subjects rewarded with high financial incentives felt a higher sense of 'satisfaction' once they accomplished the physical exercise task. This higher feeling of satisfaction fed into a sense that they had accomplished something which 'deserved a treat'. Higher satisfaction and feelings of deservingness were also induced by verbal encouragement in the 'nudge' arm, although to a lesser extent. On the one hand, this should not come as a surprise given that our experimental manipulation repeatedly exposed the participants to the idea that they were '*doing very well*', although, in fact, they did no better than in the control group. On the other hand, this brings forward more evidence in support of the idea that it was the satisfaction with the task the key mediator of the 'licensing' spillovers.[8]

[8] This last finding is generally consistent with the evidence that verbal reinforcements, that is, telling experimental subjects that they did well on a task, have strong positive effects on self-reported measures of intrinsic satisfaction (Deci et al., 1999; Fehr and Falk, 2002; Fischbach and Choi, 2012).

IV. Discussion and conclusion

In this work, we have directly explored the existence and nature of unintended short-term spillover effects of health incentives using a controlled lab-field experiment. We tested three main questions: i) whether such spillover effects exist; ii) whether they depend on the size (high versus low) and the nature (financial versus non-financial) of the incentives; and iii) which behavioural mechanisms are most likely to drive the spillover effects.

The main results suggest that financial incentives are effective in influencing physical exercise and that different incentives have a profound effect on subsequent eating behaviour. That financial incentives work, and work better than non-financial nudges, is not too surprising to economists. What is perhaps less expected is the relative high number of steps performed by the subjects in the C group, which is most likely due to the experimenter demand effect (Bardsley, 2005; Levitt and List, 2007a,b; Zizzo, 2010): participants could have tried to 'please' the experimenter, by behaving in the way they believed the experimenter wanted them to behave.[9] Our experiment also indicates that both high and low financial incentives are equally effective in inducing directly observed physical exercise. This result complements the findings by Charness and Gneezy (2009)

[9] This may be confirmed by the fact that both close scrutiny and face-to-face communication, which have been identified by Levitt and List (2007a,b) to be factors facilitating the experimenter demand effects, were indeed present in our experimental manipulation.

who found an asymmetric effect of low and high incentives on gym attendance. More generally, our result is in line with the overall findings reviewed by Camerer and Hogarth (1999): while introducing incentives often matters, '*raising incentives from some modest level L to a higher level H is more likely to have no effect*' on average performance.

The second result on the relationship between 'calories in' and 'calories out' in our experiment is particularly remarkable. In general, slightly more calories out meant considerably more calories in. This confirms the concern that the obesity problem cannot be effectively dealt solely by incentivising physical exercise. If more physical exercise disproportionately 'works up an appetite' and feeds into a bigger meal later on, it is unlikely that intervening only on the 'calories out' channel is an effective way to curb and reverse the on-going overweight trends (Church et al. 2007; Westerterp and Speakman, 2008; Swinburn et al. 2009; Foster-Schubert et al. 2012; Sonneville et al. 2012; Saint-Maurice et al. 2014). The fact that the under-estimated calories-in did not significantly differ across experimental treatments, rejects the possibility that were just different perceptions of calories consumed which led to higher caloric intakes in treatment H and E.

The resulting difference of 219.77 excess calories in the baseline group seems partly explained by distorted subjective perceptions of the number of 'calories in'. Subjects in the C group, in fact, underestimated the number of calories they consumed by an order of about 60 calories (estimated 169.03 Kcal versus actual 233.04 Kcal). For a food and health policy perspective this is generally consistent with the idea that disclosure of full nutritional information, per se, does not seem to be sufficient to automatically prevent perceptive biases: although each food and drink item clearly reported full nutritional information and was marked with information on the most salient nutritional facts, subjects still fell prey to a systematic, and conspicuous, under-estimation of their caloric intakes (Wansink and Chandon, 2006; Chandon and Wansink, 2007; Downs et al. 2009; Currie et al. 2010; Galizzi, 2014).

In a related way, the Kcal-in/Kcal-out baseline difference of 219.77 calories is also explained, to a lesser extent, by an over-estimation of 'calories out'. Subjects in the C group overestimated the number of calories they burned by a factor of four (estimated 58.94 Kcal versus actual 13.27 kcal). The over-estimation of the calories spent in the physical exercise, paired with the under-estimation of calories consumed thus appear candidate twinned reasons to explain the mis-calculation of the baseline Kcal-in/Kcal-out balance, accounting for nearly half of it (110 of 219.77 calories). This is generally consistent with the idea that over-eating may be partly favoured by confused perception about the exact caloric and nutritional intakes, and the exact relationships between burning and consuming calories, as pointed out by the literature on the obesity and the physical inactivity pandemics (Cutler et al. 2003; Finkelstein et al. 2005; Rosin, 2008; Mazzocchi et al. 2009; Kohl et al. 2012; Ruhm, 2012; Downs et al. 2013; Papoutsi et al. 2013; Canning et al. 2014).

Our results suggest the existence of spillover effects related to high financial incentives, and, to a less extent, also to 'nudges' targeting health behaviour: subjects in groups H and E, tended to consume significantly more excess calories in the lunch than in the control group. In particular, the average difference between H and C is about 200 Kcal (or about 20 minutes more exercising). Thus, clearly, the difference in the excess calories balance between the H and C group cannot be attributed to a physiological replenishment of the calories spent in the physical exercise.

The findings directly allow us to exclude some of the behavioural accounts as plausible explanations beyond the spillovers. Since high incentives led to more, rather than less, 'calories in', we can immediately rule out all forms of *promoting* spillover. The idea that financial incentives can *decrease* the likelihood of promoting spillovers has also been discussed by Thøgersen and

Crompton (2009) and Evans et al. (2013) in the context of pro-environmental behaviours.

Among the various *permitting* spillovers we are left with, the *ego depletion* account is not a fully convincing explanation here. Subjects in group L, in fact, exercised as hard as their peers in group H in terms of completed steps, manifested even higher heart rate gradients and, perhaps more importantly, *felt* more tired (actually, *Steps*, *DiffHR* and *DiffEnergy* take the highest scores in group L). Nevertheless, subjects in L did not consume significantly more excess calories than in C. Moreover, in no specification of the regression analysis the effect of the variables that should capture the 'energy depleted' feeling turns out to be significant: neither *Steps*, nor *DiffHR*, nor *DiffEnergy* are significant determinants of the choice to whether have lunch or not, nor of the decision of how many excess calories to consume.

Also the above mentioned '*take-the-most-out-of-it*' behavioural account is at odds with our evidence. In fact, if the main behavioural explanation beyond spillovers is that subjects were led to pay more attention to the 'overall reward' (as the sum of the rewards related to the physical task, plus the lunch offered afterwards) this account would manifest itself asymmetrically across treatments H and L. In particular, coherently with the idea that people eat more when they have high desire for money (Briers et al. 2006), subjects who already earned high earnings for the physical task (H) would be *less* likely to fall prey of the ‘permitting’ spillovers than subjects who received small rewards for the main task (L). This was not the case.[10]

The most plausible explanation beyond our findings are *licensing* spillovers. This behavioural explanation is mainly underpinned by subjects in H also manifesting higher levels of immediate satisfaction with the accomplished task. Being satisfied with the task is significantly associated to a higher likelihood of having lunch and consuming more excess calories. These calories, moreover, mainly took the form of tasty, energy dense, drinks and side dishes, such as crisps, which are hard to explain with the ‘glucose repletion’ physiological roots behind ego depletion. Both aspects go along quite well with the ‘licensing’ behavioural account predicting that subjects, perhaps unconsciously, felt entitled to ‘treat’ themselves after what they saw as a good deed.

[10] Such a behavioural account, moreover, is partly in contrast with the experimenter demand effect well documented in experimental economics: rather than attempting to take the most out of the experimenter, experimental subjects tend to take actions and decisions which, they believe, can please the experimenter (see also Bardsley, 2005; Levitt and List, 2007a,b; Zizzo, 2010). Furthermore, the higher cognitive attention to the rewards could not be merely explained by priming subjects to *money* (rather than health), since also in the C group subjects participated into a lottery with a chance to win money. Thus if priming is the behavioural explanation of the higher cognitive attention to rewards, it has necessarily to do with effects specifically triggered by conditional financial incentives, rather than generic money priming. The possibility of such specific effects, however, is in contrast with the evidence that a broad range of behaviours are pervasively activated by priming the generic concept of money (Vohs et al., 2006; Zhou et al., 2009). Since spillovers are documented also in treatment E, that implies no money whatsoever, such a ‘money priming’ account is not a plausible explanation either.

Our main results are thus consistent with the findings by Wisdom et al. (2010) who also documented 'deservingness' feelings and 'licensing' effects of making healthy options more salient in the menus, in particular in the form of more indulgent choices of side dishes and drinks. Our findings relate, more generally, to several other studies documenting health-related 'permitting' spillovers (Wilcox et al. 2009; Werle et al. 2010; Chiou et al. 2011a,b; De Witt et al. 2011; Van Kleef et al. 2011), and extend the evidence also to the case of financial incentives in health, an area of increasing research and policy interest (Loewenstein et al., 2007, 2012; Marteau et al., 2009; Volpp et al. 2011; Baicker et al., 2012).

From this perspective, our main results are, more broadly, in line with the rapidly growing literature documenting '*moral licensing*' in a variety of contexts (Benabou and Tirole, 2011): from discriminatory behaviour (Monin and Miller, 2001; Effron et al., 2009) to purchasing decisions (Strahilevitz and Myers, 1998; Dhar and Simonson, 1999; Khan and Dhar, 2006), from advising (Cain et al., 2005) to charitable giving (Sachdeva et al., 2009), from pro-social behaviour (Merritt et al. 2010, 2012; Jordan, et al. 2011; Conway and Peetz, 2012; Brañas-Garza et al. 2013; Cornelissen et al. 2013; Gneezy et al. 2012; Gneezy et al. 2014) to green choices and environmental behaviour (Thøgersen, 1999; Thøgersen and

Olander, 2003; Thøgersen and Crompton, 2009; Mazar and Zhong, 2009; Jacobsen et al. 2012; Evans et al. 2013; Tiefenbeck, et al. 2013).

Our evidence also contributes to the recent attempts to identify the boundary conditions that facilitate the occurrence of permitting behavioural spillovers as opposed to promoting ones (Bargh et al. 2001; Dijksterhuis and Bargh, 2001; Schwarz and Bless, 2005; Baumeister et al. 2007; Gneezy et al. 2012; Dolan and Galizzi, 2013; Fishbach et al. 2014). Gneezy et al. (2012), for instance, show that moral licensing is more likely to occur in pro-social situations when the first behaviour is costless, while consistency effects emerge when the first behaviour involves some costs. Our results are generally in line with these findings, adding the consideration that what seems to matter is the overall 'bundle' of relative costs and incentives. In our experiment, the first behaviour was inherently costly and demanding in physical terms, pointing to promoting spillovers (if any) at the baseline. The fact that the treated subjects were paid salient sums of money (or were repeatedly praised) for such costly activity could have 'washed out' the perception of costs and converted the resulting effect into a permitting spillover. This would also explain the asymmetry across high and low financial incentives: in the latter case subjects felt more tired after an objectively identical exercise, but less satisfied of their own performance.

Furthermore, our analysis shows that it is possible to empirically disentangle, to some extent, the behavioural mechanism underlying the 'permitting' spillovers. One practical way to do it is to directly assess the level of subjects' satisfaction with a performed experimental task, and to look at how this predicts the behaviour in a following task. To the best of our knowledge, ours is the first experiment exploring this avenue, and there certainly seems to be broader scope to combine and integrate measures of satisfaction and behavioural outcomes (Dolan and Galizzi, 2014).

By documenting licensing effects of high financial rewards, our work also relates to the increasing behavioural literature on the 'hidden costs' of incentives: the general message that seems to emerge is that, although financial incentives generally work well, they may also have 'unintended consequences': for instance, they can 'crowd out' intrinsic motivation, change social norms, interact in unpredictable ways with reciprocity, reputation, and social comparison concerns, or lead subjects to feel anxious and 'choke under pressure' (Frey and Oberholzer-Gee, 1997; Fehr and Gachter, 1997; Gneezy and Rustichini, 2000a,b; Fehr and Falk, 2002; Benabou and Tirole, 2003, 2006; Heyman and Ariely, 2004; Ariely et al., 2009a,b).

At the same time, our results also relate to the experimental literature on the effects and psychological consequences of money. On the one hand, our experiment confirms that, as the key target behaviour is concerned, small variations in (modest) monetary rewards do not radically affect average performances (Camerer and Hogarth, 1999; Pokorny, 2008; Ariely et al. 2009). On the other, however, it also confirms that high and low amounts of money can have different 'carryover' effects (Lea and Webley, 2006). For instance, subjects previously primed with high amounts of money, exhibited more self-sufficient behaviour (e.g. waited longer before asking for help, helped a stranger to a lesser extent) in a following task, while this did not occur for subjects primed with low amounts of money, whose behaviour did not differ from the control group (Vohs et al., 2006, 2008).

Further research is needed to explicitly explore the degree to which spillover effects are domain-specific. Most of the evidence at date, in fact, considers licensing and other behavioural spillovers occurring within the same domain, such as health, environmental, or pro-social behaviour. There is only little evidence on whether behavioural spillovers can occur across different domains, and, if so, on whether cross-domains spillovers are more likely to be permitting or promoting (Khan and Dhar, 2006; Mazar and Zhong, 2010; Sachdeva, et al. 2009; Baird et al. 2012; Dolan and Galizzi, 2013). Our lab-field setting can be naturally extended to look at spillovers effects of incentives across different domains. For instance, one can consider incentives targeting a cognitive, instead of a physical exercise, task. Subjects, for example, can be given a bunch of papers with different puzzles or maze tests, be asked to solve as many as they can, and then be kept waiting in a room arranged with various foods and drinks. Similar cross-domains extensions are left for future work.

More generally, further evidence is due in order to map the various factors and interactions that facilitate licensing and other behavioural spillovers in presence of incentives, for instance in terms of the completeness, concreteness, proximity, and complementarity of the two behaviours, or of the various trade-offs that they involve (Dhar and Simonson, 1999; Fishbach and Dhar, 2005; Fishbach and Zhang, 2008; Mazar and Zhong, 2010; Conway and Peetz, 2012; Fishbach and Choi, 2012; Gneezy, et al., 2012; Cornelissen, et al., 2013; Dolan and Galizzi, 2013; Fischbach et al. 2014).

Two clear limitations of our study are that it considers a sample of students instead of subjects who suffer from actual health problems related to overeating or obesity; and that, by its very design, it only looks at short-term spillovers of one-off incentives. In principle, in fact, it is possible that subjects who ate more in the buffet lunch outside the lab could have then eaten less later on that same day. In theory one can even speculate that the overall nutritional intake at the end of the day could be not significantly further away from the 'optimum' across the different treatments. One could also counter-argue that if spillovers are documented even in such a short time window, they would be even more, not less, likely to occur when considering longer time frames. It is possible, however, that incentives for repeated, rather than one-off behavioural change, could instead lead to habit formation, reinforcing behaviour, or other promoting spillovers.

Notwithstanding these limitations, if further confirmed in other contexts, the finding that high, but not low, financial incentives might have 'permitting' spillovers even in the short term has a number of practical implications. From a research methodology perspective, caution is due when interpreting the results of experiments whose design involves sequences of incentivised and non-incentivised tasks: for instance, subjects who perceive they have done well in a task where they could earn high financial incentives (e.g. an experimental game, incentive-compatible tests to elicitation preferences) might later feel 'licensed' to behave differently in a following unrelated task where incentives are absent (e.g. a questionnaire, the elicitation of psychological traits, a field manipulation). From a policy perspective, in several health contexts modest financial incentives can prove to work equally well as higher financial rewards, while being more cost-effective. At the same time, small financial incentives may present lower risks of unintended spillovers which might dampen, or even totally offset, the overall envisaged benefits of an intervention. Behavioural economists, especially those seeking to inform policy, should try to broadly account for how one behaviour can spillover to the next: no behaviour, after all, sits in a vacuum.

References

Acland, D., and Levy, M. (2013). "Naivete, projection bias, and habit formation in gym attendance." Working paper available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2233004

Ariely D., Bracha A. and Meier, S. (2009a). "Doing good or doing well? Image motivation and monetary incentives in behaving prosocially." *American Economic Review*, **99**, 544-555.

Ariely, D., Gneezy, U., Loewenstein, G. and Mazar, N. (2009b), "Large Stakes and Big Mistakes", *Review of Economic Studies*, **76**, 451-469.

Baicker, K., Congdon, W.J., and Mullainathan, S. (2012), "Health Insurance Coverage and Take-Up: Lessons from Behavioural Economics", *The Milbank Quarterly*, **90** (1), 107-134.

Baird, S.J., Garfein, R.S., McIntosh, C.T., Özler, B. (2012). "Effect of a Cash Transfer Programme for Schooling on Prevalence of HIV and Herpes Simplex Type 2 in Malawi: a Cluster Randomised Trial." *The Lancet*, **379**, 1320-1329.

Banerjee, A., Duflo, E., Glennerster, R., and Kothari, D. (2010), "Improving Immunization Coverage in Rural India: a Clustered Randomized Controlled Trial Evaluation of Immunization Campaigns With and Without Incentives", *The British Medical Journal*, **340**, c2220.

Bardsley, N. (2005), "Experimental Economics and the Artificiality of Alteration", *Journal of Economic Methodology*, **12** (2), 293-51.

Bargh, J.A., Gollwitzer, P.M., Lee-Chai, A., Barndollar, K., and Trotschel, R. (2001). "The automated will: non-conscious activation and pursuit of behavioral goals." *Journal of Personality and Social Psychology*, **81**, 1014-1027.

Baumeister, R.F., Bratslavsky, E., Muraven, M., and Tice, D.M. (1998). "Ego Depletion: Is the Active Self a Limited Resource?." *Journal of Personality and Social Psychology*, **74**, 1252-1265.

Baumeister, R.F., Vohs, K.D., and Tice, D.M. (2007). "The strength model of self-control." *Current Directions in Psychological Science*, **16**, 351-355.

Benabou, R., and Tirole, J. (2003), "Intrinsic and Extrinsic Motivation", *Review of Economic Studies,* **70** (3), 489-520.

Benabou, R., and Tirole, J. (2006), “Incentives and Prosocial Behaviour”, *American Economic Review*, **96** (5), 1652-78.

Benabou, R., and Tirole, J. (2011), “Identity, Morals, and Taboos: Beliefs as Assets”, *Quarterly Journal of Economics*, **126**, 805-855.

Brañas-Garza, P., Bucheli, M., Espinosa, M.P., & Garcia-Muñoz, T., (2013). “Moral cleansing and moral licensing: experimental evidence.” *Economics and Philosophy*, **29**, 199-212.

Briers, B., Pandelaere, M., Dewitte, S., and Warlop, L. (2006). “Hungry for Money: the Desire for Caloric Resources Increases the Desire for Financial Resources and Vice Versa”, *Psychological Science*, **17**, 939-943.

Cain, D.M., Loewenstein, G., & Moore, D.A. (2005). “The dirt on coming clean: perverse effects of disclosing conflicts of interest.” *Journal of Legal Studies,* **34**, 1-25.

Camerer, C.F., and Hogarth, R.M. (1999). “The Effects of Financial Incentives in Experiments: a Review and Capital-Labor-Production Framework.” *Journal of Risk and Uncertainty*, **19** (1-3), 7-42.

Canning, K.L., Brown, R.E., Jamnik, V.K., Salmon, A., Ardern, C.I., Kuk, J.L. (2014). Individuals underestimate moderate and vigorous intensity physical activity. *PLoS ONE* **9** (5): e97927. doi:10.1371/journal.pone.0097927.

Cawley, J., Frisvold, D., Meyerhoefer, C. (2013). "The impact of physical education on obesity among elementary school children." *Journal of Health Economics*, **32**(4), 743-755.

Chandon, P., and Wansink, B. (2007). "The biasing health halos of fast-food restaurant health claims: lower calories estimates and higher side-dish consumption intentions." *Journal of Consumer Research*, **34**(3), 301-314.

Charness, G., and Gneezy, U. (2009), "Incentives to Exercise." *Econometrica*, **77** (3), 909-31.

Chiou, W.B., Wan, C.S., Wu, W.H., Lee, K.T. (2011a). "A randomized experiment to examine unintended consequences of dietary supplement use among daily smokers: taking supplements reduces self-regulation of smoking." *Addiction*, **106** (12), 2221-2228.

Chiou, W.B., Yang, C.C., and Wan, C.S. (2011b). "Ironic Effects of Dietary Supplementation: Illusory Invulnerability Created by Taking Dietary Supplements Licenses Health-Risk Behaviors." *Psychological Science*, **22**, 1081-1086.

Church, S.T., Earnest C.P., Skinner J.S., Blair S.N. (2007). "Effects of Different Doses of Physical Activity on Cardiorespiratory Fitness Among Sedentary, Overweight or Obese Postmenopausal Women with Elevated Blood Pressure." *JAMA*, **297** (19), 2081-2091.

Cialdini, R.B., Trost, M.R., & Newsom, J.T. (1995). "Preference for consistency: the development of a valid measure and the discovery of surprising behavioral implications." *Journal of Personality and Social Psychology*, **69**, 318-328.

Conway, P., and Peetz J. (2012). "When does feeling moral actually make you a better person? Conceptual abstraction moderates whether past moral deeds motivate consistency or compensatory behavior." *Personality and Social Psychology Bulletin*, **38**(7), 907-919.

Cornelissen, G., Bashshur, M.R., Rode, J., Le Menestrel, M. (2013). "Rules or consequences? The role of ethical mindsets in moral dynamics." *Psychological Science*. **24**(4), 482-488.

Currie, J., Della Vigna, S., Moretti, E., Pathania, V. (2010). "The Effect of Fast Food Restaurants on Obesity and Weight Gain." *American Economic Journal: Economic Policy*, **2**, 32-63.

Cutler, D.M., Glaeser, E.L., Shapiro, J.M. (2003). "Why have Americans become more obese?" *Journal of Economic Perspectives*, **17** (3), 93-118.

Deci, E.L., Koestner, R.M., and Ryan, R. (1999). "A Meta-Analytic Review of Experiments Examining the Effect of Extrinsic Rewards on Intrinsic Motivation." *Psychological Bulletin*, **125**, 627-668.

Della Vigna, S., Malmendier, U. (2006). "Paying not to go to the gym." *American Economic Review*, **96**, 694-719.

De Witt Huberts, J.C., Evers, C., & De Ridder, D.T.D. (2011). “License to Sin: Self-Licensing as a Mechanism Underlying Hedonic Consumption.” *European Journal of Social Psychology*, **42**, 490-496.

Dhar, R., & Simonson, I. (1999). “Making complementary choices in consumption episodes: highlighting versus balancing.” *Journal of Marketing Research*, **36**, 29-44.

Dijksterhuis, A., and Bargh, J.A. (2001). “The perception-behavior Expressway: automatic effects of social perception on social behaviour.” *Advances in Experimental Social Psychology*, **33**, 1-40.

Dolan, P., Hallsworth, M., Halpern, D., King, D., Metcalfe, R., and Vlaev, I. (2012). “Influencing Behaviour: the Mindspace Way.” *Journal of Economic Psychology*, **33** (1), 264-277.

Dolan, P., Layard, R., and Metcalfe, R. (2011). “Measuring Subjective Wellbeing for Public Policy: Recommendations on Measures”. LSE CEP Special Paper 23.

Dolan, P., Galizzi, M.M. (2013). “Like Ripples on the Pond: Behavioural Spillovers and Their Consequences for Research and Policy.” London School of Economics.

Dolan, P., Galizzi, M.M. (2014). “Getting Policy-Makers to Listen to Field Experiments.” *Oxford Review of Economic Policy*, **30**(4), forthcoming.

Downs, J.S., Loewenstein, G., and Wisdom, J. (2009). “Strategies for Promoting Healthier Food Choices.” *American Economic Review*, **99** (2), 159-164.

Downs, J.S., Wisdom, J., Wansink, B., and Loewenstein, G. (2013). “Supplementing menu labelling with calorie recommendations to test for facilitation effects.” *American Journal of Public Health*, **103** (9), 1604-1609.

Dur, R., Non, A. and Roelfsema, H. (2010). “Reciprocity and incentive pay in the workplace.” *Journal of Economic Psychology*, **31,** 676-686.

Effron, D.A., Cameron, J.S., and Monin, B. (2009), “Endorsing Obama Licenses Favoring Whites.” *Journal of Experimental Social Psychology*, **45**, 590-593.

Epstein, L.H., Dearing, K.K., Roba, L.G., and Finkelstein, E. (2010). “The Influence of Taxes and Subsidies on Energy Purchased in an Experimental Purchasing Study.” *Psychological Science*, 1-9.

Evans, L., Maio, G.R., Corner, A., Hodgetts, C.J., Ahmed, S., and Hahn, U. (2013). “Self-Interest and Pro-Environmental Behaviour.” *Nature Climate Change*, **3**, 122-125.

Falk, A., and Kosfeld, M. (2006). “The Hidden Costs of Control.” *American Economic Review*, **96** (5), 1611-30.

Fehr, E., and Falk, A. (2002). “Psychological Foundations of Incentives.” *European Economic Review*, **46**, 687-724.

Fehr, E. and Gachter, S. (1997). “Reciprocity as a contract enforcement device: Experimental evidence.” *Econometrica*, **65**, 833-860.

Fehr, E., and List, J. (2004). “The Hidden Costs and Returns of Incentives: Trust and Trustworthiness Among CEOs.” *Journal of the European Economic Association*, **2** (5), 743-71.

Festinger, L., (1957). A theory of cognitive dissonance. Stanford, CA: Stanford University Press.

Finkelstein EA, Ruhm CJ, Kosa KM, 2005. “Economic Caused and Consequences of Obesity.” *Annual Review of Public Health*, **26**, 239-257.

Fishbach, A., & Choi, J. (2012). “When thinking bout goals undermines goal pursuit.” *Organizational Behavior and Human Decision Processes*, **118**, 99-107.

Fishbach, A., & Dhar, R. (2005). “Goals as excuses or guides: the liberating effect of perceived goal progress on choice.” *Journal of Consumer Research*, **32**, 370-377.

Fishbach, A., Koo, M.,, & Finkelstein, S.R. (2014). “Motivation resulting from completed and missing actions.” In Olson J., & Zanna M.P. (Eds.). Advances in Experimental Social Psychology, **50**, in press.

Fishbach, A., & Zhang, Y. (2008). “Together or apart: when goals and temptations complement versus compete.” *Journal of Personality and Social Psychology*, **94**, 547-559.

Foster-Schubert, K.E., Alfano, C.M., Duggan, C.R., Xiao, L., Campbell, K.L., Kong, A., Bain, C., Wang, C.Y., Blackburn, G., McTiernan, A. (2012). “Effect of Diet and Exercise, Alone or Combined, on Weight and Body Composition in Overweight-to-Obese Post-Menopausal Women.” *Obesity*, **20** (8), 1628-1638, doi: 10.1038/oby.2011.76.

Freedman, J.L., & Fraser, S.C. (1966). “Compliance without pressure: the foot-in-the-door technique.” *Journal of Personality and Social Psychology*, **4**, 195-202 .

Frey, B.S., and Oberholzer-Gee, F. (1997). “The Cost of Price Incentives: an Empirical Analysis of Motivation Crowding-out.” *American Economic Review*, **87** (4), 746-55.

Gachter, S. and Thoni, C. (2010). “Social comparison and performance: Experimental evidence on the fair wage-effort hypothesis.” *Journal of Economic Behavior & Organization*, **76**, 531-543.

Gailliot, M.T., Baumeister, R.F., DeWall, C.N., Maner, J.K., Plant, E.A., Tice, D.M., Brewer, L.E., and Schmeichel, B.J. (2007). “Self-Control Relies on Glucose as a Limited Energy Source: Willpower is More Than a Metaphor.” *Journal of Personality and Social Psychology*, **92**, 325-336.

Galizzi, M.M. (2014). “What is really behavioural in behavioural health policies? And, do they work?” *Applied Economics Perspectives and Policy*, **36**(1), 25-60..

Gneezy, A., Imas, A., Brown, A., Nelson, L.D., Norton, M.I. (2012). “Paying to be nice: consistency and costly prosocial behavior.” *Management Science*, **58**(1), 179-187.

Gneezy, U., Imas, A., Madarasz, K., (2014). “Conscience accounting: emotion dynamics and social behavior.” *Management Science*, in press.

Gneezy, U., and List, J.A. (2013). The Why Axis. Random House Business, New York.

Gneezy, U., Meier, S., and Rey-Biel, P. (2011). “When and Why Incentives (Don’t) Work to Modify Behaviour.” *Journal of Economic Perspectives*, **25** (4), 191-210.

Gneezy, U., and Rustichini, A. (2000a). “Pay Enough Or Don’t Pay At All.” *Quarterly Journal of Economics*, **115** (3), 791-810.

Gneezy, U., and Rustichini, A. (2000b). “A Fine Is a Price.” *Journal of Legal Studies*, **29** (1), 1-18.

Greiner, B., Ockenfels, A. and Werner, P. (2011). “Wage transparency and performance: A real-effort experiment.” *Economics Letters*, **111**, 236-238.

Heyman, J., and Ariely, D. (2004). "Effort for Payment: a Tale of Two Markets." *Psychological Science*, **15** (11), 787-93.

Jacobsen, G.D., Kotchen, M.J., and Vandenbergh, M.P. (2012). "The behavioral response to voluntary provision of an environmental public good: evidence from residential electricity demand." *European Economic Review*, **56**, 946-960.

John L, Loewenstein G, Troxel A, Norton L, Fassbender J, Volpp K (2011). "Financial incentives for extended weight loss: a randomized, controlled trial." *Journal of General Internal Medicine*, 621-626.

John L, Loewenstein G, Volpp KG (2012). "Empirical observations on longer-term use of incentives for weight loss." *Preventive Medicine*, **55**(1), S68-S74.

Jordan, J., Mullen., E., Murnighan, J.K. (2011). "Striving for the moral self: the effects of recalling past moral actions on future moral behavior." *Personality and Social Psychology Bulletin*, **37** (5), 701-713.

Keytel, L.R., Goedecke, J.H., Noakes, T.D., Hiiloskorpi, H., Laukkanen, R., van der Merwe, L., and Lambert, E.V. (2005). "Prediction of Energy Expenditure from Heart Rate Monitoring During Submaximal Exercise." *Journal of Sports Science*, **23** (3) 289-297.

Kamenica, E. (2012). "Behavioural Economics and Psychology of Incentives." *Annual Review of Economics,* **4** (13), 1-26.

Khan, U., and Dhar, R. (2007). “Licensing Effect in Consumer Choice.” *Journal of Marketing Research*, **43**, 259-266.

Kullgren JT, Troxel AB, Loewenstein G, Asch DA, Norton LA, Wesby L, Tao Y, Zhu J, Volpp KG (2013). “Individuals vs. group-based incentives for weight loss: a randomized, controlled trial.” *Annals of Internal Medicine*, **158**(7), 505-514.

Laffont, J.J., Martimort, D. (2002). “The Theory of Incentives: the Principal-Agent Model, Princeton University Press.

Lea, S.E.G., and Webley, P. (2006). “Money as a Tool, Money as a Drug: the Biological Psychology of a Strong Incentive.” *Behavioural and Brain Sciences*, **29** (2), 161-209.

Levitt, S., and List, J (2007a). “What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?” *Journal of Economic Perspectives*, **21** (2), 153-74.

Levitt, S., and List, J. (2007b). “On the Generalizability of Lab Behaviour in the Field.” *Canadian Journal of Economics*, **40** (2), 343-370.

Loewenstein, G., Troyen, B., and Volpp, K. (2007). “Asymmetric Paternalism to Improve Health Behaviors.” *JAMA*, **298** (20), 2415-17.

Loewenstein, G., Asch, D.A., Friedman, J.Y., Melichar, L.A., and Volpp, K.G. (2012). “Can behavioural economics make us healthier?” *British Medical Journal*, **344**: e3482.

Marteau, T., Ashcroft, R., and Oliver, A. (2009), "Using Financial Incentives to Achieve Healthy Behaviour." *BMJ*, **338**:b1415.

Mazar. N., & Zhong, C.B. (2010). "Do green products make us better people?" *Psychological Science,* **21** (4), 494-498.

Mazzocchi, M., Traill, W.B., and Shogren, J.F. (2009), Fat Economics: Nutrition, Health, and Economic Police, Oxford University Press.

Merritt, A.C., Effron, D., Monin, B. (2010). "Moral Self-Licensing: When Being Good Frees Us to Be Bad." *Social and Personality Psychology Compass*, **4/5**, 344-357.

Merritt, A.C., Effron, D.A., Fein, S., Savitsky, K.K., Tuller, D.M., Monin, B. (2012). "The strategic pursuit of moral credentials." *Journal of Experimental Social Psychology*, **48**, 774-777.

Monin, B., and Miller, D.T. (2001). "Moral Credentials and the Expression of Prejudice." *Journal of Personality and Social Psycholog*y, **81**, 33-43.

Müller, B.C.N., van Baaren, R.B., Ritter, S.M., Woud, M.L., Bergmann, H., Harakeh, Z., Engels, R.C.M.E., and Dijksterhuis, A. (2009). "Tell Me Why: The Influence of Self-Involvement on Short Term Smoking Behaviour." *Addictive Behaviors*, **34**, 427-431.

Papoutsi GS, Drichoutis AC, Nayga RM Jr, 2013. "The causes of childhood obesity: a survey." *Journal of Economic Surveys*, 27(4), 743-767.

Pokorny, K. (2008). "Pay, But Do Not Pay Too Much. An Experimental Study on the Impact of Incentives." *Journal of Economic Behaviour and Organization*, **66**, 251-264.

Rigdon, M. (2009). "Trust and reciprocity in incentive contracting." *Journal of Economic Behavior & Organization*, **70**, 93-105.

Rosin O. (2008). "The Economic Causes of Obesity: a survey." *Journal of Economic Surveys*. **22**(4), 617-647.

Ruhm C. (2012). "Understanding Overeating and Obesity." *Journal of Health Economics*, **31**(6), 781-796.

Sachdeva, S., Iliev, R., and Medin, D.L. (2009). "Sinning Saints and Saintly Sinners: the Paradox of Moral Self-Regulation." *Psychological Science*, **20** (4), 523-528.

Saint-Maurice, P.F., Welk, G., Russell, D.W., Huberty, J., (2014). "Moderating Influences of Baseline Activity Levels in School Physical Activity Programming for Children: the Ready for Recess Project." *BMC Public Health*, **13** (103), doi: 10.1186/1471-2458-14-103.

Sonneville K.R., Rifas-Shiman, R.L., Kleinman, K., Gortmaker, S., Gilman, M.W., Taveras, E.M. (2012). "Associations of Obesogenic Behaviors in

Mothers and Obese Children Participating in a Randomized Trial.” *Obesity*, **20** (7), doi:10.1038/oby.2012.43.

Schwarz, N., & Bless, H. (2005). “Mental construal processes: the inclusion/exclusion model.” In D. Stapel & J. Suls (Eds.) *Assimilation and contrast in social psychology* (pp. 119-141). Philadelphia: Pshychology Press.

Strahilevitz, M., & Myers, J. (1998). “Promised donations to charity as purchase incentives: how well they work may depend on what you’re trying to sell.” *Journal of Consumer Research*, **24** (4), 434-446.

Swinburn, B., Sacks, G., and Ravussin, E. (2009). “Increased Food Energy Supply Is More Than Sufficient to Explain the US Epidemic of Obesity.” *American Journal of Clinical Nutrition*, **90**, 1453-1456.

Thøgersen, J. (1999). “Spillover processes in the development of a sustainable consumption pattern.” *Journal of Economic Psychology,* **20**, 53-81.

Thøgersen, J., Olander, F. (2003). “Spillover of Environment-friendly consumer behavior.” *Journal of Environmental Psychology*, **32**, 141-163.

Thøgersen, J. and Crompton, T. (2009). “Simple and Painless? The Limitations of Spillover in Environmental Campaigning.” *Journal of Consumer Policy*, **32**, 141-163.

Tiefenbeck V., Staake T., Roth K., Sachs O. (2013). “For better or for worse? Empirical evidence of moral licensing in a behavioral energy conservation campaign.” *Energy Policy*, **57**, 160-171.

Van Kleef, E., Shimizu, M., Wansink, B. (2011). "Food compensation: do exercise ads change food intake?" *International Journal of Behavioral Nutrition and Physical Activity*, **8**(6), 661-664.

Vohs, K., Mead, N.L., and Goode, M.R. (2006). "The Psychological Consequences of Money." *Science*, **314**, 1154-1156.

Volpp, K., Troxel, A.B., Pauly, M.V., Glick, H.A., Ash, D.A., Galvin, R., Zhu, J., Wan, F., DeGuzman, J., Corbett, E., Weiner, J., and Audrain-McGovern, J. (2009). "A Randomized, Controlled Trial of Financial Incentives for Smoking Cessation." *The New England Journal of Medicine*, **360**, 699-709.

Volpp, K., John, L.K., Troxel, A.B., Norton, L., Fassbender, J., and Loewenstein, G. (2008). "Financial Incentive-Based Approaches for Weight Loss." *JAMA*, **300** (22), 2631-7.

Volpp, K., Asch, D.A., Galvin, R., and Loewenstein, G. (2011). "Redesigning Employee Health Incentives – Lessons from Behavioural Economics." *The New England Journal of Medicine*, **365** (5), 388-390.

Wansink, B., and Chandon, P. (2006). "Can 'low fat' nutritional labels lead to obesity?" *Journal of Marketing Research*, **43** (4), 605-617.

Werle, C., Wansink, B., Payne, C. (2010). "Just thinking about exercise makes me serve more food. Physical activity and calorie compensation." *Appetite*, **56** (2), 332-335.

Westerterp, K.R., and Speakman, J.R. (2008). "Physical Activity Energy Expenditure Has Not Declined Since the 1980s and Matches Energy Expenditures of Wild Mammals." *International Journal of Obesity*, **32**, 1256-1263.

Wilcox, K., Block, L., Fitzsimons, G.J., Vallen, B. (2009). "Vicarious Goal Fulfillment: When the Mere Presence of a Healthy Option Leads to an Ironically Indulgent Decision." *Journal of Consumer Research*, **36**, 380-393.

Wisdom, J., Downs, J.S., and Loewenstein, G. (2010). "Promoting Healthy Choices: Information Versus Convenience." *American Economic Journal: Applied Economics*, **2**, 164-78.

Zizzo, D.J. (2010). "Experimenter Demand Effects in Economic Experiments." *Experimental Economics*, **13**, 75-98.

Zhou, X., Vohs, K., and Baumeister, R.F. (2009). "The Symbolic Power of Money: Reminders of Money Alter Social Distress and Physical Pain." *Psychological Science*, **20** (6), 700-706.

Zwane, A.P. *et al*. (2011). Being surveyed can change later behavior and related parameter estimates. *Proceedings of the National Academy of Sciences USA*, **108** (5), 1821-1826.

Appendix: Tables.

Table 1: Distribution of subjects across timeslots and treatments.

Timeslot						
Treatment	**11.00**	**12.00**	**13.00**	**14.00**	**15.00**	***Total***
H	**6**	**10**	**8**	**7**	**9**	**40**
	15.00	*25.00*	*20.00*	*17.50*	*22.50*	*100.00*
	21.43	25.87	25.00	20.59	33.33	25.64
L	**8**	**9**	**7**	**8**	**7**	**39**
	20.51	*23.08*	*17.95*	*20.51*	*17.95*	*100.00*
	28.57	25.71	21.88	23.53	25.93	25.00
E	**8**	**5**	**11**	**10**	**5**	**39**
	20.51	*12.82*	*28.21*	*25.64*	*12.82*	*100.00*
	28.57	14.29	34.38	29.41	18.52	25.00
C	**6**	**11**	**6**	**9**	**6**	**38**
	17.95	*22.44*	*20.51*	*21.79*	*17.31*	*100.00*
Total	**28**	**35**	**32**	**34**	**27**	***156***

Table 2: Description of the Variables and Main Descriptive Statistics

	Independent and dependent variables				
	Description	*Mean*	*Std. Dev.*	*Min*	*Max*
Steps	Number of completed steps	97.583	18.239	51	145
DiffHR	Difference in directly measured heart rate before and after the stepping task (in bpm)	61.953	26.530	3	173
DiffEnergy	Difference in self-assessed 'full of energy' feeling before and after the stepping task (on a 0-10 Likert scale)	0.369	2.059	-6	6
ImSatisf	Self-assessed satisfaction with the performance in the stepping task immediately after the task (on a 0-10 Likert scale)	7.356	1.712	0	10
EndSatisf	Self-assessed satisfaction with the performance in the stepping task at the end of the experiment (on a 0-10 Likert scale)	6.981	1.753	1	10
KcalIn	Total number of calories from food and drink items consumed in the lunch (in Kcal)	335.105	299.444	0	1,504.5
KcalOut	Number of calories burned in the stepping exercise based on subject's gender, age, and directly measured weight and heart rate after the task (in Kcal)	15.582	6.587	0	26.478
ExcessKcal	Computed difference (*TotalKcalIn – CalBurned*)	319.522	299.488	-24.693	1,488.3
EstKcalOut	Subject's estimate of number of calories burned in the stepping exercise (open end)	80.430	122.558	2	1,000
EstKcalIn	Subject's estimate of number of calories from all food and drink items consumed in the lunch (open end)	196.767	238.085	0	1,700
Hunger	Self-assessed level of hunger before the lunch (on a 180 mm slider scale)	52.974	35.937	0	152

Independent and dependent variables (*continued*)					
	Description	*Mean*	*Std. Dev.*	*Min*	*Max*
KcalSandw	Total number of calories from sandwiches consumed in the lunch (in Kcal)	160.663	208.312	0	899
KcalCrisps	Total number of calories from crisps consumed in the lunch (in Kcal)	35.115	61.558	0	264
KcalSweets	Total number of calories from sweets (excluding fruits) consumed in the lunch (in Kcal)	85.692	136.694	0	587
KcalDrinks	Total number of calories from drinks consumed in the lunch (in Kcal)	53.634	63.788	0	327
Fats	Total number of fats intakes consumed in the lunch (in grams)	10.440	10.646	0	43.1
SatFats	Total number of saturated fats intakes consumed in the lunch (in grams)	3.077	3.970	0	18.9
Sugars	Total number of sugars intakes consumed in the lunch (in grams)	23.090	22.193	0	125.9
Salt	Total number of sodium salt intakes consumed in the lunch (in grams)	0.713	0.864	0	4.03
Female	Female gender (dummy)	0.655	0.476	0	1
Age	Age (in years)	25.467	6.288	18	55
Weight	Directly measured weight (in kilograms)	63.767	13.368	42.7	107
RentExpense	Weekly rent expenditure (in GBP per week)	167.229	128.474	0	800
SAH	Self-assessed health (1=very poor, 5=excellent)	3.629	0.816	2	5
HRBefore	Directly measured heart rate before the stepping task (in bpm)	78.52	13.784	11	114
Happy	Baseline happiness level (on a 0-10 Likert scale)	7.107	1.571	3	10
TreatH/L/E	Dummy variables for treatment groups H, L, E respectively	-	-	0	1
SatH/L/E	Interaction terms with *ImSatisf* variable	-	-	-	-
LastEat	Self-reported time passed from last meal (in minutes)	189.532	223.755	0	1,290

Table 3. Baseline characteristics across treatments

	H	L	E	C	Total
Age	25.62	26.51	25.12	24.63	25.47
	(5.69)	(6.53)	(7.87)	(4.72)	(6.28)
Female	0.60	0.65	0.67	0.71	0.65
	(0.49)	(0.48)	(0.47)	(0.46)	(0.47)
Weight	63.4	65.32	63.14	63.21	63.77
	(12.44)	(15.87)	(11.29)	(13.88)	(13.37)
RentExpense	175.81	158.21	173.61	160.73	167.23
	(133.76)	(131.51)	(149.71)	(98.61)	(128.47)
SAH	3.6	3.62	3.59	3.71	3.62
	(0.87)	(0.92)	(0.82)	(0.65)	(0.81)
HRBefore	82.67*	78.05	75.80	77.16	75.82
	(12.95)	(16.42)	(10.43)	(14.21)	(13.78)
Happy	7.10	7.14	6.97	7.21	7.11
	(1.69)	(1.31)	(1.58)	(1.71)	(1.57)
Hunger	48.17	62.05	52.77	49.39	52.97
	(39.41)	(35.27)	(35.24)	(33.11)	(35.94)
Observations	40	39	39	38	156

Standard deviations in parentheses.

Results of two-sample Wilcoxon rank-sum (Mann-Whitney) test between the treatment group and group C:

* p<.10, ** p<.05, *** p<.01.

Table 4. Overview of main results: Wilcoxon rank-sum test between treatment group and C

	H	L	E	C	Total
Steps	102.55***	105.54***	92.49	89.42	97.58
	(13.35)	(19.15)	(18.55)	(16.91)	(18.24)
DiffHR	64.4**	68.67**	63.39*	51.66	61.95
	(22.93)	(28.66)	(25.25)	(27.18)	(26.53)
DiffEnergy	0.225	1.128**	0.243	-0.132	0.368
	(1.746)	(2.105)	(1.716)	(2.462)	(2.059)
ImmSatisf	8.425***	7.064*	7.487***	6.394	7.355
	(1.059)	(1.857)	(1.519)	(1.701)	(1.712)
EndSatisf	7.9***	6.807**	7.128***	6.039	6.981
	(1.516)	(1.768)	(1.417)	(1.817)	(1.753)
KcalIn	432.9***	319.23*	350.12*	233.04	335.10
	(297.81)	(283.49)	(286.44)	(306.26)	(299.44)
KcalOut	16.95**	16.61**	15.402	13.27	15.58
	(5.817)	(6.424)	(6.279)	(7.385)	(6.587)
ExcessKcal	415.94***	302.62	334.72*	219.77	319.52
	(299.69)	(284.07)	(286.23)	(305.34)	(299.48)
EstKcalOut	69.39**	112.12**	84.35	58.94	80.34
	(61.42)	(183.63)	(120.58)	(98.68)	(122.56)
EstKcalIn	200.61	170.83	246.43	169.03	196.77
	(306.89)	(163.81)	(247.00)	(207.05)	(238.08)
Observations	40	39	39	38	156

* p<.10, ** p<.05, *** p<.01.

Table 5: Overview of Nutritional Outcomes, in Detail

	H	L	E	C	Total
KcalSandw	169.52	155.49	173.15	143.83	160.66
	(222.02)	(186.69)	(198.35)	(230.35)	(208.31)
KcalCrisps	79.2***	20.31	25.38	13.89	35.11
	(77.94)	(45.81)	(51.60)	(41.05)	(61.56)
KcalSweets	97.72*	95.54	109.67	38.32	85.69
	(131.72)	(153.33)	(156.08)	(87.47)	(136.69)
KcalDrinks	86.45***	47.89	41.92	37.00	53.63
	(54.25)	(71.37)	(55.35)	(62.74)	(63.78)
Fats	14.80***	9.080**	11.23*	6.431	10.440
	(11.39)	(9.449)	(9.834)	(10.353)	(10.646)
SatFats	3.932***	2.912**	3.401***	2.016	3.078
	(4.517)	(3.448)	(3.493)	(4.206)	(3.970)
Sugars	31.52***	23.60*	22.39	14.40	23.09
	(17.02)	(25.91)	(22.79)	(19.48)	(22.19)
Salt	0.843**	0.654	0.754	0.595	0.713
	(0.934)	(0.754)	(0.809)	(0.954)	(0.864)
Hunger	48.17	62.05	52.77	49.39	52.97
	(39.40)	(35.27)	(35.25)	(33.11)	(35.94)
LastEat	185.65	208.24	174.05	190.39	189.53
	(181.38)	(276.42)	(211.69)	(224.74)	(223.75)
Observations	40	39	39	38	156

Standard deviations in parentheses

Results of two-sample Wilcoxon rank-sum (Mann-Whitney) test between the treatment group and group C:

* p<.10, ** p<.05, *** p<.01.

Table 6: Two-Part Model: Probit estimates for Prob(*KcalIn>0*)
Likelihood to consume food or drink items at lunch after the task
Dependent variable: 1:(*KcalIn>0*); 0:(*KcalIn=0*).

Probability to have lunch

	m1	m2	m3	m4	m5	m6	m7
TreatH	0.727**	0.693**	0.768**	0.792**	0.792**	0.788**	0.747
	(0.345)	(0.349)	(0.358)	(0.364)	(0.377)	(0.363)	(0.36
TreatL	0.465	0.548	0.473	0.458	0.458	0.497	0.36
	(0.325)	(0.342)	(0.351)	(0.356)	(0.372)	(0.362)	(0.36
TreatE	0.181	0.157	0.155	0.270	0.270	0.279	0.27
	(0.309)	(0.311)	(0.318)	(0.330)	(0.331)	(0.331)	(0.34
Female		-0.447	-0.340	-0.469	-0.469	-0.505*	-0.4
		(0.273)	(0.282)	(0.298)	(0.298)	(0.304)	(0.30
Hunger			0.00956**	0.0116***	0.0116***	0.0117***	0.010
			(0.00377)	(0.00412)	(0.00412)	(0.00411)	(0.004
LastEat				-0.000297	-0.000297	-0.000303	-0.000
				(0.00058)	(0.00058)	(0.00058)	(0.000
Steps					- 0.0000100		
					(0.00750)		
DiffEnergy						-0.0405	
						(0.0625)	
DiffHR							0.003
							(0.005
Constant	0.555***	0.885***	0.359	0.414	0.415	0.436	0.27
	(0.215)	(0.299)	(0.362)	(0.377)	(0.775)	(0.380)	(0.44
Observations	156	154	154	152	152	152	14
Pseudo R^2	0.0352	0.0579	0.104	0.124	0.124	0.127	0.11

Standard errors in parentheses
* p<.10, ** p<.05, *** p<.01

Table 7: Two-Part Model: Probit estimates for Prob*(KcalIn>0)* (*Continued*)
Likelihood to consume food or drink items at lunch after the task
Dependent variable: 1:(*KcalIn>0*); 0:(*KcalIn=0*).

Probability to have lunch

	m8	m9	m10	m11
TreatH	0.788**	0.659*	0.374	0.451
	(0.363)	(0.388)	(0.400)	(0.427)
TreatL	0.497	0.236	0.475	0.282
	(0.362)	(0.424)	(0.380)	(0.435)
TreatE	0.279	0.258	0.0560	0.120
	(0.331)	(0.364)	(0.347)	(0.382)
Female	-0.505*	-0.286	-0.403	-0.241
	(0.304)	(0.341)	(0.312)	(0.346)
Hunger	0.0117***	0.0146***	0.0114***	0.0140***
	(0.00411)	(0.00466)	(0.00422)	(0.00470)
LastEat	-0.000303	-0.000595	-0.000355	-0.000612
	(0.000583)	(0.000709)	(0.000586)	(0.000703)
DiffEnergy	-0.0405	-0.0612	-0.0338	-0.0523
	(0.0625)	(0.0713)	(0.0658)	(0.0724)
EstKcalOut		0.0134**		0.0118**
		(0.00520)		(0.00525)
ImSatisf			0.199**	0.105
			(0.0795)	(0.0899)
Constant	0.436	-0.292	-0.864	-0.896
	(0.380)	(0.459)	(0.645)	(0.695)
Observations	152	143	152	143
Pseudo R^2	0.127	0.228	0.172	0.238

Standard errors in parentheses
* p<.10, ** p<.05, *** p<.01

Table 8: Two-Part Model: OLS Estimates for E(*LnExcessKcal*| *ExcessKcal>0*)
Balance between calories consumed and burned
Dependent variable: *LnExcessKcal*

Log of excess calories

	m12	m13	m14	m14	m16
TreatH	0.599**	0.568**	0.596**	0.568**	0.593**
	(0.257)	(0.259)	(0.254)	(0.260)	(0.255)
TreatL	0.305	0.281	0.244	0.246	0.220
	(0.261)	(0.265)	(0.272)	(0.265)	(0.271)
TreatE	0.625**	0.625**	0.614**	0.625**	0.614**
	(0.246)	(0.245)	(0.249)	(0.246)	(0.251)
Female		-0.365**	-0.313*	-0.373**	-0.331**
		(0.164)	(0.172)	(0.158)	(0.165)
Hunger			0.00399		0.00388
			(0.00247)		(0.00267)
LastEat				0.0000608	-0.000062
				(0.000491)	(0.000489)
Constant	5.266***	5.509***	5.254***	5.503***	5.284***
	(0.207)	(0.242)	(0.277)	(0.253)	(0.271)
Observations	126	125	125	124	124
R^2	0.0695	0.105	0.127	0.110	0.130
Adjusted R^2	0.0467	0.0751	0.0903	0.0720	0.0849

Heteroskedatisc-robust standard errors
Standard errors in parentheses
* p<.10, ** p<.05, *** p<.01

Table 9: Two-Part Model: OLS Estimates for E(*LnExcessKcal*| *ExcessKcal>0*) (*Continued*)
Balance between calories consumed and burned
Dependent variable: *LnExcessKcal*

Log of excess calories

	m17	m18	m19	m20	m21	m22	m23
TreatH	0.596**	0.635**	0.636**	0.613**	0.661**	0.650**	0.634**
	(0.254)	(0.265)	(0.256)	(0.253)	(0.266)	(0.255)	(0.264)
TreatL	0.244	0.295	0.305	0.296	0.337	0.357	0.326
	(0.272)	(0.299)	(0.278)	(0.287)	(0.300)	(0.291)	(0.308)
TreatE	0.614**	0.620**	0.650**	0.642**	0.651**	0.678**	0.638**
	(0.249)	(0.254)	(0.255)	(0.263)	(0.257)	(0.265)	(0.263)
Female	-0.313*	-0.323*	-0.346**	-0.312*	-0.350**	-0.349**	-0.321*
	(0.172)	(0.171)	(0.168)	(0.178)	(0.169)	(0.174)	(0.177)
Hunger	0.00399	0.00401	0.00428*	0.00405	0.00426*	0.00434	0.0039
	(0.00247)	(0.00249)	(0.00247)	(0.00267)	(0.00248)	(0.00265)	(0.0027
Steps		-0.00293			-0.00225		-0.0025
		(0.00510)			(0.00519)		(0.0060
DiffEnergy			-0.0374		-0.0333	-0.0386	
			(0.0439)		(0.0453)	(0.0453)	
DiffHR				-0.00111		-0.000881	-0.00032
				(0.00346)		(0.00346)	(0.0040
Constant	5.254***	5.521***	5.235***	5.308***	5.442***	5.278***	5.503**
	(0.277)	(0.489)	(0.277)	(0.312)	(0.510)	(0.317)	(0.514)
Observations	125	125	125	120	125	120	120
R^2	0.127	0.130	0.132	0.123	0.134	0.128	0.125
Adjusted R^2	0.0903	0.0856	0.0881	0.0763	0.0821	0.0740	0.0698

Heteroskedatisc-robust standard errors
Standard errors in parentheses
* p<.10, ** p<.05, *** p<.01

Table 10: Two-Part Model: OLS Estimates for E(*LnExcessKcal*| *ExcessKcal>0*) (*Continued*)
Balance between calories consumed and burned
Dependent variable: *LnExcessKcal*

Log of excess calories

	m24	m25	m26	m27	m28	m29
TreatH	0.596**	0.582**	0.335	-2.010	-1.444	
	(0.254)	(0.259)	(0.294)	(1.325)	(1.392)	
TreatL	0.244	0.370	0.128	0.254	0.822	
	(0.272)	(0.276)	(0.280)	(0.567)	(0.684)	
TreatE	0.614**	0.604**	0.468*	-1.646**	-1.076	
	(0.249)	(0.253)	(0.261)	(0.801)	(0.887)	
Female	-0.313*	-0.299*	-0.254	-0.237	-0.241	-0.249
	(0.172)	(0.180)	(0.171)	(0.171)	(0.172)	(0.170)
Hunger	0.00399	0.00256	0.00416	0.00443*	0.00434*	0.00434*
	(0.00247)	(0.00250)	(0.00251)	(0.00249)	(0.00250)	(0.00249)
EstKcalOut		0.000245				
		(0.000418)				
ImSatisf			0.149**		0.0853	0.114*
			(0.0587)		(0.0781)	(0.0677)
SatH				0.311**	0.225	0.0513
				(0.147)	(0.167)	(0.0380)
SatL				-0.00112	-0.0867	0.0169
				(0.0752)	(0.107)	(0.0405)
SatE				0.297***	0.211*	0.0713**
				(0.0966)	(0.123)	(0.0359)
Constant	5.254***	5.306***	4.215***	5.179***	4.621***	4.426***
	(0.277)	(0.289)	(0.479)	(0.277)	(0.509)	(0.476)
Observations	125	118	125	125	125	125
R^2	0.127	0.112	0.176	0.211	0.216	0.188
Adjusted R^2	0.0903	0.0642	0.135	0.157	0.155	0.147

Heteroskedatisc-robust standard errors
Standard errors in parentheses
* p<.10, ** p<.05, *** p<.01

Table 11: Main nutritional intakes in each food and drink item

Nutritional intakes					
	Kcal	*Fats*	*Saturated Fats*	*Sugars*	*Salt*
Bacon, lettuce, and tomato	396	14.3	4.1	7.1	1.5
Chicken and bacon	470	16	3	5	1.68
Ham and cheese	459	18.7	7.8	6.2	1.78
Tuna and sweet corn	298	6.7	0.7	4.2	1.09
Roast chicken salad (*low fat*)	281	3.7	0.5	2.6	0.9
Bacon and chicken (*low fat*)	364	5.6	2	2.9	1.61
Tuna and cucumber (*low fat*)	290	4.6	0.7	3.4	1.3
Cheese and tomato	401	19.6	12.7	3.9	1.74
Cheese and celery	380	18.9	9.8	3.7	1.36
Crisps	132	8.1	0.5	0.3	0.19
Muffin	293	13.9	2.5	20.4	0.29
Chocolate bar	247	12.5	6.2	25.5	0
Apple	47	0.1	0	11.8	0
Orange juice	94	0.2	0	20.0	0
Coke	139	0	0	35	0

CENTRE FOR ECONOMIC PERFORMANCE
Recent Discussion Papers

1285	Swati Dhingra	Reconciling Observed Tariffs and the Median Voter Model
1284	Brian Bell Anna Bindler Stephen Machin	Crime Scars: Recessions and the Making of Career Criminals
1283	Alex Bryson Arnaud Chevalier	What Happens When Employers are Free to Discriminate? Evidence from the English Barclays Premier Fantasy Football League
1282	Christos Genakos Tommaso Valletti	Evaluating a Decade of Mobile Termination Rate Regulation
1281	Hannes Schwandt	Wealth Shocks and Health Outcomes: Evidence from Stock Market Fluctuations
1280	Stephan E. Maurer Andrei V. Potlogea	Fueling the Gender Gap? Oil and Women's Labor and Marriage Market Outcomes
1279	Petri Böckerman Alex Bryson Jutta Viinikainen Christian Hakulinen Laura Pulkki-Raback Olli Raitakari	Biomarkers and Long-term Labour Market Outcomes: The Case of Creatine
1278	Thiemo Fetzer	Fracking Growth
1277	Stephen J. Redding Matthew A. Turner	Transportation Costs and the Spatial Organization of Economic Activity
1276	Stephen Hansen Michael McMahon Andrea Prat	Transparency and Deliberation within the FOMC: A Computational Linguistics Approach
1275	Aleksi Aaltonen Stephan Seiler	Quantifying Spillovers in Open Source Content Production: Evidence from Wikipedia
1274	Chiara Criscuolo Peter N. Gal Carlo Menon	The Dynamics of Employment Growth: New Evidence from 18 Countries

1273	Pablo Fajgelbaum Stephen J. Redding	External Integration, Structural Transformation and Economic Development: Evidence From Argentina 1870-1914
1272	Alex Bryson John Forth Lucy Stokes	Are Firms Paying More For Performance?
1271	Alex Bryson Michael White	Not So Dissatisfied After All? The Impact of Union Coverage on Job Satisfaction
1270	Cait Lamberton Jan-Emmanuel De Neve Michael I. Norton	Eliciting Taxpayer Preferences Increases Tax Compliance
1269	Francisco Costa Jason Garred João Paulo Pessoa	Winners and Losers from a Commodities-for-Manufactures Trade Boom
1268	Seçil Hülya Danakol Saul Estrin Paul Reynolds Utz Weitzel	Foreign Direct Investment and Domestic Entrepreneurship: Blessing or Curse?
1267	Nattavudh Powdthavee Mark Wooden	What Can Life Satisfaction Data Tell Us About Discrimination Against Sexual Minorities? A Structural Equation Model for Australia and the United Kingdom
1266	Dennis Novy Alan M. Taylor	Trade and Uncertainty
1265	Tobias Kretschmer Christian Peukert	Video Killed the Radio Star? Online Music Videos and Digital Music Sales
1264	Diego Battiston Richard Dickens Alan Manning Jonathan Wadsworth	Immigration and the Access to Social Housing in the UK
1263	Alexandre B. Cunha Emanuel Ornelas	Political Competition and the Limits of Political Compromise
1262	Alexander C. Lembcke	The Impact of Mandatory Entitlement to Paid Leave on Employment in the UK

The Centre for Economic Performance Publications Unit
Tel 020 7955 7673 Fax 020 7404 0612
Email info@cep.lse.ac.uk Web site http://cep.lse.ac.uk